Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies
Lunn, Brian Kirkegaard

Publication date:
2008

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
? You may not further distribute the material or use it for any profit-making activity or commercial gain
? You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies:
Towards the design of surrogate records

Brian Kirkegaard

PhD thesis from Research Programme Information Interaction and Information Architecture
Royal School of Library and Information Science, Denmark
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies: Towards the design of surrogate records

Brian Kirkegaard
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies: Towards the design of surrogate records / Brian Kirkegaard. – Aalborg, Denmark: Royal School of Library and Information Science, 2008. – XVII, 311, plus Appendix volume.
Available: http://biblis.db.dk/uhtbin/hyperion.exe/db.brikir07
ISBN 978-87-7415-305-4
Medievidenskabelige forskere og studerendes foretrukne metadataelementer ved søgning og relevansvurdering af tv-udsendelser i relation til design af surrogatposter

Brian Kirkegaard

Ph.d.-afhandling fra Forskningsprogrammet Informationsinteraktion og Informationsarkitektur
Danmarks Biblioteksskole
Acknowledgement

I wish to express my acknowledgement of, and indebtedness to, all of my colleagues, friends, and family who have supported me in my doctoral work. The encouragement, inspiration, moral support, advice, and constructive criticism have been priceless. The list of persons to whom I owe my gratitude is extensive, and with the risk of neglecting to mention persons that genuinely deserve the expression of my indebtedness, I will put a name to a few of these supportive people.

First of all, I want to thank my principal supervisor Associate Professor and Head of Research Programme Pia Borlund. Pia, you have taking me under your wing in every way, you have been a wonderful supervisor, and I owe everything to you. I am gratefully indebted to you for the immense work you have put into it, and during the last three years you have truly made me appreciate red ink, owed to the fact that your comments have been pearls of wisdom. Your eternal confidence in me and my work have provided an invaluable encouragement in good times as well as when my own faith experienced lapses. Your teaching and dedication to LIS inspired me to apply for the doctoral scholarship in the first place, and you have been a first-class guide on the, at times, bumpy PhD journey.

In addition to my principal supervisor, I am indebted to numerous colleagues and friends. Especially Associate Professors Marianne Lykke Nielsen, Jesper W. Schneider, and Birger Larsen, PhD student Mette Skov, and Professor Peter Ingwersen. Thank you Marianne and Jesper for always being there for me and for providing encouragement whenever it was needed. Mette, despite the distance you have been a dear colleague, and knowing you have really made me appreciate Graham Bells 130 year old “electrical speech machine”. Peter I am very grateful for your kindness and your willingness to chair my assessment committee. While we are at it, I want to thank Professor Eero Sormunen and Associate Professor Morten Hertzum for joining the review committee. I know it has been hard work reviewing the thesis, and I really appreciate your evaluation and constructive criticism.

I would also take the opportunity to thank the many kind and helpful people I have meet outside the research community of the Royal School of Library and Information Science. In particular, I am thankful to the NorsLIS community for providing constructive and knowledge stimulating doctoral courses and workshops, and for allocating grants. Senior researchers as well as fellow PhD students have provided the constructive and thought stimulating environment found in the courses. In
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

particular, I would take the opportunity to fellow doctoral students, now PhDs, and friends Kora Golub and David Elsweiler.

I am also grateful for my research visits at Dublin and Aarhus University. Thank you Professor Alan Smeaton, Senior Lecturer Dr. Gareth Jones, and Lecturer Dr. Cathal Gurrin for making the research visit at the Centre of Digital Video Processing at Dublin University possible. I am also thankful to Associate Professor Jørgen Bang concerning the research visit at the Department of Information and Media Studies at Aarhus University. This research visit was essential for my data collection and I am indebted to the test participants from the Media Studies departments at Aarhus University as well as at Copenhagen University, and in particular I am grateful to the nine test participants who participated in interviews.

I want to thank the State and University Library for providing the topical frame for the thesis, as well as for the corporation during the project, and especially for ensuring a workplace for me at the library. Several persons from the library have contributed directly or indirectly to my work. I want to thank Director of Development, Birte Christensen-Dalsgaard for being my project supervisor. The Research Forum at the library, have shown a keen interest in my work and in particular, I wish to express my gratitude to Head of Research and Special Collections Harald von Hielmcrone and Senior Research Fellow, now retired, Helge Clausen for taking a genuine interest in my work. The staff at the State Media Archive has been very open and helpful throughout the project, and I owe all of you my gratitude.

I also owe gratitude to the administration and IT department at the Royal School of Library and Information Science. Thank you Bo, Lisbeth, Dorte, Mimi, Tove, Lisbeth, Jette, Lise Lotte, and Frank. In particular, I am thankful for Clerk Mimi Lindegaard Hansen’s support in formatting the troublesome word-file, and Librarians Jette Thise Pedersen and Hanne Guldhammer Møller for providing swift library support when needed. I would also like to thank Assistant Professor Jan Steffensen for programming the web-questionnaire in ASP.

Lastly, I would take the opportunity to express my deepest gratitude to my friends and family for holding in throughout the long periods where socializing with me has been drudgery. Thank you for your patience despite the fact that no, no, and absolutely no have been my stock answer for any social activity, and consequently my attendance mainly have been of an imaginary kind. Your persistence is an assurance of your true friendship.

Warmest of all is my gratitude to Tine, my wife in spe. Tine your unconditional love, support and encouragement in good as well as bad times have been tremendous and incomparable. Thank you for always being there and for always believing in me.
Abstract

The present doctoral work concerns the investigation of three aspects of scholars’ and students’ information seeking behaviour in a television broadcast context, and the associated implications for design and construction of metadata elements in surrogate records in future broadcast retrieval systems. The three aspects of information seeking behaviour in focus are information need characteristics, preferred search entries, and application of relevance criteria.

The thesis provides considerations regarding the construction of surrogate records for broadcasts in the Danish national collection of television broadcasts at the State and University Library in Aarhus. At present, no surrogate records exist for the collection, and hence no retrieval facilities are available for users. To gather knowledge about the construction of surrogate records for a future broadcasts retrieval system, the present work focuses on three interconnected research areas: 1) aspects of users’ information seeking behaviour in relation to television broadcasts; 2) appropriate access points for television broadcasts; and 3) construction of surrogate records for television broadcasts.

Users’ information seeking behaviour is the behaviour connected to the purposive actions when seeking information in order to fulfil an underlying work task (Wilson, 2000, p. 49). The present work inquires into three aspects of academics’ information seeking behaviour in a television broadcast context, namely: 1) characteristics of the users’ information needs; 2) the users’ search entry preferences; and 3) the users’ application of relevance criteria. Investigation of these aspects of the academics’ information seeking behaviour reveals which television broadcast dimensions that are of importance for effective retrieval of television broadcasts. Knowledge about these dimensions is translated to knowledge about the access points that are appropriate in surrogate records in future broadcast retrieval systems. The appropriate access points are the metadata elements in surrogate records that ensure that users are provided with effective retrieval for fulfilment of their work task at hand. Subsequently to identification of appropriate access points, we consider indexing approaches for construction of the access point in surrogate records. The primary focus is on exploitation of information in readily available external sources, namely television schedules. In addition, we cursorily discuss the benefits of exchanging surrogate records with operating archives.

The epistemological foundation for the doctoral work is the cognitive integrated framework for information seeking and information retrieval (IR) (Ingwersen & Järvelin, 2005). The framework builds upon the fundamental conception that
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

concurrent consideration of research in information seeking and IR is advantageous for Library and Information Science research. Further, the framework contests that in order to create an effective broadcast retrieval system, it is important to gather knowledge about users’ information seeking behaviour in the context of television broadcast.

Methodologically the thesis is based on grounded theory (Glaser & Strauss, 1967; Strauss & Corbin, 1998), which is in concordance with the cognitive integrated framework. Grounded theory emphasises inductive generation of theories, models and/or hypothesis from raw empirical data, rather than a deductive test of already established theories, models and/or hypothesis. The thesis applies a triangulation of data collecting and data analysing methods, e.g., web questionnaire and in-depth follow-up interviews to reveal the academics’ information needs, and preferred access points for searching and assessing the relevance of television broadcasts.

With reference to the first research area concerning aspects of users’ information seeking behaviour in relation to television broadcasts, the investigation identifies five characteristics of the academics’ information needs. These characteristics include that the academics’ broadcast needs are for known broadcasts, factual data about the broadcasts, or of a topical nature, whether the perception of the topic is clear or muddled. With reference to research area two concerning identification of appropriate access points for television broadcasts, we find that broadcasts are mainly needed at the programmes level of granularity, but access at the serial and feature levels of granularity is also needed. Further, the investigation identifies 28 search entries, and 32 relevance criteria. Subsequently, 24 access points are identified to be appropriate in relation to a future broadcast retrieval system, including novel as well as previously identified access points. The function of each access point is discussed, and recommendations provided regarding the value of the access points in a future broadcast retrieval system. With reference to the third research area concerning the construction of access points in surrogate records for television broadcasts, the results show that the information available in the television schedules is generally useful for descriptive indexing, while of no or minor value for subject indexing. Exchange of surrogate records is valuable as a supplement for construction of the identified appropriate access points, only.

In brief, the thesis expands our knowledge on academics’ information needs and access points for searching and relevance assessment in a television broadcast context, and the implications for design of surrogate records that are essential for facilitating the recognized behaviour. Hereby, it provides novel research, which is important for construction of surrogate records in future broadcast retrieval systems.
Resumé

Afhandlingens formål er at undersøge tre aspekter af akademikeres informationssøgeadfærd i relation til genfinding af tv-udsendelser, og identificere de implikationer adfærdenden har for design af surrogatposter i et fremtidigt genfindingssystem. De tre aspekter er karakteristika ved brugernes informationsbehov, foretrukne søgeindgange og anvendelsen af relevanskriterier i en tv-udsendelseskontekst.

Afhandlingen tager afsæt i Statsbibliotekets nationale samling af tv-udsendelser. For nuværende er der ikke konstrueret surrogatposter for tv-udsendelserne i samlingen, og der er derfor ingen søgefaciliteter til rådighed for brugerne. For at indsamle viden om konstruktion af surrogatposter for et fremtidigt genfindingssystem til tv-udsendelser, fokuseres der på tre forskningsområder: 1) aspekter ved brugerens informationssøgeadfærd i relation til tv-udsendelser, 2) hensigtsmæssige dataelementer til tv-udsendelser og 3) konstruktion af surrogatposter for tv-udsendelser.

Brugerens informationssøgeadfærd er den adfærd der er knyttet til bevidste handlinger i forbindelse med søgning af information med henblik på at fuldføre en underliggende arbejdspåfald (Wilson, 2000, p. 49). I afhandlingen fokuseres der på tre aspekter af akademikeres søgeadfærd i relation til tv-udsendelser: 1) karakteristika ved informationsbehov, 2) foretrukne søgeindgange og 3) anvendelse af relevanskriterier. En afdækning af brugerens informationssøgeadfærd er vigtig i forhold til at kunne identificere de dimensioner, brugerne finder væsentlige ved genfinding af tv-udsendelser. Denne viden anvendes til at bestemme de metadataelementer, der er hensigtsmæssige at inkludere i surrogatposter. Hensigten hermed er at sikre, at brugerne oplever effektiv genfinding. Efterfølgende undersøges det hvordan de identificerede metadataelementer kan indekseres. Det primære fokus ligger på anvendeligheden af informationer, der er tilgængelige i tekstruelle eksterne kilder, og specifikt tv-programoversigter. Sekundært er der fokus på værdien ved udveksling af surrogatposter med eksisterende arkiver.

Epistemologisk er afhandlingsarbejdet funderet i det kognitive synspunkt (Ingwersen & Järvelin, 2005). Dette kommer til udtryk i en grundlæggende antagelse om, at viden omkring brugerens informationsbehov er en forudsætning for at konstruere effektive informationsgenfindingssystemer.

primære data indsamlingsmetoder der anvendes er web-spørgeskema og dybdegående interviews.


Afhandlingen bidrager med ny viden omkring akademikeres informationsbehov, og foretrukne metadataelementer ved søgning og relevansvurdering i en tv-udsendelseskontekst, samt de implikationer den identificerede adfærd har for design af surrogatposter i fremtidige effektive genfindingssystemer for tv-udsendelser.
# Table of Content

1 Introduction ........................................................................................................................................ 1

1.1 Background .................................................................................................................................. 3
1.1.1 Background – the Danish national collection of television broadcasts ....................................... 3
1.1.2 Background – the cognitive viewpoint ......................................................................................... 4
1.1.3 Background – information seeking behaviour ............................................................................. 5
1.1.4 Background – indexing ............................................................................................................. 7
1.2 Objectives of the thesis ................................................................................................................. 7
1.3 Research questions ....................................................................................................................... 9
1.4 Structure of the thesis .................................................................................................................. 11

2 A cognitive framework for information seeking and information retrieval ........................................ 15

2.1 Information seeking and information retrieval ............................................................................ 16
2.2 The cognitive foundation of the integrated framework .................................................................. 18
2.2.1 Processing takes place in senders and recipients .................................................................... 20
2.2.2 Processing takes place at different levels .................................................................................. 21
2.2.3 Cognitive actors are influenced by their experiences and environments .................................... 22
2.2.4 Cognitive actors influence their environment ........................................................................... 25
2.2.5 Information is situational and contextual ............................................................................... 26
2.3 The integrated framework for information seeking and retrieval ............................................... 26
2.4 The cognitive viewpoint and the thesis ....................................................................................... 29
2.5 Summary ..................................................................................................................................... 32

3 Information seeking behaviour ........................................................................................................ 35

3.1 The potentially developing and dynamic information need .......................................................... 36
3.1.1 Information need types ............................................................................................................. 38
3.2 Models and theories of information seeking behaviour ............................................................... 41
3.2.1 B. Dervin – The Sense-Making Approach ............................................................................. 42
3.2.2 D. Ellis – Information seeking features .................................................................................... 43
3.2.3 C.C. Kuhlthau – The Information Search Process .................................................................... 46
3.2.4 K. Byström & K. Järvelin – Task complexity, information types, and information sources ........ 48
3.2.5 P. Vakkari – Task based information seeking ......................................................................... 49
3.3 Humanists’ information seeking behaviour ................................................................................... 52
3.3.1 The humanist researcher’s work stages .................................................................................... 52
3.3.2 Information seeking in the humanities ..................................................................................... 54
3.4 Information seeking in the context of still and moving images ................................................... 57
3.4.1 Typologies of requests ............................................................................................................ 57
3.4.2 Searching behaviour ................................................................................................................ 65
3.4.3 Work with and use of television broadcasts ............................................................................. 67

VII
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

3.5 Relevance .................................................................................................................. 71
  3.5.1 The concept of relevance ............................................................................... 72
  3.5.1.1 Multidimensional and situational relevance .............................................. 73
  3.5.2 Relevance criteria ......................................................................................... 77
    3.5.2.1 Relevance criteria and moving images ............................................... 79
3.6 Summary ............................................................................................................... 82

4 Construction of access points in surrogate records: indexing .................. 85
  4.1 The process of indexing .................................................................................... 86
    4.1.1 The concept of aboutness .......................................................................... 87
    4.1.2 Indexing of different types of information content ...................................... 89
  4.2 Approaches for indexing of moving images ....................................................... 91
    4.2.1 The concept-based approach ..................................................................... 91
    4.2.2 The content-based approach .................................................................... 94
  4.3 External sources for indexing of primary information objects ....................... 99
4.4 Summary ............................................................................................................ 101

5 Empirical context ........................................................................................................ 105
  5.1 Television broadcasting in Denmark ................................................................. 105
  5.2 Access to Danish television broadcasts ............................................................. 107
    5.2.1 The collection of television broadcasts at the State and University Library .................................................................................................................. 107
    5.2.2 Other alternatives for searching and accessing Danish television broadcasts .................................................................................................................. 110
  5.3 User target group ............................................................................................... 113
    5.3.1 Users of television programmes from the State and University Library .................................................................................................................. 113
    5.3.2 Selection of user target group ...................................................................... 116
    5.3.3 Introduction to the field of Media Studies ................................................. 117

6 Methodical approach ............................................................................................... 121
  6.1 Methodological background ............................................................................. 121
    6.1.1 Triangulation of research strategies .......................................................... 122
  6.2 Domain knowledge: informal information gathering ....................................... 124
    6.2.1 Television broadcasting and archiving ....................................................... 124
    6.2.2 The need for and work with television broadcasts .................................... 126
    6.2.3 The field of Media Studies ........................................................................ 129
  6.3 Semi-structured web questionnaire survey ....................................................... 131
    6.3.1 Design and construction of the questionnaire ............................................ 131
    6.3.2 Piloting ...................................................................................................... 134
    6.3.3 Collecting the data: respondents and invitations to participate .................. 135
    6.3.4 Respond rates ............................................................................................ 136
    6.3.5 Analysing the collected data ...................................................................... 138
6.4 Open-ended interviews ................................................................. 139
6.4.1 The interview guide .................................................................. 140
6.4.2 Collecting the data: respondents ............................................. 143
6.4.3 Transcribing, coding and analysing the collected data............. 147
6.5 Group discussion with scholars .................................................. 150
6.6 Methodical overview: timeline for data collection ...................... 150

7 Research area 1: aspects of information seeking behaviour ............... 153

7.1 Research question 1: what characterises a given group of users’ information needs in the context of television broadcasts? ......................... 153
7.1.1 The work tasks underlying the nine interviewees’ television needs .... 154
7.1.2 Information needs expressed in the nine interviews .................. 158
    7.1.2.1 Interviewee A - Television programme format versions ............ 158
    7.1.2.2 Interviewee B - Children’s reception and use of moving images .. 159
    7.1.2.3 Interviewee C - The management style in two private companies. 161
    7.1.2.4 Interviewee D - The genre of traveller guides ..................... 162
    7.1.2.5 Interviewee E - The historical development of Danish entertainment programmes ........................................ 164
    7.1.2.6 Interviewee F - The historical development of Danish sports journalism ................................................................. 166
    7.1.2.7 Interviewee G - The historical development of Danish news broadcasts ................................................................. 167
    7.1.2.8 Interviewee H - The historical development of Danish documentaries ................................................................. 168
    7.1.2.9 Interviewee I - Genre and reception analysis of Danish soap operas ........................................................................ 172
    7.1.2.10 Summary of the nine interviewees’ information needs ......... 174
7.1.3 Summary and concluding statements ........................................ 176
7.2 Research question 2: which search entries are preferred for searching of television broadcasts, and why and when are they preferred? ........... 191
    7.2.1 Bibliographic search entries .................................................... 194
    7.2.2 Screening search entries .......................................................... 196
    7.2.3 Content search entries ............................................................. 197
    7.2.4 Archival search entries .............................................................. 199
    7.2.5 Relational search entries .......................................................... 199
    7.2.6 Comparison between search entries and affiliation or academic status ................................................................. 201
    7.2.7 Discussion and summarising statements ................................... 203
7.3 Research question 3: which relevance criteria are applied when evaluating television broadcasts, and why and when are they applied? ............ 208
    7.3.1 Bibliographic relevance criteria ................................................ 210
    7.3.2 Screening relevance criteria .................................................... 211
    7.3.3 Content relevance criteria ....................................................... 212
    7.3.4 Archival relevance criteria ....................................................... 214

IX
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

7.3.5 Relational relevance criteria ........................................................................... 214
7.3.6 Comparison between relevance criteria and affiliation or academic status ........................................ 215
7.3.7 Discussion and summarising statements ......................................................... 216
7.4 Summary and discussion of aspects of seeking behaviour: research area 1 221

8 Research area 2: access points for television broadcasts 225

8.1 Research question 4: which access points in a surrogate record are appropriate for television broadcasts, and what are their functions? 226
8.1.1 Bibliographic access points ........................................................................... 227
8.1.1.1 Title and Statement of responsibility .......................................................... 227
8.1.1.2 Production ................................................................................................. 229
8.1.1.3 Physical description .................................................................................. 230
8.1.2 Screening access points .................................................................................. 231
8.1.3 Content access points .................................................................................... 233
8.1.3.1 Textual content access points .................................................................. 234
8.1.3.2 Visual content access points ..................................................................... 236
8.1.4 Archival access points .................................................................................... 236
8.1.5 Relational access points .................................................................................. 237
8.1.5.1 Serial ......................................................................................................... 237
8.1.5.2 Features .................................................................................................... 238
8.1.5.3 External .................................................................................................... 238
8.1.6 Discussion and summarising statements ......................................................... 239

9 Research area 3: construction of surrogate records 245

9.1 Research question 5: to what extent does external textual and other sources contain or provide information valuable for constructing the identified access points? 246
9.1.1 Segmentation of broadcasts ........................................................................... 249
9.1.2 Bibliographic access points ........................................................................... 251
9.1.2.1 Title and statement of responsibility .......................................................... 251
9.1.2.2 Production ................................................................................................. 253
9.1.2.3 Physical description .................................................................................. 254
9.1.3 Screening access points .................................................................................. 255
9.1.4 Content access points .................................................................................... 256
9.1.4.1 Textual content access points .................................................................. 257
9.1.4.2 Visual content access points ..................................................................... 259
9.1.5 Archival access points .................................................................................... 261
9.1.6 Relational access points .................................................................................. 261
9.1.6.1 Serial ......................................................................................................... 261
9.1.6.2 Features .................................................................................................... 262
9.1.6.3 External .................................................................................................... 263
9.1.7 Access points in existing IR systems for television broadcasts ..................... 263
9.1.8 Discussion and summarising statements ......................................................... 266
10 Summary and conclusions ............................................................................................................................................ 271

10.1 Summary of thesis objectives and results .................................................................................................. 271
10.2 Discussion and recommendations to future work ..................................................................................... 281
10.3 Concluding remarks ................................................................................................................................... 283

References ................................................................................................................................................................. 285
## List of Appendices (Part II of the thesis)

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 1</td>
<td>The coverage of Danish television broadcasting at the State and University Library</td>
<td>1</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>Electronic requisition form for television broadcasts from the State and University Library</td>
<td>9</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>Media Studies student papers</td>
<td>11</td>
</tr>
<tr>
<td>Appendix 4</td>
<td>Web questionnaire</td>
<td>15</td>
</tr>
<tr>
<td>Appendix 5</td>
<td>Television schedule from DR1</td>
<td>21</td>
</tr>
<tr>
<td>Appendix 6</td>
<td>E-mail invitation to participate in piloting the web questionnaire</td>
<td>37</td>
</tr>
<tr>
<td>Appendix 7</td>
<td>E-mail invitation to participate in the web questionnaire</td>
<td>39</td>
</tr>
<tr>
<td>Appendix 8</td>
<td>Interview guide</td>
<td>49</td>
</tr>
<tr>
<td>Appendix 9</td>
<td>Invitation to participate in interview</td>
<td>51</td>
</tr>
<tr>
<td>Appendix 10</td>
<td>Verbatim Danish version of the quotes used in the thesis, the full Danish transcription, and the questionnaire responses</td>
<td>53</td>
</tr>
<tr>
<td>Appendix 11</td>
<td>Presentation at the group discussion session</td>
<td>63</td>
</tr>
<tr>
<td>Appendix 12</td>
<td>Statistics for number of types of television broadcasts marked by respondents</td>
<td>71</td>
</tr>
<tr>
<td>Appendix 13</td>
<td>Test of association between academic status and previous experience with the State and University Library</td>
<td>73</td>
</tr>
<tr>
<td>Appendix 14</td>
<td>Test of association between types of broadcasts and level of granularity</td>
<td>75</td>
</tr>
<tr>
<td>Appendix 15</td>
<td>Statistics for the number of search entries marked by respondents</td>
<td>77</td>
</tr>
<tr>
<td>Appendix 16</td>
<td>Example of a ‘Based on the book’ relation in a motion picture bibliographic record</td>
<td>79</td>
</tr>
<tr>
<td>Appendix 17</td>
<td>Statistics for the number of relevance criteria marked by respondents</td>
<td>81</td>
</tr>
<tr>
<td>Appendix 18</td>
<td>Test of association between search entries and relevance criteria</td>
<td>83</td>
</tr>
</tbody>
</table>
Appendix 19: Example of horizontal and vertical programming 85
Appendix 20: Statistical relation between the granularity level of ‘Sequences’ and the relevance criteria ‘Description of images sequences’ 87
Appendix 21: Surrogate records from DR 89
Appendix 22: Surrogate record from Centre for Educational Service at University College Jutland 95
Appendix 23: Surrogate records from TV-Meter 97
Appendix 24: Surrogate records from the Registrant 101
List of Figures

Figure 2.1: Nested model of information behaviour (Ingwersen & Järvelin, 2005, p. 198). .................................................................17

Figure 2.2: The holistic cognitive framework for information seeking and IR seen from a social and physical level, and as perceived from the information seekers cognitive-emotional level (Ingwersen & Järvelin, 2005, p. 278). Arrow legend corresponds to Ingwersen and Järvelin (2005, p. 274). .................................................................28

Figure 2.3: The interrelationship between the judgment of situational relevance and the development of the information need during a dynamic and interactive IR session (Borlund, 2000, p. 44)...........................31

Figure 3.1: A stage process version of Ellis’ behavioural framework (Wilson, 1999b, p. 255). ............................................................................44

Figure 3.2: Model of the Information Search Process (ISP) (Kuhlthau, 2004, p. 82). .............................................................................47

Figure 3.3: Vakkari’s model of the information search process in task performance (Vakkari, 2001a, p. 308). ................................................50

Figure 3.4: A process version of literary critics’ research phases. ......................54

Figure 3.5: Basic classification of user requests for (moving) images. Examples taken from a television broadcast context.................................64

Figure 3.6: Markkula and Sormunen’s stage model of work processes in television broadcasts making (Markkula & Sormunen, 2006, p. 195)......67

Figure 4.1: Structural units of moving images in a television context ..................96

Figure 5.1: Number of requests from individual users ........................................114

Figure 5.2: Distribution of academic users’ affiliation to branches of science.........116

Figure 7.1: Four part of the test participants’ television broadcast needs. ............189
List of Tables

Table 3.1: Examples of Enser and McGregor’s four categories of visual requests and the distribution of request in each category. Adapted from Enser and McGregor (1993, pp. 17-18) and Enser (1993, p. 30)..........................58

Table 3.2: Armitage and Enser’s categorisation of visual information requests. Examples from Armitage and Enser (1997, pp. 290, 292), and results from Armitage and Enser (1996, p. 31), and Sandom and Enser (2002, p. 38). ........................................................................................................60

Table 3.3: Preferred facilities for searching moving images. Adapted from Sandom and Enser (2002, p. 46). ........................................................................................................66

Table 3.4: Relevance criteria for moving images identified. Modified from Sandom and Enser (2002, p. 46) by including percentages. .................................80

Table 4.1: Types of aboutness in relation to Soergel’s (1985, pp. 226-227) two kinds of indexing, and whether the information objects are being analysed or not. ........................................................................................................88

Table 4.2: Comparison of terminology for kinds of attributes for indexing and retrieval of moving images, and the relation to indexing and types of aboutness. ........................................................................................................95

Table 5.1: Broadcasts in the Danish national collection of television broadcasts distributed on station, and channel, and with indication of coverage in hours and percentages. ................................................................................109

Table 5.2: User affiliations........................................................................................................115

Table 6.1: Respond rate for web questionnaire survey, distributed in accordanceance to academic department and academic status. .................................136

Table 6.2: Six criteria for purposive sampling ........................................................................145

Table 6.3: General overview of the data collection process...................................................151

Table 6.4: Overview of the most active part of the data collection process, which occurred in the spring 2006 ...........................................................................152

Table 7.1: Listing of the work tasks of the nine interviews .......................................................157

Table 7.2: Summary of the nine interviewees’ information needs, the levels of granularity of needed broadcasts, and the type of broadcasts needed. The numbers in the ‘Types of information need’ column refer to Ingwersen and Järvelin’s (2005, pp. 292-293) typology of information needs.............................................................................................................175

Table 7.3: Distribution of the typically broadcasts needed by the respondents. The table is arranged similar to the arrangement of the types of
broadcasts in the questionnaire, which is alphabetical according to the Danish terminology.................................................................177

Table 7.4: Distribution of types of broadcasts typically needed for fulfilling the work task at hand, as indicated by the respondents in the questionnaire. The distribution shows the types marked by affiliation, and academic status, and the pool of all respondents. The types are arranged in accordance to their overall frequency.................................178

Table 7.5: Distribution of the level of granularity. The distribution shows the granularity as preferred by affiliation, and academic status, and the pool of all respondents, as well as the nine interviewees. The table is arranged in accordance to the levels of granularity........................................182

Table 7.6: Relation between respondents marking of type of broadcast and level of granularity. Types of broadcasts are arranged according to overall frequency. The types of broadcasts are arranged according to the overall frequency in marking by the respondents.............................................184

Table 7.7: Search entries preferred by the respondents. The table is arranged similar to the arrangement of search entries in the questionnaire, which is alphabetical according to the Danish terminology.........................192

Table 7.8: Comparison of search entries preferred by respondents according to affiliation, academic status, and the pool of all respondents. The search entries are listed in accordance to the overall marking by the 108 respondents....................................................................................202

Table 7.9: Categorisation of search entries for television broadcasts.................................................204

Table 7.10: Summarisation of the importance of search entries in relation to the four types of information needs.................................................................207

Table 7.11: Relevance criteria applied by the respondents. The table is arranged similar to the arrangement of the types of broadcasts in the questionnaire, which is alphabetical according to the Danish terminology.................................................................209

Table 7.12: Comparison of relevance criteria applied by respondents according to affiliation, academic status, and the pool of all respondents. The relevance criteria are listed in accordance to the overall marking by all 108 respondents....................................................................................216

Table 7.13: Categorisation of relevance criteria for television broadcasts.................................................217

Table 7.14: Summarisation of the importance of relevance criteria in relation to the four types of information needs.................................................................219

Table 8.1: Summary of significant access points for retrieval of television broadcasts. Listed according to category.................................................................240

Table 8.2: The function of the appropriate access points for television broadcasts.................................................................241
<table>
<thead>
<tr>
<th>Table 9.1:</th>
<th>Extraction of descriptions of DR1 television schedule from week 16, 2007. For the full schedule, please see Appendix 5.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 9.2:</td>
<td>Presence of identified access points in existing television surrogate records.</td>
</tr>
<tr>
<td>Table 9.3:</td>
<td>Value of information in the television schedule for construction of broadcast access points.</td>
</tr>
</tbody>
</table>
1 Introduction

This doctoral thesis is about facilitating effective storage and retrieval of television broadcasts. More specifically, the thesis reports on the study of how scholars and students seek and use broadcasts, and how this affects the construction of searchable records to ensure effective retrieval of television broadcasts. Recent technological developments have made it possible to provide users with electronic access to a vast amount of broadcasts (e.g., BBC News, 2003; Bjørkeng, 2006; British Film Institute, 2006; Frostholm, 2006; Munck, 2006). However, Ingwersen and Pejtersen stress that “[i]f we wish to reach the stage where capabilities of the present technological development are used to design information systems which support people’s problem-solving efforts, we must improve our knowledge of how man seeks and processes information” (Ingwersen & Pejtersen, 1986, p. 111). Despite the fact that Ingwersen and Pejtersen neither had the dissemination of television broadcasts in mind, nor knew the capabilities of our present technologies, we do believe the statement is of great relevance when dealing with electronic access to television broadcasts.

There are several ways to investigate how users seek and process information. In the present work, we investigate users’ information seeking behaviour by focusing on three aspects of users’ behaviour: 1) their information needs; 2) their preferred search entries; and 3) their application of relevance criteria. Thus we follow the recommendation by Wilson, who points out that in order to “…uncover the determining factors of [information seeking] behaviour we must do so by first undertaking in-depth studies of well-defined categories of persons, developing explanatory concepts and then testing these concepts in related but different settings” (Wilson, 1981, p. 11). Our work takes this approach and the well-defined category of persons investigated is Danish scholars and students within the academic field of Media Studies. The in-depth studies give indication as of the cognitive processes of users when seeking television broadcasts. The work concentrates on a well-defined category of persons, and does not attempt to test the derived knowledge in related but different settings, as suggested by Wilson. Instead, the thesis focuses on the implications for design and construction of an effective information retrieval system (IR system) for television broadcasts. In the present work, we do not focus on information seeking behaviour per see, but rather on the three aspects of users’ information seeking behaviour just mentioned. For instance, since no IR system is available for the collection under scrutiny in the present work, we do not investigate users’ interactivity with a search system.
Knowledge about the three aspects of users’ information seeking behaviour is important because it informs how the present technological developments can be used to design an effective IR system. We apply such knowledge to suggest which access points\(^1\) are appropriate in surrogate records for television broadcasts. Further, the present work investigates to what extent external textual and other sources (e.g., television schedules) contain information valuable for construction of surrogate records for future television broadcast retrieval systems.

Motivated by a need for better access to television broadcasts the intention of the present work is to contribute to our knowledge about providing effective storage and retrieval of television broadcasts. In this way, the thesis contributes to the continuing development and refinement of insight about users’ information seeking behaviour. More specifically, the thesis enriches our knowledge in the following areas of research:

1) **Aspects of users’ information seeking behaviour in relation to television broadcasts;**
   a. Users’ information needs;
   b. Users’ preferred search entries;
   c. Users’ application of relevance criteria;

2) **Appropriate access points for indexing and retrieval of television broadcasts; and**

3) **Construction of surrogate records for television broadcasts.**

We do not attempt the complete design and construction of an IR system for television broadcasts, nor do we provide a review or an evaluation of automatic approaches for indexing of television broadcasts. Instead, the empirically identified search entry and relevance criteria preferences are discussed as design implications for surrogate records for a future television broadcast retrieval system.

The assumption behind the present work is that in order to create an IR system that assist users with effective retrieval of television broadcasts, it is important to improve our knowledge about users’ information needs, search entry preferences, and application of relevance criteria. This assumption derives from our theoretical background in the cognitive viewpoint as expressed by Ingwersen and Järvelin (2005), which is introduced in the succeeding section.

\(^1\) We define access point as any data elements in a surrogate record that enables a user to search, and evaluate whether or not a specific information object is relevant to his or her task at hand.
1.1 Background

In this section, four elements of background for the present work are introduced. These four elements are:

- The Danish national collection of television broadcasts and its limitations;
- The cognitive viewpoint, which constitutes the epistemological basis for our work;
- Information seeking behaviour, which constitutes the first of the three research areas of the thesis; and
- Indexing, which is the second of the three research areas and of importance for the third research area concerning construction of surrogate records.

Hence, the following four sub-sections cursorily introduce the collection and its conditions, the cognitive viewpoint in Library and Information Science (LIS), previous research in information seeking, and previous research on indexing, respectively. The Danish national collection of television broadcasts is described in more detail in Chapter 5. The cognitive viewpoint, users’ information seeking behaviour, and indexing are dealt with further in Chapters 2, 3, and 4.

1.1.1 Background – the Danish national collection of television broadcasts

The research project is carried out in the context of the Danish national collection of television broadcasts. The collection is located at the State Media Archive at the State and University Library of Aarhus, Denmark. The State Media Archive comprises several collections of audiovisual character (e.g., sound recordings from the Danish Parliament and Danish video productions), but we are concerned with the collection of television broadcasts, only. The State and University Library collects, preserves, and gives access to the Danish cultural heritage as it is expressed in television broadcasts. The collection was established in 1987 (Ministeriet for Kulturelle Anliggender, 1987), and it contains more than 350,000 hours of broadcasts from Danish television stations (e.g., see Appendix 1). The State and University Library collects all transmitted broadcasts (e.g., news, documentary, commercials, spots), and the collection is accordingly heterogeneous in nature.

The access to the Danish national collection of television broadcasts is restricted in concordance with the copyright legislation, and it has been revised on several occasions (Fønss-Jørgensen, 2002). At present a non-restricted access is granted to academic users with documented research activities, whereas the broader public can
access the collection at the premises of the State and University Library, only (Kirring, 2006; Statsbiblioteket, 2007). However, broadcasts that have been published independently (e.g., a motion picture) are neither available for researchers nor for the broader public.

Until January 1st 2006 all broadcasts were received and stored analogically on VHS-tapes. The indexing of television broadcasts is traditionally very labour intensive, because it involves manual processing of each individual television broadcast (Brunelli, Mich & Modena, 1999, p. 79; Smeaton, 2004, p. 380). The inevitable need of manual resources has so far retained the State and University Library from indexing the collection of television broadcasts. Therefore, searchable records representing the broadcasts do not exist for the Danish national collection of television broadcast. Consequently, neither librarians nor users are able to search the collection in a systematic and electronically manner. Retrieval of television broadcasts is entirely based upon the shelving arrangements of the analogous VHS-tapes. The VHS-tapes are arranged chronologically according to date of broadcasting for each television channel. For retrieval to be feasible substantial knowledge on formal data of the requested television broadcast is needed. In addition, retrieval of broadcasts is facilitated at the granularity level of programmes, only. That is, whatever the level of granularity a user’s need addresses (e.g., a series of television programmes, a specific television programme, or a feature from a television programme), retrieval must be made at the level of television programmes as a whole. Users are, at present, required to provide the librarian with the formal data of the requested television programme, including the name of the transmitting channel, the title of the broadcasted programme, and the exact date and time of transmission, as well as additional administrative information (e.g., see the requisition form in Appendix 2).

1.1.2 Background – the cognitive viewpoint

The theoretical foundation for the present work is the cognitive viewpoint to LIS. The viewpoint represents a holistic approach to information seeking and IR (Ingwersen & Järvelin, 2005, p. 24). The viewpoint puts emphasis on the users, and accordingly the personal and potentially dynamic nature of the users’ information need, as expressed by Belkin and colleagues in the anomalous state of knowledge (ASK) hypothesis (Belkin, 1980; Belkin, Oddy & Brooks, 1982). Along with this emphasis on users, the holistic viewpoint recognises the importance of IR applications. Ingwersen points out that “[…] the cognitive viewpoint suggests that we investigate the variety of individual world models and knowledge structures that underly the surface structures of the variables of interaction with one another […]”. Further, it proposes the study of their quality and
Chapter 1: Introduction

limitations, in order to produce IR theories and applications that may optimize IR interaction, make us understand the nature of individual users’ actual desire for information and support its fulfilment” (Ingwersen, 1992, p. 18). The assumption of the cognitive viewpoint is not to bring the different world models and knowledge structures into harmony with each other, but rather to accept, understand, and exploit the differences between such models and structures (Ingwersen & Järvelin, 2005, p. 30). Insight as of the nature and characteristics of models and structures is a prerequisite for this exploitation. The present work is an attempt to create insight in order to make effective storage and retrieval of television broadcasts possible.

1.1.3 Background – information seeking behaviour

In continuation of the cognitive viewpoint’s emphasis on the individual actors, an increased focus on users’ information seeking behaviour is found in the research literature. Examples of this focus is given by the extensive work of Pejtersen (1980; 1989); Kuhlthau (1991; 1993); Byström and Järvelin (1995; Byström, 1999); and Vakkari (1999).

In line with the cognitive viewpoint Pejtersen studies the users’ cognitive perceptions as a prerequisite for construction of representations in surrogate records, in the ‘Book House’ project. With the objective of designing an effective IR system, Pejtersen is concerned with the understanding of the users and their search for information. The purpose of the ‘Book House’ project is to provide users with better access to fiction literature. Pejtersen analyses genuine information needs and on this basis she develops a classification scheme for fiction literature, indexes a test collection of fiction by use of the classification scheme, develops the icon based ‘Book House’ IR system, which is then evaluated. The evaluation of the ‘Book House’ system shows that the users approve of the system, and that the developed classification scheme improves users’ retrieval satisfaction of fiction literature (Pejtersen, 1980; 1989).

Kuhlthau focuses on the task underlying the users’ information needs and she derives a model of the different stages in users’ information search process. These stages are dependent on how clearly the user understands the underlying work task (Kuhlthau, 1991; 1993). Byström and Järvelin further investigate the users’ perceptions of the complexity of the underlying work tasks (Byström & Järvelin, 1995; Byström, 1999). This is furthermore elaborated on by Vakkari who associate task performance with information actions: need and use of information types; search strategies; and relevance assessments (Vakkari, 1999). We build on this line of work on users’ information seeking behaviour. That is, the thesis gathers knowledge about characteristics of users’ information needs, and search entry and relevance criteria
preferences in the context of a specific type of information objects, namely television broadcasts. In line with the work by Kuhlthau, Byström and Järvelin, and Vakkari, the underlying work task is seen in relation to the users’ information needs. In addition, similar to Pejtersen, we aim at identifying implications for future design and construction of broadcast retrieval systems. This we do by relating the knowledge learned in our empirical investigation of user needs, search entry preferences and relevance application to the information available in external textual sources, e.g., in the form of television schedules.

The above-mentioned cognitive viewpoint puts emphasis on the user and the notion of the user’s individual and potential dynamic information need. The notion of relevance is understood in connection to the personal and potential dynamic information need of the user. The relevance of an information object\(^2\) should be judged in accordance with the individual and potentially dynamic information need as opposed to the request or the query, only (Borlund, 2003a, p. 922). Hereby, the concept of relevance is recognised as a dynamic and multidimensional concept. For instance defined by Schamber, Eisenberg and Nilan (1990) and Borlund (2003a) as situational relevance. When evaluating the relevancy of information objects, users apply several criteria. Relevance criteria have mostly been studied with respect to textual information objects (e.g., Schamber, 1994). An increased focus is given to relevance criteria in respect to non-textual information objects (e.g., Barry & Schamber, 1998; Choi & Rasmussen, 2002). However, only limited research exist with specific emphasis on moving images (e.g., Yang, 2005). In a comparative analysis of studies of relevance criteria, Barry and Schamber concluded that the research provides “[…] evidence for the existence of a finite range of criteria that are applied across types of users, information problem situations, and information sources” (Barry & Schamber, 1998, p. 234). This said, Barry and Schamber also conclude that different types of information objects, users, and situational contexts, reveal different relevance criteria, and that further research is needed which involves other types of users and use environments. In the present work, we explore and expand our knowledge on relevance criteria in relation to scholars and students in Media Studies. We do this by investigating the metadata elements users apply when assessing the relevancy of television broadcasts in relation to their information need and underlying work task.

\(^2\) We define information objects similarly to Schneider’s (2004, p. 1) definition of documents. The term information objects include all types of documents in a semiotic sense (i.e., messages represented by organized sets of symbols) no matter their form (text, image, sound, moving images, etc.). Hereby, information is treated as ‘information-as-thing’ in Buckland’s (1991) terminology.
Knowledge about users’ cognitive processes, including their information needs and relevance assessments, is important in order to design effective indexing of information objects, and subsequently provide effective retrieval.

1.1.4 Background – indexing

Basically, two approaches can be applied for indexing of textual information objects: 1) manual intellectual analysis; and 2) automatic algorithmic analysis (Anderson, 1985; Anderson & Pérez-Carballo, 2001b; 2001a). Similarly, two approaches are applied for indexing of (moving) images (e.g., television broadcasts): 1) a manual or concept-based approach, based on intellectual analysis of (audio)visual data; and 2) an automatic or content-based approach, based on computational processing of (audio)visual data (Cawkell, 1992; Rasmussen, 1997; Jørgensen, 1999; 2001; Smeaton, 2004; Persson, 2005).

The manual approach depends heavily on human intellectual efforts, as “[...] images and the objects therein are manually identified and described in terms of what they are and what they represent” (Rasmussen, 1997, p. 170). In the manual approach, textual indexing terms are assigned to surrogate records. The manual approach is the traditional approach for indexing of broadcasts in operational archives at television stations (e.g., Green & Klasén, 1993).

The automatic approach is primarily based on computational production of numerical representations of attributes (Rasmussen, 1997, p. 178). The prevailing portion of the automatically indexed moving image IR systems are experimental systems (Hollink et al., 2004). Automatic approaches can also incorporate information derived from external sources. This technique stems from research which focuses on automatic indexing of images on the Internet (e.g., Dunlop & van Rijsbergen, 1993; Harmandas, Sanderson & Dunlop, 1997). Knowledge about indexing is relevant since we touch upon the applicability of information in external sources for construction of surrogate records for television broadcasts.

1.2 Objectives of the thesis

The overall objective of the thesis is to explore and extend our knowledge in order to assist users in effective retrieval of television broadcasts. We present research in order to aid in the establishment of retrieval facilities for television broadcasts. The State and University Library is interested in remedying the present unsatisfying situation of limited access to the Danish national collection of television broadcasts. Hence, the State and University Library is interested in answering the following two questions:
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

- How can a user be supported in retrieval of television broadcasts?; and
- How can this support be created with low utilization of resources?

However, the present work is not capable of giving full answers to these questions. Instead, the thesis attempts to gather explorative knowledge to assist the State and University Library to remedy the present unsatisfying situation for retrieval of television broadcasts. As a mean to reach this aim, the present work contributes with research in relation to the research areas presented above:

1) Aspects of users’ information seeking behaviour in relation to television broadcasts (including information needs, search entries, and relevance criteria);
2) Appropriate access points for indexing and retrieval of television broadcasts; and
3) Construction of surrogate records for television broadcasts.

In LIS, researchers have long recognised the value of gathering knowledge about users’ and their information seeking behaviour (e.g., Dervin, 1983b; Ellis, 1987; Kuhlthau, 1991). The research, however, is mainly concerned with searching for and use of textual information object, and only a small amount of research is undertaken to add to our knowledge about users’ behaviour when seeking non-textual information objects (e.g., Schamber, 1991b; Jörgensen, 1998). The attention towards users’ behaviour when seeking moving images is increasing in line with technological developments for dissemination of moving images (e.g., Yang, 2005; Lee et al., 2006). Nevertheless, the information behaviour of users seeking television broadcasts is still a neglected area of research. The present work aims at contributing research in this area. That is, the main part of the thesis is concerned with investigation of the nature and characteristics of users’ information needs for television broadcasts, their preferred search entries, as well as the relevance criteria they apply when evaluating television broadcast. Hereby, we attempt to expand our knowledge about aspects of users’ information seeking behaviour in relation to television broadcasts, the specific group of users in question (scholars and students in Media Studies), as well as in general.

The present work pursues the connection of the identified behaviour and preferences of the test participants with implications for design and construction of IR systems for television broadcasts. To be more precise, the work aims at identifying appropriate access points in surrogate records for television broadcasts. Appropriate
access points are those data elements in surrogate records that ensure optimized retrieval to support the fulfilment of users’ genuine need for information. Identification of appropriate access points is based on knowledge about users’ behaviour when seeking television broadcasts. In this way, the thesis aims at generating knowledge for the construction of surrogate records for television broadcasts, which is important for the future users of systems for retrieval of television broadcasts, as well as systems for retrieval of comparable types of information objects, e.g., radio broadcasts.

Finally, the aim of the thesis is to investigate whether external contextual sources (e.g., television schedules) contain information that is applicable in the construction of surrogate records. This knowledge is important as a prerequisite for indexing in relation to the construction of a future broadcast retrieval system, though considerations regarding the specific indexing in relation to such a future IR system is outside the scope of the present work.

1.3 Research questions

The overall research aim of the doctoral work is to gather knowledge about the appropriateness of access points in relation to a future broadcast retrieval system. The present work pursues this overall research aim by answering five main research questions:

1. What characterises a given group of users’ information needs in the context of television broadcasts?
2. Which search entries are preferred for searching of television broadcasts, and why and when are they preferred?
3. Which relevance criteria are applied when evaluating television broadcasts, and why and when are they applied?
4. Which access points in a surrogate record are appropriate for television broadcasts, and what are their functions? and
5. To what extent do external textual and other sources contain or provide information valuable for constructing the identified access points?

In line with the recommendations put forward by Wilson (1981, p. 11) the research questions focus on a given group of users. The group of users is Danish scholars and students in Media Studies. The first research question will scrutinize the nature, and characteristics of the users’ information needs in relation to television broadcasts. To some extend we focus on the users’ underlying work tasks to operationalize the
somewhat intangible concept of information need. Information about users’ information needs is important because it gives information of the users’ cognitive perceptions of television broadcasts, as well as an indication of the functionalities that an effective future IR system for television broadcasts should support. The second research question focuses more specifically on search entries for television broadcasts. It gives clarification as to which search entries users find important when they search for television broadcasts. This knowledge is significant since it informs about the search facilities that should be available in an effective IR system for television broadcasts. The third research question puts emphasis on the relevance criteria users apply when seeking television broadcasts. In the present work, we consider relevance criteria to be the metadata elements users apply when they assess the relevancy of television broadcasts. The research question verifies which metadata elements users apply when evaluating whether or not a television broadcast is relevant. This verification is essential because it provides the information users require for evaluating whether or not a television broadcast is relevant to pursue, in order to fulfil their information need at hand. Together, the first three research questions investigate and verify aspects of the users’ behaviour when seeking television broadcasts. More specifically, they identify which dimensions of television broadcasts, the users’ seeking behaviour relates to. The answering of the three first research questions contributes to our knowledge in the first research area concerning aspects of users’ information seeking behaviour in relation to television broadcasts.

The results from these three research questions are essential for recognising the dimensions of television broadcasts that ought to be described in surrogate records. This is investigated in the fourth research question. More specifically, the fourth research question is concerned with the correspondence between the identified dimensions of television broadcasts and access points in surrogate records. In this way, the fourth research question contributes to our knowledge in the second research area about appropriate access points for indexing and retrieval of television broadcasts. This knowledge, again, is related to investigation of the fifth research question. The fifth research question gathers knowledge about the correspondence between information in external and other sources and the information needed to construct the identified access points. The fifth research question is related to the third research area on construction of surrogate records for television broadcasts.

The five research questions are strongly interconnected, and rooted in the assumption that for an IR system to be effective it must be based on insight into aspects of the users’ behaviour when seeking information. More specifically, the research questions are connected as knowledge about users’ information needs, search entry
preferences, and application of relevance criteria is essential for identification of appropriate access points. Identification of appropriate access points is again essential for investigating whether external or other sources contain information valuable for indexing of television broadcasts. Together the five research questions give clarification with reference to the overall research aim concerning the appropriateness of access points in relation to a future broadcast retrieval system. The thesis does not aim at constructing an operating IR system for television broadcasts, nor to investigate specific algorithms for extraction of information from external sources. Rather, the present work intends to gather insight that is a prerequisite for such future research e.g., concerned with the automatic extraction of information from external sources for indexing of television broadcasts based on knowledge about aspects of scholars’ and students’ information seeking behaviour.

1.4 Structure of the thesis

The structure of the doctoral thesis falls in two main parts, a theoretical and an empirical part. The theoretical part comprises four chapters (Chapters (1), 2, 3, 4), and together they provide the theoretical foundation for the work presented in the thesis. The empirical part comprises five chapters (Chapters 5, 6, 7, 0, and 9), and together they present the empirical study undergone in this doctoral work. Chapter 10 comprises summary and conclusions. Each chapter is represented by a short indicative description below.

Chapter 2 provides an introduction to the cognitive viewpoint in LIS. The theoretical foundation of our work is the cognitive viewpoint as it is expressed by Ingwersen and Järvelin (2005). The chapter discussed five central and interrelated features of the viewpoint, namely that: 1) information processing takes place in senders and recipients; 2) processing takes place at different levels; 3) any communicative actor is influenced by its experiences and contextual environment; 4) individual actors influence their contextual environment; and 5) information is situational and contextual. Further, the chapter outlines Ingwersen and Järvelin’s holistic framework for an integrated cognitive approach for information seeking and retrieval and places our work within the cognitive viewpoint.

Chapter 3 focuses on information seeking behaviour. The chapter begins by discussing the notion of a personal and potentially dynamic information need, which is in line with the cognitive viewpoint. We propose an expansion of the concept of information need, and the chapter explains the theoretical basis for this proposal. The chapter further discusses general models of information seeking behaviour, including
models that emphasise the importance of the underlying work task situation. Given that, the empirical part of the thesis focuses on scholars and students in Media Studies and aspects of their information seeking behaviour in the context of television broadcasts, the chapter reviews and discusses previous work in both of these areas. More precisely, the chapter discusses literature on the information needs of academics, and mainly academics in the humanities, and the chapter outlines and reviews literature on users information needs in the context of (moving) images. Further, due to our focus on relevance criteria in research question three, Chapter 3 briefly outlines our conception of relevance as a multidimensional, situational, and potentially dynamic concept. The review has an emphasis on previous work that deals with relevance criteria in the context of (moving) images. In this way, Chapter 3 addresses the first research area on aspects of users’ information seeking behaviour in relation to television broadcasts, including users’ information needs, preferred search entries, and application of relevance criteria.

Chapter 4 discusses different approaches for indexing of television broadcasts. The chapter opens by outlining and discussing central aspects of the process of indexing in a traditional textual context, including the concept of aboutness. Afterwards the chapter pays attention to indexing in the context of (moving) images and the chapter concerns both the manual intellectual approach for indexing, and the technical automatic approach for indexing. Finally, the chapter focuses on the application of external sources for indexing of primary information objects. The chapter discusses previous work from a variety of areas, which have applied external sources for describing different types of information objects.

The first four chapters explain the theoretical foundation for the empirical studies presented in this thesis. Chapters 5 to 9 explicate the empirical part of the thesis. Chapter 5 serves two functions: 1) it illustrates the context of the empirical study; and 2) it describes the selection of target group for these empirical studies. In the first part, we present the context of the empirical study by explaining the current situation for gaining access to television broadcasts in Denmark. This presentation has special emphasis on the Danish national collection of television broadcasts at the State and University Library. The selection of user target group is mainly based on knowledge about the present genuine requests for television broadcasts put forward to the State and University Library. The chapter describes these genuine requests, and we identify scholars and students in Media Studies as the user target group. Finally, the chapter gives an introduction to the academic field of Media Studies.

Chapter 6 outlines the methodical approach applied to answer our five research questions. Grounded theory (Glaser & Strauss, 1967; Strauss & Corbin, 1998) is
presented as the methodological basis. The chapter illustrates the concordance between the underlying methodological approach and the epistemological foundation in the cognitive viewpoint described in Chapter 2. A web questionnaire and in-depth interviews are the main methods for collection of empirical data. Chapter 6 clarifies the reasons for applying these methods, and explains thoroughly the application of the methods in the present work. Further, the chapter explains the methods applied for analysis of the collected data.

The Chapters 7, 8, and 9 report on the results of the empirical investigation, concerning each of the three research areas, respectively. Chapter 7 provides answers to the three research questions concerning aspects of the test participants’ information seeking behaviour, namely: 1) what characterises a given group of users information needs in the context of television broadcasts?; 2) which search entries are preferred for searching of television broadcasts, and why and when are they preferred?; and 3) which relevance criteria are applied when evaluating television broadcasts, and why and when are they applied? The chapter shows that the test participants’ needs for television broadcasts are for known broadcasts, factual data about the broadcasts, or of a topical nature, whether the test participants’ perception of the topic is clear or muddled. The broadcasts needed are mainly at the programmes level of granularity, but the test participants’ also express a need to access broadcasts at the serials and features level of granularity. The chapter identifies 28 search entries in the television broadcast retrieval context. Similarly, 32 relevance criteria are identified to be of importance for broadcast retrieval.

Chapter 8 focuses on the second research area concerned with appropriate access points for indexing and retrieval of television broadcasts, and the chapter brings forth answers to the fourth research question, stating: which access points in a surrogate record are appropriate for television broadcasts, and what are their functions? The search entries and relevance criteria identified in Chapter 7 are translated to access points, and the value of each access point is considered. Based on this consideration the chapter identifies 24 access points, which we regard as appropriate in a future broadcast retrieval context.

Chapter 9 focuses on the third research area concerned with construction of surrogate records for television broadcasts, and the chapter answers the fifth and final research question stating: to what extent do external textual and other sources contain or provide information valuable for constructing the identified access points? The chapter’s focal point is television schedules, and the chapter shows that information available in television schedules is mainly valuable for construction of access points associated with descriptive indexing. Further, the chapter discusses how other
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

approaches for construction of access point might be applied as a complement to the information available in television schedules. This includes approaches for automatic indexing as well as how an exchange of surrogate records with existing and operating archives and indexes can enhance the quality of indexing of television broadcasts.

Chapter 10 summarises the contributions of the thesis and provides recommendations for future work. Lastly, Chapter 10 presents our final conclusion.
Chapter 2: A cognitive framework for information seeking and information retrieval

2 A cognitive framework for information seeking and information retrieval

The cognitive viewpoint constitutes the epistemological foundation of the present work. The intention of this chapter is to introduce the cognitive viewpoint to research in LIS, and to position our work within this viewpoint. In short, the viewpoint states that one must consider the cognitive changes of all participants in an IR interaction (Brookes, 1980; Ingwersen, 1992). Due to possible learning and cognition during IR interaction, an information need should not be treated as static. Instead, it should be regarded as being potentially dynamic, which means allowing for it to change and mature over time during IR interaction. This further implies that relevance assessments should not be treated in isolation, but in relation to the potential dynamic information need. Hence, the relevancy of information objects to a user’s information need is similarly of a potential dynamic nature (Borlund, 2000; Ingwersen & Järvelin, 2005, p. 3). The cognitive viewpoint is not particular to LIS, but a so-called cognitive revolution emerges concurrently in the 1970s in several scientific disciplines (e.g., De Mey et al., 1977; Miller, 2003). Our focus is on the cognitive viewpoint in relation to LIS, only.

The cognitive viewpoint has been discussed in, and applied to several researchers’ work in the past three decades, which have helped to shape, develop, and mature the approach. We regard Ingwersen and Järvelin’s (2005) recent proposal of an integrated cognitive framework for information seeking and IR, as the latest ripening of a cognitive approach to LIS. Accordingly, our understanding of the viewpoint corresponds to the conception expressed in this framework, and the framework put forward by Ingwersen and Järvelin is the focal point of this chapter. Our review of the cognitive viewpoint is obviously a great simplification, which overlooks several significant results. However, the aim of our review is merely to introduce the viewpoint as Ingwersen and Järvelin manifest it. For more thorough discussions of the origin and development of the cognitive viewpoint, we refer to the extensive works provided by Ingwersen (1992; 1996) and Borlund (2000).

The chapter consists of four main sections. In the 1st section, we cursorily discuss the interrelatedness of information seeking and IR. This is relevant since integration of the two is the main goal of the cognitive framework proposed by Ingwersen and Järvelin, as well as the focal point of our work. The 2nd section briefly discusses the cognitive foundation of the integrated framework. In the 3rd section, we outline the framework, and discuss the main model of the framework. The 4th section places our
work within the cognitive viewpoint, and the chapter ends with a short summary in the 5th section.

2.1 Information seeking and information retrieval

The motivation behind the cognitive framework proposed by Ingwersen and Järvelin is the integration of information seeking and IR research. The purpose of the integration is to strengthen LIS research by considering three interconnected research aspects, namely:

a) Generation of further theoretical understanding;

b) Description and explanation of empirical evidence; \textit{and}

c) Supporting the development of technology.

(Ingwersen & Järvelin, 2005, p. 3).

The generation of new theoretical understanding is important in order to further broaden our knowledge concerning our empirical results. However, the empirical evidence is also important as a mean to broaden our theoretical understanding. Further, on one hand the development of technology is blind if it does not consider the theoretical and empirical results (Ingwersen & Järvelin, 2005, p. 3). On the other hand, development of technology is important in order to empirically investigate new theories. A holistic cognitive approach to LIS research is found in these three interrelated aspects and the integration of information seeking and IR.

Information seeking and IR are both concerned with information behaviour activities, but from different perspectives. Information seeking is concerned with a user’s or group(s) of users’ purposive seeking for information, in order to satisfy an underlying goal (Wilson, 2000, p. 49). IR, more narrowly, concerns the processes involved in representation, storage, searching, finding, filtering, and presentation of information. That is, information which is assumed to be relevant to a user’s requirement for information (Ingwersen & Järvelin, 2005, p. 21).

Ingwersen and Järvelin’s integration of information seeking and IR is a further elaboration of Wilson’s nested model of the information seeking and IR research areas (Wilson, 1999b, p. 263). Ingwersen and Järvelin’s modification of Wilson’s original model is depicted in Figure 2.1. The model stresses the interrelatedness between information seeking and IR, by emphasising that every incident of IR occurs in the context of an information seeking process. This goes whether the IR process is of an interactive or algorithmic nature. Further, information seeking is encompassed within a
broader information behaviour activity (Ingwersen & Järvelin, 2005, p. 198). Figure 2.1 emphasises that a user’s perception of an underlying work task is what instigates information behaviour. In this way, a perceived work task is the driving force for any information seeking or (interactive) IR activity, whether the task is a natural work task, an assigned work task, or a simulated work task. The term work task is not restricted to denote tasks originating from job contexts alone, but it also refers to tasks originating from leisure or non-job related contexts. In contrast, the term search tasks is applied to refer to the instrumental seeking or retrieval activities that are triggered by the underlying work task (see e.g., Section 3.2.5 on task based information seeking).

Research in information seeking and IR share the same ultimate goal of facilitating effective human information behaviour. However, despite earlier attempts (e.g., Saracevic, 1997; Spink, 1997), research from the two areas have had a hard time being integrated. Ingwersen and Järvelin (2005, p. 2) ascribe these difficulties to the fact that information seeking, in general terms, arises from a Social Science approach to human behaviour, while IR originates from a Computer Science focus on technological developments. Consequently, the results from one community are not considered in the other community. That is, information seeking researchers view IR research as too narrowly bound to technological achievements, while IR researchers view information seeking research as an unnecessary academic exercise (Ingwersen & Järvelin, 2005, p. 2). However, integration of results from the two research areas will broaden our
knowledge about human information behaviour, and hereby further illuminate the significance of LIS research. This is related to the coherency and interconnectedness of the three aspects listed above. As an example, clarification of empirical evidence is important for theoretical understanding and/or development of new technologies. Ingwersen and Järvelin (2005, p. 3) maintain that a focus on one of the three research aspects in isolation makes the research blind.

The integrated framework is rooted in a cognitive approach to LIS. Ingwersen and Järvelin list five central and interrelated features of the cognitive viewpoint, which are of particular importance for their integrated information seeking and IR framework:

“1. Information processing takes place in senders and recipients of messages;
2. Processing takes place at different levels;
3. During communication of information any actor is influenced by its past and present experiences (time) and its social, organizational and cultural environment;
4. Individual actors influence the environment or domain;
5. Information is situational and contextual”

In the succeeding section, we outline the cognitive viewpoint. We first cursorily give a brief summary of the foundation of the viewpoint, including its origin as an alternative to the traditional system-driven approach to LIS. Then we discuss each of the five central features in five sub-sections e.g., by relating central developments in the cognitive viewpoint to these features. The aim of our discussion is to introduce the cognitive foundation of Ingwersen and Järvelin’s integrated framework, prior to our further outline of the framework in Section 2.3.

2.2 The cognitive foundation of the integrated framework

The relatively short history of LIS is characterised by a wish to discuss the discipline’s research focus and approach, and the cognitive viewpoint is proposed as an alternative to the mainstream system-driven approach to research in LIS (Ingwersen, 1992, pp. 1-14; Borlund, 2000, p. 11). The cognitive viewpoint grew out of a user-oriented approach to LIS, and it is considered a further development and maturation of a focus upon users in LIS. In the following, we focus on the system-oriented approach because it has been prominent in implementation and evaluation of IR systems. Both the system-driven and the cognitive approach are concerned with access to information
objects and investigation of retrieval effectiveness. The system-driven approach is concerned with algorithmic matching of document representations and information need representations. Document representations are constructed by automatically processing specific document features, e.g., words in a textual document. Information needs are represented by static queries, often denoted topics. Objective relevance assessments are made by assessors, and the assessments are objective indications of the topical relationship between a document representation and a query (Borlund, 2008). The system-driven approach adheres to the so-called Cranfield model, which aims at conducting benchmark evaluation of IR system performance without interference from intangible operational variables (Cleverdon, Mills & Keen, 1966, p. 8). Static information needs and objective relevance judgments are elements of the Cranfield model, and the model utilises the performance measures of recall and precision (e.g., Cleverdon & Keen, 1966; Cleverdon, Mills & Keen, 1966). We agree with Järvelin (2007, p. 973), that the system-driven approach abstracts a mechanical component out of human-information communication, and makes it the isolated object of inquiry. The removal of individual subjectivity promotes a homogeneous quantitative treatment of IR systems, to the detriment of being able to handle problems raised by human behaviour. In brief, the system-driven approach disregards human cognitive behaviour, while the behaviour of human cognitive actors is the central element in the cognitive viewpoint.

The origin of the cognitive viewpoint in LIS cannot be ascribed any one specific person or any one specific event, and it is not possible to specify any general agreed upon definition of the viewpoint (De Mey, 1982). According to De Mey (1982, p. 228), elements of the viewpoint can be dated back to Piaget’s pioneering work in the 1920s on identification of the development of cognitive structures in children, as a mean to understand knowledge construction. However, such cognitive structures cannot be generalised across populations without, for instance, considering the cultural and social context of the people involved. The first conference on the cognitive viewpoint in LIS is held in 1977 (De Mey et al., 1977). The multidisciplinary conference is critical in the development of the cognitive viewpoint, not least because it included contributions from prominent researchers such as B.C. Brookes, N.J. Belkin, M. De Mey, and P. Ingwersen (Belkin, 1977; Brookes, 1977; De Mey, 1977; Harbo, Ingwersen & Timmermann, 1977). We agree with Borlund (2000, p. 11), that the work by these four researchers is leading in developing, shaping, and maturing the cognitive viewpoint in LIS (e.g., Brookes, 1975a; 1980; De Mey, 1977; 1980; 1982; Belkin, 1978; 1980; Ingwersen, 1982; 1992; 1996). In addition, we find T.D. Wilson’s work on information seeking, central for the development of the cognitive viewpoint in LIS (e.g., Wilson,
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

1980; 1981; 1984; 1999b). Wilson’s work is important because he is the proponent for investigation of information seeking behaviour in workplace settings.

In the following five sub-sections (Sections 2.2.1 - 2.2.5) we discuss each of the five central features of the cognitive viewpoint, which we briefly introduced in the previous section. Our discussion is mainly focused on significant contributions by the five prominent researchers, mentioned above. The five central features are of an interrelated nature, and one researcher’s contribution may consequently be relevant to more than one of the five features.

2.2.1 Processing takes place in senders and recipients

In order to stimulate discussions on the scope of LIS, Brookes expresses his perception of the cognitive viewpoint in a formalised pseudo-mathematical equation of Information Science (Brookes, 1975a; 1975b; 1977; 1980). The 1980 version of the equation states that:

\[ K[S] + \Delta I = K[S + \Delta S] \]

Knowledge structures (\( K[S] \)) change into new modified structures (\( K[S + \Delta S] \)) by the addition of information (\( \Delta I \)). Brookes’ equation should be interpreted in the light of his understanding of human communication to be mediated by a ‘channel’, and that cognitive processing takes place at both ends of such a ‘channel’ (Brookes, 1977, p. 195). Hereby, the equation gains a dynamic element in which both senders and recipients are affected by the communicative process. Ingwersen (1992, pp. 32-33) later accentuates this dynamic element of the equation by adding the notion of potential information (\( pI \)). From the cognitive actor’s point of view, Ingwersen constructs the following equation:

\[ pI \rightarrow \Delta I + K[S] \rightarrow K[S + \Delta S] \rightarrow pI' \]

Ingwersen’s modification implies that a recipient receives potential information (\( pI \)), only. The information is considered potential because it may just be a part of it, which is perceived, or perceivable, by the recipient (\( \Delta I \)). The received information may for instance be incomplete and/or the recipient may be incapable or uninterested in perceiving all the received information. The outcome of the equation is also potential information in that “[t]he modified state of knowledge may generate, e.g., answer back or later create, new information (\( pI' \)), potential to other recipients” (Ingwersen, 1992,
By adding potential information ($pI'$) at the end of the equation, Ingwersen emphasizes that for any communicative activity in which information systems are involved, cognitive processes take place at both ends of the ‘channel’. In this way, the equation relates to the first central feature of the cognitive viewpoint stating that information processing occurs in senders and recipients of information.

Belkin (1990, p. 11) is explicitly inspired by Brookes’ work on the fundamental equation, and he agrees that humans build up internal structured representations of their world, and that meaningful communication changes these knowledge structures (Belkin, 1975, pp. 81-82). Belkin refers to the knowledge structures as states of knowledge and to him “[...] the essence of the cognitive viewpoint is that it explicitly considers that the states of knowledge, beliefs and so on of human beings (or information-processing systems) mediates (or interacts with) that which they receive/perceive or produce” (Belkin, 1990, pp. 11-12). He follows Brookes’ line of thought in that processing of information takes place in both senders and recipients of information. In addition, he stresses that mediation or processing is not limited to human beings, but that it also encompasses processing undertaken by computers. Hereby, he touches on the second central feature of the cognitive viewpoint concerned with different levels of processing, which is discussed in the succeeding section.

2.2.2 Processing takes place at different levels

In an explanation of the implications of his understanding of the cognitive viewpoint, De Mey extends Michie’s (1974, p. 113) three levels of interpretive capability for information processing devices. De Mey distinguishes between four stages of processing:

1. A monadic stage during which information units are handled separately and independently of each other as if they were simple self-contained entities;

2. A structural stage where the information is seen as a more complex entity consisting of several information units arranged in some specific way;

3. A contextual stage where in addition to an analysis of its structural organization of the information-bearing units, there is required information on context to disambiguate the meaning of the message; and

4. A cognitive or epistemic stage in which information is seen as supplementary or complementary to a conceptual system that represents the information-processing-system's knowledge of its world” (De Mey, 1980, p. 49).
The processing gradually becomes more and more complex as we move downwards through the stages. In the monadic stage, data is perceived in isolation, whereas data is perceived through the eyes of the beholder, in the cognitive stage. This entails subjective understanding, and hereby relates to the third central feature of the cognitive viewpoint, which is discussed in the following section. Human actors are capable of processing information at all four stages, whereas computers, due to their algorithmic predictiveness, can not reach the fourth stage, unless they receive direct support from human actors (Ingwersen & Järvelin, 2005, p. 26).

Ingwersen’s modification of Brookes’ fundamental equation of Information Science is also related to the second central feature on levels of processing. In the previous section, we discussed the equation from a cognitive actor’s point of view. From the system’s point of view, Ingwersen (1992, pp. 32-33) adds the notion of data \(D\), to the equation. The equation from a system’s point of view is hence:

\[
pI \rightarrow D + K(S) \rightarrow K(S) + \Delta S \rightarrow pI'
\]

Data \(D\) is applied instead of information \(\Delta I\) to denote that a system is incapable of processing the meaning of unrecognised data. The system is unable to process data at De Mey’s cognitive stage of processing. Consequently, the system’s state of knowledge \(K(S)\) remains unchanged. This version of the equation may also be applicable from a human recipient’s point of view, on the rather theoretical assumption that no information is perceived by the recipient, and hence that the recipient’s state of knowledge remains unchanged all through the communicative activity. Here, one should keep in mind that even a ‘no hit’ respond from an IR system alters a recipient’s cognitive structures. The modified equation follows Belkin’s thoughts underlying his model of the communication system of IR (Belkin, 1977, p. 190; 1978, p. 81; 1980, p. 135; Belkin, Oddy & Brooks, 1982, p. 65), in that processing occurs at two levels: a linguistic level and a cognitive level. The linguistic level incorporates the first three stages of De Mey’s four stages of information processing, namely the monadic stage, the structural stage, and the contextual stage, while the cognitive level is comparable to De Mey’s cognitive stage.

2.2.3 **Cognitive actors are influenced by their experiences and environments**

Brookes’ fundamental equation states that the addition of information changes the recipient’s knowledge structures \(K(S)\). However, the growth of information is not simply accretive. The increment of information \(\Delta I\) does not necessarily imply an
equivalent increment of the recipient’s knowledge structure \( (K[S]) \), but instead the increment denoted \( \Delta S \) (Brookes, 1980, p. 131). By using \( \Delta S \), and not \( \Delta I \), Brookes stresses that the same information \( (\Delta I) \) may cause different adjustments or modifications \( (\Delta S) \) of the knowledge structures \( (K[S]) \) for different cognitive actors, or at different points in time. For instance, it signifies that two users who get the same information (e.g., reading the same scientific article) may perceive the information differently, and that the information may cause different modifications or adjustments to their knowledge structures. Hence, the fundamental equation emphasise cognitive actors’ subjective perception of information. That is, the increment of knowledge structures are influenced by the recipient’s previously acquired knowledge, including his or her past and present experiences.

A similar potential dynamic change in knowledge structures is found in Belkin’s model of the communication system of IR. According to Belkin (e.g., 1980), the instigation of any IR activity is an anomalous state of knowledge known as the ASK hypothesis. The ASK is an elaboration of Wersig’s conception of the problematic situation as that which underlies and prompts a need for information (Wersig, 1971, pp. 68ff). The situation is considered problematic because the person recognises that his or her cognitive structures are inadequate for reaching a goal, and therefore information is needed to solve the problematic situation. That is, the person’s “[…] recognition of an inadequacy in her/his state of knowledge (an anomaly with respect to the problem)” (Belkin, Oddy & Brooks, 1982, p. 65). In order to remedy the inadequacy in his or her state of knowledge, the person may instigate communication with an IR system. However, each instance of communication may alter the person’s perception of what is needed to resolve the problematic situation. The person’s ASK, and hence his or her perception of the problematic situation, may consequently change, and communication with the IR system may be resumed, commenced anew, or be terminated if the person considers the problem solved. The person’s experiences and contextual environments constantly influence the cognitive structures, and therefore experiences and environments play an important role in a person’s recognition and perception of an anomalous situation.

In an elaboration of a principle proposed by Craik (1943, p. 61), De Mey explains that the central point of the cognitive viewpoint is that “[…] any processing of information, whether perceptual or symbolic, is mediated by a system of categories or concepts which, for the information-processing device is a model of his [or her] world” (De Mey, 1977, pp. XVI-XVII). This goes whether the processing device is a human being or a computer. The inclusion of world models implies that the cognitive viewpoint encompasses the social and organisational environment of the individual.
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

processing devices. The world model is an image of the world in the way it is perceived by the processing device, which is established by the processing device through its interpretation of, or experience within, its social and cultural context. De Mey’s notion of world model is equivalent to Brookes’ notion of knowledge structure, Belkin’s notion of a state of knowledge, and what we, in line with Ingwersen, denote cognitive structure. With reference to the cognitive level of information processing, discussed in the previous Section 2.2.3, the recipient processes the incoming data by applying his or her cognitive structures and hereby reaches an interpretation of the incoming data. The recipient’s own cognitive structures and the context of the received data combine into an interpretation of the received data. Hence, ‘meaningful’ data or information is constructed. This subjective interpretation is comparable to the alteration of knowledge structures in the fundamental equation of Information Science by Brookes. Thus, different actors interpret specific information differently. Such semantic openness may influence the uncertainty in relation to the meaning of data, but it may also have an effect on the potential value of the data for the recipient as well as the sender (Ingwersen & Järvelin, 2005, p. 26). Thereby relating it to the first central feature of the cognitive viewpoint discussed in Section 2.2.1.

Meaning and understanding are two prominent concepts in Wilson’s comprehension of the cognitive viewpoint (e.g., Wilson & Streatfield, 1977; Wilson, 1980). Wilson focuses explicitly on information seeking behaviour, and to him “[…] the essence of the cognitive approach is the idea of human perception, cognition, and structures of knowledge” (Wilson, 1984, p. 197). Though the conception could also be applied to computers, the focal point for Wilson is human beings. From Wilson’s point of view, the goal of an interaction between an individual’s world model and information is to produce meaning or understanding of the received information. This is comparable to the (modified) fundamental equation of Information Science, which states that new information may change the recipient’s current cognitive structures, and the dynamic nature of Belkin’s communicative system. The recipient tries to interpret the incoming information in relation to the knowledge he or she has already acquired, influenced by earlier experiences and the contextual environment.

In brief, the cognitive structure of any actor is under influence by its past and its social and cultural environment. The dimension of time hence plays an important role for interpretation and cognition. The cognitive viewpoint stresses that the influence is not of a deterministic character, but rather that it is based on individual actors’ subjective interpretations.
2.2.4 Cognitive actors influence their environment

In the previous section, we discussed how the individual actor is influenced by his or her experience and environmental context. This influence is of a bi-directional nature, in that the cognitive actor’s perception of the situation at hand may influence the contextual environment as well. To be exact, it is the actions (e.g., communicational) of the cognitive actor that are based on his or her perceptions, which may, over time, contribute to modifications of the contextual environment. Similar to the influence stemming from the environment, the influence from a cognitive actor is of a potential and non-deterministic character.

The mutual influence between individual cognitive actors and the environment may be compared to our discussion of the first central feature of the cognitive viewpoint stating that information processing takes place in senders and recipients of messages. For instance, Ingwersen’s modified equation of Information Science states that the result of receiving potential information may be the generation of new potential information. Potential information that can be perceived by one or more cognitive actors, and hence it may possibly generate modification in their cognitive structures and so on. The bi-directional influence may be most visible in cases where a cognitive actor is also a group member. The behaviour of the individual group member can influence the behaviour of other members of the group. In addition, the behaviour may affect other groups or other elements of the contextual environment. This is illustrated in Hyldegård’s (2006, pp. 94-95) modification of Ingwersen and Järvelin’s model of interactive information seeking, retrieval and behavioural processes (see the lower half of Figure 2.2 and our discussion of the model in Section 2.3). Hyldegård accentuates that depending on ones perspective, a group can be considered the closest contextual environment to the individual group members, or as a cognitive unit on its own with collective cognitive structures. Collective cognitive structures denote that individual cognitive actors have a shared perception of the situation at hand (Ingwersen & Järvelin, 2005, pp. 27-28). These collective cognitive perceptions may, similar to the individual cognitive structures, be influenced by the contextual environment, or by the participating cognitive actors’ past and present experiences. Ingwersen and Järvelin denote this mutual influence the principle of complementary social and cognitive influence (Ingwersen & Järvelin, 2005, pp. 30-31), which touches upon the fifth central feature stating that information is contextual and situational.
2.2.5 Information is situational and contextual

According to Ingwersen (1996, p. 13), Wilson is the leading proponent for settling with the “[…] inherent and silent agreement that information behaviour, seeking and retrieval mainly takes place among academics”. Drawing on Schultz (1970), Wilson (1984, p. 200) emphasises that the significant parts of an actor’s cognitive structures refers to the actor’s social heritage, or what we denote the contextual environment. In addition to the third central feature about influence from experiences and the environment, discussed in Section 2.2.3, Wilson emphasises that information is determined by situation and context. Information should be considered in relation to the recipient’s frame of reference or social heritage. Consequently, Wilson was prominent in taking information seeking investigations into the work place in order to empirically derive knowledge about people’s work tasks and their information seeking behaviour (e.g., Wilson & Streatfield, 1977; Wilson, 1980). Wilson’s work has been important in the development of practical as well as methodical aspects of research of information seeking behaviour from a cognitive point of view. If LIS is concerned with how and why people engage in information seeking behaviour, and of how they use information, then the cognitive viewpoint, as approached by Wilson, is concerned with understanding the development and structure of people’s cognitive structures of everyday life and to relate them to that life which is necessarily social (Belkin, 1990).

Ingwersen (1984; 1992; Ingwersen & Järvelin, 2005) ascribe the focus on the situational context to Luria’s (1976) investigations of social categorisation dating back to the 1920s. In this way, he follows the thoughts of Wilson, and adds to De Mey’s conception of the cognitive viewpoint, that cognitive structures are “[…] determined by the individual and social/collective experiences, education, training etc” (Ingwersen, 1982, p. 168). The current cognitive state is affected by the individual’s environmental context along with the situational context, and hence situation, context, and time is important for interpretation and cognition (Ingwersen & Järvelin, 2005, p. 27).

By discussing each of the five central features of the cognitive viewpoint, we have introduced the cognitive foundation of Ingwersen and Järvelin’s integrated framework for information seeking and IR. In the following section, we outline the framework further.

2.3 The integrated framework for information seeking and retrieval

In retrospection, the integrated framework seems a natural continuation of the development, maturation, and shaping of the cognitive viewpoint in LIS. Above, we
discuss the cognitive viewpoint’s emphasis on information as being contextual and situational. The situational and contextual importance is found in Ingwersen’s further development of Belkin’s model of the cognitive communication system of IR. Ingwersen and Järvelin (2005, p. VII) consider their integrated framework for information seeking and retrieval to be an explicit continuation of Ingwersen’s earlier work (e.g., Ingwersen, 1992). Ingwersen elaborates on the model several times (e.g., Ingwersen, 1982; 1992; 1996), and a modified version of this model is central in the integrated framework for information seeking and IR (Ingwersen & Järvelin, 2005, p. 274). The model is depicted in the lower half of Figure 2.2.

The focus of Figure 2.2 is the information seeker, which is illustrated with the information seeker’s central position in the social and physical level and the information seeker’s domination of the cognitive and emotional level. With reference to our discussion on levels of processing in Section 2.2.2, the social and physical level is analogous to Belkin’s linguistic level of processing, while the cognitive-emotional level is analogous to Belkin’s cognitive level of processing. At the social and physical level, Figure 2.2 consists of five components (from right to left: 1) the contextual component; 2) the cognitive space component; 3) the interface component; 4) the information object component; and 5) the information technology component). The social and physical level illustrates that the five components are observable in the physical world, and hence that these components can be studied with different investigative methods. Each component consists of cognitive structures or ‘models’ originating from various actors (e.g., system designers, authors, indexers or information seekers). The cognitive structures are created in interaction with the cognitive actor’s contextual environment (indicated with bi-directional arrows in Figure 2.2). In line with the third central feature, discussed in Section 2.2.3, this includes influences from the information seeker’s previous experiences and contextual environment.

From the information seeker’s perspective, the cognitive-emotional level consists of his or her cognitive model, which has been developed over time through interaction with the contextual component, as well as his or her perception of the situation at hand, including the work task situation. This perception prompts the information seeker’s perception of the problematic situation, and his or her perceived information need. The cognitive-emotional level depicts the information seeker’s interpretation of the five central components at the social and physical level.
The five components depicted at the social and physical level of Figure 2.2 can all be studied in isolation, but due to interaction between the components, a holistic cognitive approach regards such isolated investigations to be rather ineffective. This is related to the interconnectedness between theoretical understanding, empirical investigations, and technological developments, mentioned in Section 2.1, and the fact that IR processes always are embedded in broader information seeking activities, as illustrated in Figure 2.1. Figure 2.2 aims at illustrating all information seeking and IR processes, whether work or leisure related. Most real-life information seekers consider the seeking and/or retrieval process to be instrumental, and not a goal in itself. Hence, the real issue “[...] in IR systems design and evaluation is not whether a proposed method or tool is able to improve recall / precision by an interesting percentage with statistical significance. The
real concern is whether it helps the searcher better solving the seeking and retrieval task” (Ingwersen & Järvelin, 2005, p. 314). For instance, Turpin and Scholer (2006) show that a performance score from 0.55 to 0.95 mean average precision hardly affects the user’s task performance. This further supports the observation by Buckley and Voorhees (2005, p. 59), that there is no single user application that directly motivates mean average precision. Therefore, designers of IR systems need to consider the usability of their system in a broader context. That is, the contribution of an IR system could be evaluated at the end of the information seeking process.

The integrated framework subsumes the two components of focus for the traditional system-driven approach, namely the information object component and the information technology component, and it is considered as a holistic approach to LIS research. An example of research within this holistic approach is Borlund (2000), who develops a hybrid framework of interactive IR evaluation which manages to combine strengths from the system-driven and the cognitive approach, e.g., by maintaining experimental control and realism, respectively.

Ingwersen and Järvelin’s cognitive framework denotes a shift from believing that it is possible to bring the variety of cognitive and functional structures into accord with one another, to the acceptance of the diversity of such structures, and that research should aim at exploiting this diversity (Ingwersen & Järvelin, 2005, p. 30). An example of exploitation of diversity is the principle of polyrepresentation. The principle explicitly takes advantages of the diversity in cognitive representation, and hereby exploits the inconsistencies in such representations (Ingwersen & Järvelin, 2005, p. 208). The principle was introduced in 1992 (Ingwersen, 1992), fully expanded in subsequent work (Ingwersen, 1994; 1996; Ingwersen & Järvelin, 2005), and further developed in 2006 (Larsen, Ingwersen & Kekäläinen, 2006). The principle of polyrepresentation is applied in several recent empirical investigations (e.g., Larsen, 2004; Kelly, Dollu & Fu, 2005; Skov, Larsen & Ingwersen, 2006).

With this, we close the overall introduction of the cognitive viewpoint and the revealing of the interrelationship of the five central features of the viewpoint, and turn to how the present work builds on the cognitive viewpoint.

2.4 The cognitive viewpoint and the thesis

In this section, we outline how our work builds on the cognitive viewpoint. The objective of our work is threefold:
a) To gain a better understanding of aspects of users’ information seeking behaviour in relation to television broadcasts;
b) To support the design and construction of television broadcast retrieval systems by identifying appropriate access points for indexing and retrieval of television broadcasts; and, subsequently
c) To learn whether textual external and other sources contain information, which can be exploited for construction of surrogate records for television broadcasts

The former objective is related to the research questions 1, 2, and 3, while the second objective is connected to the research question 4, and the third objective is related to research question 5. The thesis is grounded on an empirical investigation of how and why users request television broadcasts, and our work meets the call by Ingwersen and Järvelin (2005, p. 356) to investigate the access points users are able and willing to use, and how such access points relate to representations of information objects in IR systems. Knowledge about aspects of information seekers behaviour is gathered in order to give an indication of the data elements that ought to be present in information object representations in a future IR system for television broadcast.

The influence from the cognitive viewpoint is particularly traceable in three characteristics of the experimental design of our investigation:

1) Involvement of genuine users;
2) Acceptance of personal and potential dynamic information needs; and
3) Acceptance of subjective, multidimensional, and potentially dynamic relevance assessments.

The connection between the three characteristics and our work is elaborated below.

The involvement of genuine users is one of the characteristics that differentiate the cognitive viewpoint from the system-driven approach. Our work is based on empirical investigations of genuine users in the academic field of Media Studies. The second and third characteristic is concerned with the user’s personal and subjective perceptions of the situation at hand, and accordingly the characteristics underline the importance of investigation of genuine users in the first place.

Information needs are based on the user’s perception of the work task at hand, and therefore an information need is related to the individual user. A user may change his or her perception of the underlying work task and/or the information needed in order to handle the work task. Hence, information needs are potentially dynamic, which is in
When the information need is personal and dynamic so becomes the relevance assessments. The relevancy of an information object is the result of the user’s perception of the information need at the time of the situation at hand. This implies that dynamic relevance assessments in real-life settings exist. That is, the assessments are not static, but potentially dynamic in concordance with the user’s work task situation. In addition, different users may perceive and assess relevance differently, and the same user may assess the same information object differently over time, as illustrated in Borlund’s figure of the interaction between relevance judgments and the information need (Borlund, 2000, p. 44), as depicted in Figure 2.3.

Figure 2.3: The interrelationship between the judgment of situational relevance and the development of the information need during a dynamic and interactive IR session (Borlund, 2000, p. 44).

Figure 2.3 illustrates that the same work task may be perceived differently over time (CW-CWⁿ), this may instigate different perceptions of the information need (N-Nⁿ) which again may result in different relevance assessments (↔). It is important to stress that the dynamism illustrated in Figure 2.3 is of a potential nature, and hence not a determinant necessity. In this chapter, we devote a few words to the close connection between the concepts of information need and relevance, but the concepts are thoroughly discussed in the following chapter. The information need is discussed in Section 3.1 and the concept of relevance in Section 3.5.

In order to gain a better understanding of users and their behaviour when seeking television broadcasts, we focus on the information seeker, and consequently, we are mostly concerned with the cognitive actor component of Figure 2.2. More specifically,
our emphasis is on the information seeker’s perception of the situation at hand and the information objects, at the cognitive-emotional level. Ingwersen and Järvelin (2005, pp. 277-278) maintain that though the cognitive-emotional level is somewhat intangible due to the subjective perceptions, different investigative methods may be applied to obtain (although obtrusive) data on the information seeker’s perception. Further, the implications of our knowledge gained about information seekers at the cognitive-emotional level are discussed in relation to the social-physical level. We accomplish this by emphasising the ideal access points in surrogate records, and discussing the value of information for constructing such access points.

2.5 Summary

The cognitive viewpoint is concerned with human perception, interpretation, and transfer of information, and how human behaviour relates to acquisition of information. The viewpoint is based on the conception that concurrent consideration of research in information seeking and IR is advantageous for LIS. More specifically, the viewpoint emphasises that though important results are produced in each community the effect of such research would be even stronger if it is approached from the same overarching theoretical foundation. The strength would most notably be a mutual exploitation of results across the demarcation of the two communities. A demarcation that is ascribed the communities’ ancestry in Social Science and Computer Science, respectively. In addition, the viewpoint emphasises that a holistic approach is pertinent in order to advance LIS research theoretically, empirically, and technologically. In this way it is a further maturation of Ingwersen’s previous work attempting to bind research in human information behaviour into one overarching holistic cognitive approach (e.g., Ingwersen, 1996).

The aim of the cognitive viewpoint is not to identify one ‘best practice’ (i.e., methods for representation or IR techniques) for information seeking and IR, but rather to gain insight about characteristics and implications of each practice in order to understand which combination of active, passive and conceptual system structures is most suitable in a given situation (Ingwersen, 1992, p. 37). In the same way, the framework does not dictate the cumbersome and unrealistic conception that research at all times should encompass all of the five components and their interrelatedness. The individual researcher should focus on specific components and/or relations, but bear the remaining components and relations in mind. Therefore, our main focus on two of the five components (the cognitive actor and the information object) and their interrelation is within the cognitive framework. With reference to the three research aspects put
Chapter 2: A cognitive framework for information seeking and information retrieval

forwards by Ingwersen and Järvelin (2005, p. 3), and presented in Section 2.1, our work aims at the following:

a) Theoretically understanding aspects of information seeking in the context of television broadcasts;
b) Description and explanation of empirical investigation of aspects of information seeking behaviour in the context of television broadcasts; and
c) Supporting the development of future technology, by providing identification of appropriate access points for indexing and retrieval of television broadcasts.

We aim at providing a proposal for technological developments by identifying suitable access points for television broadcasts, based on description and explanation of empirical evidence of aspects of the information seeking behaviour of users in Media Studies in relation to television broadcasts. This again is based on a theoretical understanding of information seeking of academic users’ in the context of (moving) images. With a focus on investigation of genuine users’ behaviour, it is possible to discuss the consequences for design of broadcast retrieval systems. That is, how appropriate access points can support indexing and retrieval of television broadcasts. We approach the viewpoint from a seeking perspective and the following Chapter 3 focuses on previous research on information seeking behaviour.
3 Information seeking behaviour

The literature on information seeking behaviour is immense. Case (2002, p. 224) estimates it to comprise more than 10,000 publications. Further, the diversity of foci range from the development of a general theory or model (e.g., Vakkari, 2001b), to investigation of information seeking behaviour in very specific contexts (e.g., female police officers working with undercover prostitution (Baker, 2004)). We neither have the space, nor would it be relevant for our work, to present a broad overview of information seeking research as such. Instead, this chapter focuses on theoretical and empirical developments mainly stemming from the 1980s and forth. Developments that we believe are important in the context of the present doctoral work.

Research questions 1, 2, and 3 are all concerned with aspects of users’ behaviour when seeking information. Though our focus is on a specific type of information objects (television broadcasts) and a specific domain (Media Studies), earlier research of a general nature is important for the present work. The chapter comprises six sections. The 1st section is concerned with the information need, and introduces and analyses the development and understanding of information needs in LIS. In this way, Section 3.1 is a natural continuation of the previous chapter on the cognitive viewpoint, and it relates directly to research question 1. In the 2nd section, we briefly outline proposals for modelling and theorising users’ information seeking behaviour, at a general level. These models are of importance for our analysis in Chapter 7. Section 3.3 and Section 3.4 focus on previous information seeking literature in relation to the specific domain and the specific type of information objects in question, and are relevant in relation to research questions 1 and 2. Media Studies is understood as a humanistic research field (see Section 5.3.3), and therefore Section 3.3 introduces and discusses previous research on humanists’ seeking behaviour. In Section 3.4, we present earlier research on information seeking behaviour in the context of moving images, since television broadcasts are a particular type of moving images. The visual element of moving images is believed to be very important in relation to the users’ behaviour. Consequently, the 4th section also contains discussions of previous research.

---

3 For reviews of the literature on information seeking we refer to one of the many chapters in Annual Review of Information Science and Technology (ARIST), the earliest chapters (Menzel, 1966; Herner & Herner, 1967; Paisley, 1968; Allen, 1969; Lipetz, 1970; Crane, 1971; Lin & Garvey, 1972; Martyn, 1974; Crawford, 1978), or the more recent chapters with emphasis on a cognitive approach (Dervin & Nilan, 1986; Hewins, 1990; Allen, 1991; Jacob & Shaw, 1998; Ingwersen, 2001; Vakkari, 2003). In addition, the monograph by Case (2002) provides a good review of previous research on information seeking behaviour.
on information seeking in the context of still images. The 5th section focuses on relevance and most notably the criteria users’ apply for assessing the relevancy of information objects. The section is important in relation to our third research question, which concerns the identification of users’ relevance criteria when seeking television broadcasts. The 6th and final section shortly summarises the chapter.

3.1 The potentially developing and dynamic information need

Wilson’s (2000, p. 49) definition of information seeking to be concerned with the purposive seeking for information in order to satisfy an underlying goal, ties it to the concept of information need, and our point of departure for this chapter is the maturing, potentially developing, and hence dynamic nature of the user’s information need. In the previous chapter we touched upon Belkin’s ASK hypothesis and the information need. Belkin perceives his ASK hypothesis to build on previous work by Taylor (1968), among others. Taylor further builds on, and develops Mackay’s psychological assumptions about the state of the human intellect in the question formulation process. Mackay (1960, p. 789) proposes that a question raises out of an inadequacy in what we might call the cognitive actor’s ‘state of readiness’ to interact purposefully with the world around him. The purpose or intention for formulation of a question is, according to Mackay (1960, p. 790), to cause a change in the questioner’s own state of readiness. Mackay’s idea of a ‘state of readiness’ is comparable to Belkin’s (e.g., 1980) ‘state of knowledge’ and Wersig’s (1971) ‘problematic situation’, as the underlying element that prompts a user’s need for information.

Taylor investigates the information need formation process empirically through interviews with reference librarians, and he proposes that the process develops in four stages. The information need stages are:

**Q1:** The visceral need, which “[…] is the conscious or even unconscious need for information not existing in the remembered experience of the inquirer. It may be only a vague sort of dissatisfaction. It is probably inexpressible in linguistic terms. This need (it really is not a question yet) will change in form, quality concreteness and criteria as information is added, as it is influenced by analogy, or as its importance grows with investigation” (Taylor, 1968, p. 182);

**Q2:** The conscious need, which “[…] is a conscious mental description of an ill-defined area of indecision. It will probably be an ambiguous and rambling statement. The inquirer may, at this stage, talk to someone else to sharpen his focus. He presumably hopes that two things will happen in this process: (a) his
colleague will understand the ambiguities; and (b) these ambiguities will gradually disappear in the course of the dialogue” (Taylor, 1968, p. 182);

Q3: **The formalized need**, at which “[…] an inquirer can form qualified and rational statements of his question. Here he is describing his area of doubt in concrete terms and he may or may not be thinking within the context or constraints of the system from which he wants information” (Taylor, 1968, p. 182); and

Q4: **The compromised need**, at which “[…] the question is recast in anticipation of what the files can deliver” (Taylor, 1968, p. 182).

These stages should be understood as interwoven points along a continuum. In IR activities the intermediary is initially presented with the user’s compromised need (Q4) and from here, the intermediary should aim at moving upwards in the continuum. The intermediary, whether human or computer, should try to generate an understanding of the user’s formalized need (Q3), as well as conscious need (Q2). This understanding is important as a mean to support the user in his or her fulfilment of the need. Understanding of the user’s visceral need (Q1) is unfeasible for any intermediary. Taylor (1968, pp. 183) discusses five filters intermediaries apply for analysing the expressed need, in order to understand the user’s need for information. The filters are: 1) determination of subject; 2) objectives and motivation; 3) personal characteristics of inquirer; 4) relationship of inquirer description to file organization; and 5) anticipated or acceptable answers. Particularly the second and the fifth filters are concerned with the user’s intentional cause that underlies the information need (e.g., the impending work task).

Hjørland (1997, pp. 162-165) sees Taylor’s levels to be stages of cognition. We agree with Hjørland, that modifications occurs in the user’s knowledge and perception of the situation, during the problem solving process, and that these modification are instigated by social interaction. This is related to the principle of complementary social and cognitive influence discussed in the previous chapter. The recognition of an information need matures out of a perception of an inadequacy, anomalousness, uncertainty, problematic situation, or wrongness, into a visceral need, which further matures into a comprised representation of the need, in verbal or literal form. The information seeking process is considered a cognitive process in which e.g., learning may occur. However, the formation of the information need is not of a predefined determinant character. The user’s perception of the instigative anomaly may change during the maturation, and the perception of what is needed may consequently be altered. This change may include the initiation of a new need (the recognition of a new
anomaly), a perception that the need is fulfilled, or a decision not to pursue the need further, though not fulfilled. Further, Taylor’s theory about the formation of an information need does not claim that any particular time-line exist for the formation or maturation, neither within the same, nor across different individual actors. One need may take years to be matured, while another may take seconds, only. The formation of an information need is closely related to the subjective development of cognitive structures. This said, the complexity of an information need, and the underlying work task situation, may very well affect the difficulties a user experiences when coping with his or her information need. Taylor’s second filter might in particular point to the fact that he acknowledges the formation of the need to be subjective in nature, due to its relation to the user’s cognitive structures.

The work by Taylor has motivated further research. For instance Ingwersen empirically investigates the compromised need (Q4) in protocol analysis of user-librarian negotiations, which leads to the identification of the ‘label effect’ (Ingwersen, 1982). He later classifies information need types (Ingwersen, 1986; 1992; Ingwersen & Järvelin, 2005), which are further discussed in the succeeding sub-section.

3.1.1 Information need types

Based on Belkin’s ASK hypothesis, Taylor’s theory on the development of the information need, the label effect, and previous empirical results, Ingwersen (1986, p. 223; 1992, pp. 116-117) proposes a classification of three fundamental types of information needs:

1) **Verificative needs**, where the user wants to verify or locate specific items. The user possesses the bibliographic or formal data necessary for requesting the information objects. In comparison to Taylor’s theory, a verificative need is at the level of the formalized need (Q3);

2) **Conscious topical needs**, where the user wants to clarify, review or pursue aspects of known subject matter. The user knows the terminology to express the need, and like the verificative need, the conscious topical need is related to Taylor’s formalized need (Q3); and

3) **Muddled topical needs**, where the user wants to explore some new concept or concept relations outside his known subject area. The cognitive structures of the user, concerning the required potential information, are weak or muddled, and the need is assumed to relate to Taylor’s conscious need (Q2).
According to Ingwersen, the label effect is usually present in muddled topical information needs, but may also occur in verificative and conscious topical needs. Therefore, it is of great importance that intermediaries gather knowledge about the user’s underlying work task, e.g., by applying Taylor’s filters, and in particular the second and the fifth filter (Ingwersen, 1986, p. 223).

In an elaboration of the above classification of information need types, Ingwersen and Järvelin (2005, pp. 291-292) categorise eight types of information needs:

1) **Known item**, where the cognitive actor seeks information objects by means of known formal or bibliographic features;

2) **Known data element**, where the cognitive actor seeks information entities by means of known structured data elements;

3) **Known topic or contents**, where the cognitive actor seeks information objects by means of known keys or features of potential information sources;

4) **Factual data**, where the cognitive actor seeks informative information (facts) by known content-associated or aboutness-related data;

5) **Muddled item**, where the cognitive actor seeks information objects by means of insufficient formal or bibliographic features;

6) **Muddled data element**, where the cognitive actor seeks or explores for information entities by means of vague or ill-defined data elements;

7) **Muddled topic or contents**, where the cognitive actor seeks or explores for information objects by means of ill-defined or vague knowledge of keys or features of potential information sources; and

8) **Muddled factual**, where the cognitive actor seeks informative information (facts) by ill-defined or vaguely known conceptual data.

The needs are classified with respect to three dimensions: 1) intentionality or goal of the information seeker (seeking information source content, e.g., journal articles, images or passages of such information objects, or seeking informative data entries, e.g., client address, author name); 2) the type of knowledge possessed by the information seeker (declarative information seeking and retrieval knowledge e.g., bibliographic or relational data, or declarative and procedural domain knowledge, procedural knowledge is e.g., knowledge about organisation of literature in a domain or search procedure skills); and 3) the quality of current knowledge, how well is the information seeker able to describe the features of the information objects which are perceived to solve the information need (well-defined, or ill-defined) (Ingwersen & Järvelin, 2005, p. 291). The first four information need types are well-defined and can be characterised as being
of a specific nature, while the latter four need types are of a more explorative nature. Though the three dimensions are described as dichotomous, the values in the third dimension (quality of known data) have infinite points between the two extremes of well-defined and ill-defined.

The eight types of information needs are not mutually exclusive in relation to a particular information seeker, and his or her work task at hand. Rather, the eight types may very well intermingle, and a combination of the different types of information needs may be applied at the same time. In addition, the eight information needs are not considered to be of a static nature, but because they are related to the user’s (changing) perception of the underlying work task, the needs are of a potential dynamic nature (see e.g., Figure 2.3, and our discussion of dynamic relevance in Section 3.5.1).

The point of departure for Ingwersen and Järvelin’s identification of information need types is the relationship between task complexity and types of knowledge (declarative and procedural) on one side, and information need formation and development on the other. The aim is to explore the features of a user’s knowledge and underlying tasks that generate which type of information need, because information seeking and retrieval “[…] deals with providing information for work task/interest performance – not directly with the solving process itself. Hence, if we can establish properties of information needs we are better capable of designing IS&R [Information Seeking & Retrieval] environments that may act on such properties during interaction with natural work and search tasks” (Ingwersen & Järvelin, 2005, p. 290).

In the present work, we propose a further expansion of the conception of the information need by incorporating task processes or actions in relation to retrieving information objects. Hereby, the concept of information need is related to Kuhlthau’s perception of the progression in the information search process, which is discussed in Section 3.2.3. Further, the need is connected to Vakkari’s expansion of Kuhlthau’s model, and in particular the relation between the types of information objects needed and the progression of the information search process. Vakkari’s work is further discussed in Section 3.2.5. Our conception of information need to incorporate task actions is in line with the perception expressed by the test participants in the present empirical work. In fact, tasks or actions are interwoven into the test participants’ conception of their information need, which is reported on in Chapter 7. Hereby, the present empirical investigation provides empirical evidence for the proposed extension of the concept of information need. The proposed expansion is mentioned here in order for the reader to be familiar with the theoretical basis of our conception. The conception is further discussed in Chapter 7 in connection to our discussion of the
empirical results of the characteristics of the test participants’ information needs in a television broadcast context.

In the following section, we introduce and analyse theoretical work concerned with modelling the user’s information seeking behaviour.

3.2 Models and theories of information seeking behaviour

The literature on information seeking has produced several models of, and theories about user behaviour. In this section, we introduce the models and theories we find most relevant for the present work. That is, models and theories that fit within the epistemological stance of the present work in the integrated framework for information seeking and IR.

Wilson argues that “[a] model may be described as a framework for thinking about a problem and may evolve into a statement of the relationships among theoretical propositions” (Wilson, 1999b, p. 250). Further, he states that most of the models that are concerned with information behaviour are statements of the former kind. They attempt to depict an information-seeking activity, the causes and consequences of the activity, or the stages of and relationship between stages in humans seeking behaviour. The models in information seeking do generally not involve fully formed theories, or directly suggest hypothesis to be tested. Rather, they are at a pre-theoretical stage, but nonetheless, they may propose relationships that are fruitful to examine.

The section comprises five sub-sections (Sections 3.2.1 - 3.2.5) each focusing on the work by prominent researchers in the field of information seeking behaviour. These researchers are B. Dervin, D. Ellis, C.C. Kuhlthau, K. Byström and K. Järvelin, and P. Vakkari. The work by these researchers has contributed considerably to the development of research on information seeking behaviour. Dervin is prominent in paying attention to the user’s feelings in relation to information behaviour. Ellis more specifically focuses on the characteristics of users’ seeking behaviour. Kuhlthau incorporates the emotional perspective in her focus on the search process. Byström and Järvelin, and Vakkari are proponents for investigation of the relationship between work tasks and information behaviour. Byström and Järvelin focus more specifically on the relationship between task complexity, types of information, and information sources. Vakkari elaborates on the work by Ellis, Kuhlthau, and Byström and Järvelin, and encompasses it into a model for task based information seeking. Apart from contributing to the field themselves, these researchers’ work has had an immense influence on later theoretical and empirical work conducted within information seeking research.
3.2.1 B. Dervin – The Sense-Making Approach

B. Dervin is the primary proponent of the Sense-Making Approach to information needs and seeking. Whether the approach constitutes a theory, a set of methods, a methodology, or something else, is open for discussion (Case, 2002, p. 146). We consider it a theoretical approach, which among others, builds on Dewey’s (1933; 1960) theories of constructivist learning in science, and Kelly’s (1991) analogue for formation of the individual in his Personal Construct Theory. That is, theories on the phases a human being progresses through in the process of constructing ‘sense of the world’, or cognitive structures. Kelly’s Personal Construct Theory is cursorily outlined in our discussion of Kuhlthau’s contributions in Section 3.2.3.

From Dervin’s Sense-Making perspective, an information need is connected to a person’s need to ‘make sense’ of a current situation. The need is recognised because a state transpire “[…] within a person, suggesting some kind of gap that requires filling” (Dervin, 1983a, p. 156). According to Dervin (2003) the information need is situated in a person’s perception of a particular situation (or work task), and the identified gap between what the person understands, and what he or she needs, to make sense of the current situation. The gap is manifested in questions (internal and/or external) to find help, or information, to bridge the identified gap. The outcome of the bridge building is so-called verbings or ‘sense’, which is comparable to the increment of cognitive structures in the cognitive viewpoint, albeit sense also incorporates emotional responses. Hence, the Sense-Making Approach highlights the cognitive actor’s emotional states when seeking information.

The Sense-Making Approach is a holistic information seeking approach, which emphasises that information systems should be understood from the cognitive actor’s point of view, instead of from the systems’ perspective. An important aspect of the Sense-Making Approach is to refute the quantitative notion of information, e.g., found in Shannon and Weaver’s communication theory (Shannon, 1948; Shannon & Weaver, 1949), and Buckland’s metaphor of ‘information-as-thing’ (Buckland, 1991). In the quantitative notion, information is comparable to ‘bricks’, to be filled into ‘empty buckets’, representing the passive human mind. From the Sense-Making point of view “[…] the empty bucket has evolved into a thinking, self-controlling human being. And information changes from brick to clay, moved and shaped in unique ways by each perceiver” (Dervin, 1983a, p. 169). The quotation by Dervin (1983a, p. 156) reproduced above might be interpreted in a quantitative way, but the Sense-Making Approach maintains that the cognitive actor, on the influence of experience and context, constructs information internally. The use of information is contextual and situational.
and in this way, the Sense-Making Approach directly supports the five central features of the cognitive viewpoint discussed in Chapter 2.

In the succeeding section, we introduce and discuss Ellis’ work on patterns of users’ information seeking behaviour.

3.2.2  
**D. Ellis – Information seeking features**

The point of departure for D. Ellis is the design of effective IR systems. More specifically, he focuses on the derivation of recommendations for IR systems design from analysis of actual users’ information seeking behaviour (Ellis, 1987, pp. 64-74; 1989a, pp. 171-172; 1989b, p. 237). Ellis’ focus on investigation of information behaviour as a prerequisite for design of effective retrieval systems is comparable to Pejtersen’s (e.g., 1989) work on fiction retrieval. Based on empirical investigations, Ellis (1987; 1989a; 1989b) proposes a behavioural model consisting of six features of information seeking patterns, and expands it to eight features in later empirical work (Ellis, Cox & Hall, 1993). The features are derived through empirical investigations of the information seeking behaviour of social scientists (Ellis, 1987, 1989a, 1989b), humanists (Ellis, 1993), and scientists (Ellis, 1993; Ellis, Cox & Hall, 1993) in academia, and engineers and research scientists in an industrial environment (Ellis & Haugan, 1997). The eight identified features are:

1) **Starting**: activities characteristic of the initial search for information in a new topic or area;
2) **Chaining**: following chains of referential connection between materials;
3) **Browsing**: semi-directed searching in an area of potential interest;
4) **Differentiating**: using differences between sources as filters on the nature and quality of the material examined;
5) **Monitoring**: maintaining awareness of the development in a field through the monitoring of particular sources;
6) **Extracting**: systematically working through a particular source to locate material of interest;
7) **Verifying**: activities associated with checking the accuracy of information;
8) **Ending**: activities characteristic of information seeking at the end of a project.

Ellis deliberately does not use the term stages in order to emphasise a non-deterministic and non-hierarchical relationship between the features in individual information seeking activities. The specific pattern of features is closely related to the circumstance associated with the information behaviour of a particular information seeker at a
Metadata elements preferred in searching and assessing relevance of archived
television broadcast by scholars and students in media studies

particular point in time. However, Ellis does indicate the existence of a relationship
between the features at a more general level (Ellis, 1989a, p. 178). Wilson elaborates
on this general relationship, suggests an ordering of the features, and proposes a
diagrammatic presentation of the model, which we have depicted in Figure 3.1.

![Diagram of Information Seeking Features]

Figure 3.1: A stage process version of Ellis’ behavioural framework (Wilson, 1999b, p. 255).

In the starting feature, the information seeker builds up a picture of the new topic or
area of interest, and aims at establishing familiarisation with the key aspects (e.g., ideas,
people or publications). Informal contacts, citations, and references are mentioned as
important means for gathering information in this feature. As indicated in the choice of
terminology, starting is an initial feature, which is followed by other features. Chaining
follows connections between information objects, most notably references and citations
in scientific texts, and it is hence strongly related to the application of citation
databases. Browsing implies that the information seeker looks through information,
which has been grouped or collocated, based on some kind of similarity. The similarity
may be based on a range of different characteristics (e.g., authors, titles, subjects)
depending on the information seeker, the work task at hand, and the type of information
objects being sought. Browsing is often applied in relation to familiarisation with a new
area. Monitoring implies that the information seeker is already familiar with a topic or a
research area. Monitoring is related to differentiation in that the information seeker’s
knowledge about the sources of interest is a prerequisite for the feature. Informal
contacts are important resources in relation to this feature. The feature of differentiating
is concerned with filtering information objects based on the information seeker’s (tacit)
knowledge about characteristics of (non-)relevant information objects. Filtering can be
based on the seeker’s knowledge about the quality of sources, relevancy of types of
information objects, the perspective or approach applied in the source etc. The feature
of extracting requires the previous identification of relevant sources e.g., by browsing
and/or differentiating, and it is strongly related to the feature of monitoring. Verifying
is comparable to the feature of chaining, the difference being that it requires some initial
information, which the seeker aims at verifying in other information objects (e.g., the original source). Ending is somewhat similar to the starting and the monitoring features but it is conducted at the end of a project to check if something has appeared during the last period of the project, e.g., during the final preparations of a manuscript for publication.

The strength of Ellis’ stage model is that it is based on empirical investigations in different contextual environments, and the general nature of the model makes it applicable for describing information seeking activities. But as Ingwersen and Järvelin (2005, p. 64) point out, if one is to explain information seeking behaviour “[…] in terms of the work tasks the subjects are engaged with, or their knowledge of the task, the features fall short because they are not explicitly related to such external possible causative factors. Neither is there any possibility to predict the order in which the categories appear in an individual seeking process”.

Slightly different features are found in the various studies. The differences are two-fold. The main part is terminological and is derived from the grounded theoretical methodology applied. More specifically, due to the fact that “[…] the latter studies were not simple verification studies, as, in each case, the models were developed from the data and then compared to the original model” (Ellis, 1993, p. 483). A smaller part is due to subject differences between the groups investigated. This is related to the results by Allen (1977), and Garvey, Tomita and Woolf (1979), that because of diverse problem solving functions, different information sources are needed in the various phases of a project or work task. The work by Ellis and colleagues lend support to the perception that an information seeker adopts different approaches and reveals various behaviours at different stages of an information seeking activity, and that the notion of information seeking transpire in stages that link to the accomplishment of an underlying work task, including the type of information objects being sought.

Ellis and colleagues’ investigations are all conducted in the context of textual scientific information objects, and the type of information objects under scrutiny may very well affect the model. For instance, the chaining feature is strongly connected to the referential characteristic of scientific publications, and when investigating aspects of seeking behaviour in the context of non-textual, non-scientific information objects (e.g., television broadcasts) which do not have the same referential characteristic, a behavioural difference is expected. Further, the characteristics of the information objects, which are important in relation to the features, are potentially different for different domains and types of information objects sought. For instance, citations are identified as indicators of quality for scientific articles. Television broadcasts do not
contain citations instead audience ratings might be expected to have a similar quality function for this type of information objects.

3.2.3 C.C. Kuhlthau – The Information Search Process

Similar to Ellis, C.C. Kuhlthau proposes a model of the Information Search Process (ISP). Kuhlthau’s model derives through longitudinal empirical studies in educational as well as work place settings (e.g., Kuhlthau, 2004). Like Dervin, Kuhlthau builds on Kelly’s (1963; 1991) Personal Construct Theory, which elaborates Dewey’s (1933) five philosophical phases of reflective thinking, into five psychological phases of construction. The five phases are summarised by Kuhlthau (1993, pp. 20-21) to be: 1) confusion and doubt (when confronted with new information); 2) mounting confusion and possible threat (when the new information cannot be easily assimilated with the existing cognitive structures); 3) tentative hypothesis (where the person choose to create a hypothesis to assimilate the new information); 4) testing and assessing (where the hypothesis is assessed against the established cognitive structures); and 5) reconstruing (where the meaning of the new information is assimilated, and new cognitive structures created). At any point during the process, the person may choose to give up on assimilation of the new information, and break off the process. The assimilation of new information in Kelly’s Personal Construct Theory is comparable to the increment of knowledge in Brookes’ fundamental equation for information science. However, Kelly emphasises the feelings associated with each phase in the process of the construction of new cognitive structures. It is exactly the relation between feelings and cognition, which is a central element in Kuhlthau’s work, and along with Dervin, she is the prime proponent for focus on the information seeker’s feelings in information behaviour studies.

Kuhlthau’s (2004, pp. 44-50) model suggests that the information search process consists of six stages:

1) Task initiation, when it is recognizes that information will be needed to complete the assignment;
2) Topic selection, when a general topic is selected;
3) Pre-focus exploration, where information on the general topic is investigated to extend personal understanding and form a focus;
4) Focus formulation, where the task is to form a focus from encountered information;
5) Information collection, where the task is to gather information pertaining to the focused topic; and
6) **Presentation**, where the task is to complete the search and to prepare to use the findings.

The six stages of the Information Search Process model are depicted in Figure 3.2. The model associates the user’s feelings, thoughts, and actions to each stage. The user is initially uncertain in relation to his or her tasks and information seeking is explorative. Towards the end of the search process, the user has established a focus, he or she is confident, and information seeking is structured and directed towards the work task at hand. As describe by Kuhlthau “[...] a dip in confidence is commonly experienced once an individual has initiated a search and begins to encounter conflicting and inconsistent information. A person “in the dip” is increasingly uncertain until a focus is formed to provide a path for seeking meaning and criteria for judging relevance” (Kuhlthau, 1999a, pp. 12-13).

<table>
<thead>
<tr>
<th>Stages</th>
<th>Task initiation</th>
<th>Topic selection</th>
<th>Pre-focus exploration</th>
<th>Focus formulation</th>
<th>Information collection</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feelings (affective)</td>
<td>Uncertainty</td>
<td>Optimism</td>
<td>Confusion/frustration/doubt</td>
<td>Clarity</td>
<td>Sense of direction/confidence</td>
<td>Satisfaction, or disappointment</td>
</tr>
<tr>
<td>Thoughts (cognitive)</td>
<td>Vague</td>
<td>Focused</td>
<td>Increased interest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actions (physical)</td>
<td>Seeking relevant information</td>
<td>Exploring</td>
<td>Seeking pertinent information</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.2:** Model of the Information Search Process (ISP) (Kuhlthau, 2004, p. 82).

Kuhlthau’s model of the Information Search Process is theoretically constructed from a synthesis of the theories of Kelly (1991), Taylor (1968), and Belkin (1980), all previously presented in this thesis (Kuhlthau, 1988a, pp. 233-234; 1991, p. 363). However, the model has been verified and further developed through several empirical investigations of e.g., high school students (Kuhlthau, 1988b), academic, public, and school library users (Kuhlthau et al., 1990), and recently in the workplace (e.g., securities analyst (Kuhlthau, 1999b) and lawyers (Kuhlthau & Tama, 2001)), where information is used for complex work tasks.

Kuhlthau and Tama’s (2001) recent empirical investigation in the workplace indicates that a user’s experience at the different stages of the Information Search Process is related to the user’s familiarity with, and perceived complexity of the work task at hand. A novice user perceives more tasks to be complex than an expert user. To
Kuhlthau the initial uncertainty, confusion, and anxiety especially among novice users verifies Kelly’s Personal Construct Theory in that encounters with new information begins with negative feelings. Kuhlthau (1999b, p. 409) concludes that it is the perception of the complexity of a task, rather than the actual, objective complexity of the task, that seems to be the critical factor in experiencing uncertainty. A user’s affective experience is related to his or her perception of the complexity of the task at hand. Kuhlthau’s claim that the user’s experience is related to the user’s perception of complexity fits nicely the results by Byström and Järvelin on work task complexity, presented in the succeeding section.

3.2.4 K. Byström & K. Järvelin – Task complexity, information types, and information sources

K. Byström and K. Järvelin focus on the complexity of a task and its relation to a person’s information seeking activities (Byström & Järvelin, 1995; Byström, 1997; 1999). Derived from empirical investigations of civil servants from the city secretarial office, Byström and Järvelin divide task complexity into categories, which are based on the a priori determinability of tasks input, process, and outcome. Byström and Järvelin propose five categories of increasing complexity: 1) automatic information processing tasks, (routine tasks which in principle could be automated); 2) normal information processing tasks, (almost completely a priori determinable, though a part of the process and information needed is a priori indeterminable); 3) normal decision tasks, (quite structured, including primarily case-based arbitration); 4) known genuine tasks (type and structure of result is known a priori, but procedures for performing the task is largely indeterminable); and 5) genuine decision tasks, (neither results, nor information requirements are a priori determinable) (Byström & Järvelin, 1995, pp. 194-195; Byström, 1997, p. 130; 1999, pp. 43-44).

Three different information types are applied: 1) problem (or task) information, considers the structures, properties, and requirements of the problem (or task) at hand, mainly in factual form; 2) domain information, considers common facts and other knowledge in the specific domain in question; and 3) problem (or task) solving knowledge, is instrumental knowledge about performance of tasks in the domain. In comparison to Ingwersen and Järvelin’s types of knowledge possessed by the information seeker, the two former are declarative, while the latter is procedural. Byström and Järvelin identify six different types of sources in three categories in their domain of study. The location of the sources may be internal or external in relation to the organisation.
Byström and Järvelin’s study shows, convincingly, that task complexity is related to various dimensions of information source use. They demonstrate that when task complexity is increased, so is the variety of different information types needed for solving the task, the number of sources, the share of general purpose sources, and the share of sources located internally. The more complex a task is perceived by a person, the greater is that person’s need for task solving information. According to Ingwersen and Järvelin, this corresponds to Luria’s (1976) discoveries that “[…] situational problem solving knowledge commonly is easier to learn and remember – and least likely to slip the memory – than categorical hierarchical conceptual structures on contents, which are the first to be forgotten” (Ingwersen & Järvelin, 2005, p. 288).

The differences between novices and experts have been discussed earlier in the literature. For instance, De Mey stated that “[c]omparisons of novices and experts focus on the changes in accessibility of knowledge resources that are manifest in the transition from novice to expert. Whereas both groups are expected to master the required knowledge sources, experts express a degree of flexibility and organization which is manifestly lacking in novices” (De Mey, 1984, p. 108). Further, Isenberg (1986) shows that differences between experts and novices exist in that experts, in Kuhlthau’s terminology, reach certainty faster than novices, which is verified in the findings by Byström and Järvelin.

In collaboration with Kuokkanen, Vakkari further develops the work by Byström and Järvelin by examining the growth of theories from Wagner and Berger’s (1985) conception of sociological theories (Vakkari & Kuokkanen, 1997; Vakkari, 1998). Vakkari (1998, p. 379) finds that Byström and Järvelin’s work meant a growth in theory in relation to increasing precision, broadening scope, and more empirical evidence. In addition, Vakkari and Kuokkanen point to new relations between concepts not identified in the original proposal from Byström and Järvelin (Vakkari & Kuokkanen, 1997). Having already brought up the contributions by Vakkari, it is appropriate to continue with a proper introduction of his work.

3.2.5 P. Vakkari – Task based information seeking

For P. Vakkari information seeking and retrieval is essentially a task based activity. A work task is the underlying driving force of any seeking or retrieval activity (Vakkari, 1999). Vakkari puts forward a model of the information search process in task performance, based on the hypothesis “[…] that the stages of task performance are connected to the types of information searched for, to the changes of search tactics and terms and relevance judgments” (Vakkari, 2001a, p. 296). The model is depicted in Figure 3.3. Vakkari mainly elaborates on Kuhlthau’s work on the Information Search
Process, but he also explicitly notes the empirical investigations by Byström and Järvelin (1995), Ellis and Haugan (1997), and Wilson (1999a) on the relationship between information search tactics and different phases of work task performance, as important for the development of his model. Theoretically, Vakkari’s model is based within the cognitive viewpoint, and indicates a pattern between the stages of task performance, IR tactics, and relevance judgments (Vakkari, 2001a, p. 296).

<table>
<thead>
<tr>
<th>Domain knowledge</th>
<th>Task performance process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage</td>
<td>Information</td>
</tr>
<tr>
<td></td>
<td>Search Process</td>
</tr>
<tr>
<td>Mental model</td>
<td>Differentiation</td>
</tr>
<tr>
<td></td>
<td>Integration</td>
</tr>
<tr>
<td>Use of information</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IR system knowledge</th>
<th>Specificity of information sought</th>
<th>Expected contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Query formulation support</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search tactics</th>
<th>Search keys</th>
<th>Operators</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Relevance judgments of references</th>
<th>Documents obtained and used</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Degree of relevance</td>
<td>- Degree of relevance</td>
</tr>
<tr>
<td>- Criteria used</td>
<td>- Type of contribution</td>
</tr>
<tr>
<td>- Precision and recall</td>
<td></td>
</tr>
</tbody>
</table>

| Legend: Arrows indicate the direction of impact. |

**Figure 3.3:** Vakkari’s model of the information search process in task performance (Vakkari, 2001a, p. 308).

The model supports and integrates Kuhlthau’s stages as part of the work task performance process, and expands it into IR. The user’s perception (the mental model) of the work task is reflected in the stage of the work task performance process. Integration of the mental model, and thereby the stages of task performance affects the user’s ability to express his or her expectations to the information needed and its contribution to the task at hand. The user’s mental model affects what happens during
IR interaction, and it also affects the user’s relevance judgments. A comparable relationship between information seeking, information need development, and relevance judgments is previously proposed by Ingwersen and Borlund (1996) and is depicted in Figure 2.3. Vakkari’s model predicts that stages of task performance and the mental model of the searcher, affects the “[…] information types sought for, search tactics and term choices for identifying the documents as well as patterning of relevance assessments and consequently, the success of searches in terms of the relevance of documents found” (Vakkari, 2001a, p. 306).

Vakkari’s model of information seeking and work task performance is empirically based and later verified in investigations of university students (Vakkari, 1999; 2000; Vakkari & Hakala, 2000; Vakkari, 2001a; 2001b; Pennanen & Vakkari, 2003; Serola & Vakkari, 2005). The model separates declarative domain knowledge from procedural searching knowledge, and work tasks from search tasks, which is a strong point of the model. Also the ‘expected contribution’ that refers to the user’s gained experience, is considered an asset of the model. Though the model is associated with operational IR systems, it could be seen in a broader perspective (Ingwersen & Järvelin, 2005, pp. 199-200).

We agree with Vakkari (2001a, p. 309), that the model demonstrates that the information sought, search tactics, term choices and relevance judgments and the contributory types of information in documents, systematically depend on the stage of the task performance process, and the mental model of the searcher. The obvious conclusion is that it is productive to study information searching and information retrieval in particular, as a process in connection with the underlying work task. Knowledge about the searchers perception of the task enables us to obtain research results, which will also provide useful information for designing information systems.

Vakkari’s model is the last general model of information seeking behaviour to be discussed in this section. In the following sections, we introduce, outline, and discuss work on information seeking in relation to specific contexts. The literature on information seeking behaviour is very broad, and it can be divided into several categories depending on ones purpose. The present work investigates the behaviour of scholars and students in Danish University departments of Media Studies, when seeking television broadcasts. Hence, we are concerned with information seeking behaviour from two perspectives: occupation and type of information objects sought. Therefore, in addition to the general theories and models presented in this section, we are concerned with literature with a more precise focus in relation to the contexts that frame our work. In the following two sections, we first discuss information seeking literature
in relation to academics in the humanities, and later outline literature on information seeking behaviour in relation to still and moving images.

### 3.3 Humanists’ information seeking behaviour

Investigations of information seeking behaviour have predominantly been framed within an occupational context, with academia as the most frequent category (Julien & Duggan, 2000). The literature is immense and diverse, and the results from the various disciplines under scrutiny have consequently been aggregated into meta-disciplines (Case, 2002, p. 232). The 1950s and 1960s gave emphasis on the needs of scientists and engineers. The needs of social scientists was attended to in the 1970s, and attention turned to the arts and humanities in the 1980s and 1990s (Bates, 1996c, p. 155). The application of meta-disciplines will naturally bring along discussion of categorisation of specific disciplines (e.g., is Library and Information Science a humanistic, social science, or science discipline). No previous research has devoted specific attention to information seeking behaviour of scholars and/or students in Media Studies. Media Studies is understood as a humanistic discipline in the present work (see Section 5.3.3 for a presentation of the field of Media Studies), we are therefore inspired by earlier results on information seeking of academics in the meta-discipline of humanities.

This section consists of two sub-sections. In the first sub-section, we present a model of the stages of work tasks of humanists’ researchers, and the second sub-section presents more specific results about the seeking behaviour of humanists.

#### 3.3.1 The humanist researcher’s work stages

Chu focuses on the work tasks and associated processes of academics in the academic field of Literary Criticism, as a mean to investigate information seeking behaviour. She identifies six stages, or research phases, of the literary critic’s work. Chu’s (1999, p. 259) six stages are:

1) **Idea**: where the literary critic, e.g., generates an idea, initiates the project, creates initial formulation of the idea, decides on literary texts for study, and discusses the idea(s);
2) **Preparation**: where the literary critic, e.g., searches for primary and secondary materials or resources, uses/reads primary and secondary sources, takes notes from primary and secondary sources, tries out idea(s), and applies for funding;
3) **Elaboration**: where the literary critic, e.g., thinks of a focus for each section of the work, maps/sketches idea(s) for work, creates an outline, organises notes to represent structure of work, creates shape of argument, discusses idea(s), and applies for funding;

4) **Analysis and writing**: where the literary critic, e.g., drafts the work, revises the work, obtains help with work, searches for more information, re-reads text(s), notes and other material, and preliminarily explores dissemination channels;

5) **Dissemination**: where the literary critic, e.g., explores dissemination channels, publishes or presents work, and applies for funding; and

6) **Further writing and dissemination**: where the literary critic, e.g., explores dissemination channels, searches for more information, reads and/or re-reads text(s), notes and other material, re-writes work, publishes or presents reworked document.

In the ‘preparation’ and ‘analysis and writing’ phases the literary critic has the most extensive need for information, while in the ‘further writing and dissemination’ phase he or she has a moderate need for information, and in the ‘idea’, ‘elaboration’, and ‘dissemination’ phase only a small amount of information is needed (Chu, 1999, pp. 259-262). For example, Chu (1999, p. 266) finds that the ideas for literary criticism often originate from issues arising from previous work or from teaching. This illustrates that for many critics, a new project builds on previous knowledge, which indicate that literary critics do minimal information searching in the generation of new ideas.

In similar vein to Ellis, Chu takes pain in emphasising that the phases do not occur in pre-determined linear sequence, and that all phases do not necessarily occur. She proposes several variations of the pattern of the phases, and leaves it open for additional sequences to be possible. She also states that a literary critic may very well have several intermingling research project going on at the same time. Chu’s model is not to be considered a linear deterministic model, but rather an indication of various possible variations of the pattern, moving from initiation (generation of idea) to ending (dissemination of results). In addition, Chu’s model includes the possibility of further writing and dissemination, and hereby the model indicates a cyclic behaviour. In Figure 3.4, we propose a diagrammatical depiction of Chu’s model of the humanistic researchers. The arrows designate the dynamic and cyclic characteristic of the model.
Figure 3.4: A process version of literary critics’ research phases.

Figure 3.4 resembles Wilson’s depiction of Ellis’ model of information searching activities (see Figure 3.1). However, it is important to emphasise the difference in focus of the two models. Ellis focuses on information seeking, while Chu investigates the underlying research phases, and, among other things, the associated information seeking activities. Chu contrasts her results with earlier models of work stages, and she finds that her results are comparable to the five stages of Stone’s more general model of work tasks in the humanities (Stone, 1980, p. 16).

The following section focuses on more specific results concerned with the information seeking behaviour of humanist researchers.

3.3.2 Information seeking in the humanities

Several authors (e.g., Stone, 1982; Wiberley & Jones, 1989; Watson-Boone, 1994; Bates, 1996b) report that humanist scholars’ use of technology is low. It is a common point of view that the low utilization of technology in the humanities is caused by technophobia (Case, 1991, p. 75). Wiberley and Jones (2000, p. 426) question this line of reasoning and argue that the low usage of technology is caused by the limited value of technology to advance in humanistic scholarship. Technology is merely a mean to an end, and if usage of technology is perceived as beneficial for the scholar, then the scholar uses technology. This behaviour is comparable to Zipf’s Principle of Least Effort or the similar Mooers’ Law. The former suggests that people will “[...] strive to minimize the probable average rate of work-expenditure (over time). And in so doing he will be minimizing his effort” (Zipf, 1972, p. 1), while the latter advocate that “[a]n information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer [or user] to have information than for him not to have it” (Mooers, 1960, p. II). The Principle of Least Effort is widely recognised to exist in
information seeking behaviour, e.g., as found in Poole’s work on theory construction in LIS (Poole, 1985).

Bates conduct a longitudinal study of 27 humanists’ online searching behaviour, she claims that the low usage of IR systems in the humanities is connected to the fact that the needs and behaviours of scientist are the basic assumption for IR theory and research. She further claims that scientists’ needs and behaviour may not be applicable when studying academics in the humanities. Bates finds that a humanist expects to be familiar with the majority of the retrieved results, and for instance, a 200-item search is considered to be of great success if it produces one or two novel references (Bates, 1996a, pp. 699-700). Grover and Hale similarly observe that humanist scholars aim at adding new knowledge to a topic in which they are already knowledgeable, and in which they have previously completed some research (Grover & Hale, 1988, p. 11). A humanist is broadly speaking familiar with most authors and titles in their field. A knowledge which they mostly obtain by chaining, in Ellis’ terminology, that is, gathering references from seed documents, from interaction with their everyday colleagues, or interaction with colleagues at conferences (e.g., Tibbo, 1993, pp. 116-121; Bates, 1994; Wiberley & Jones, 1994, p. 506). Humanists are broadly speaking expected to know the items they are searching for, and in comparison to the types of information needs proposed by Ingwersen and Järvelin, this behaviour is comparable to the ‘known item’ need. A frequent reuse of literature in the humanities is also found in citation studies (e.g., Diodato & Smith, 1993). In addition, in her review of research on humanists’ information seeking behaviour, Watson-Boone (1994, pp. 206-207) identifies browsing as a central feature of humanists’ behaviour. Not least due to the element of serendipity, in browsing. In browsing researchers expect to be familiar with most items, and are happily surprised when they encounter unfamiliar items.

Wiberley and Jones maintain that humanist scholars gradually will become more involved with electronic information technology, but their involvement will always be influenced by considerations of time and will always be less than that of scholars whose fields are paradigmatic and who direct the creation of the evidence they use (Wiberley & Jones, 2000, p. 430). In their argument they persist an earlier definition of the humanities to be “…[...] those fields of scholarship that strive to reconstruct, describe, and interpret the activities and accomplishments of men and women by establishing and studying documents and artefacts created by those men and women” (Wiberley & Jones, 1994, p. 503). The differentiation of the humanities in relation to so-called paradigmatic fields is described as a decrease in control of the data material to be analysed. An important element of the definition, and thereby the distinctiveness of the humanities, is that information objects, or artefacts, created by humans, is the primary
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

sources for analysis, with the purpose of reconstructing, describing, and/or interpreting human activities. In popular terms, one might say that the library is for the humanist what the laboratory is for the scientist. We agree with this perception of the humanities, which we further outline in our description of the field of Media Studies in Section 5.3.3.

Because of this focus on primary data sources, Wiberley and Jones claim that the most important development in the humanities will be the extent to which humanists use electronic information technology to access the primary sources (Wiberley & Jones, 2000, p. 428). However, since most primary data sources in the humanities originated prior to digitalisation, the share of humanists that use digital data sources will be smaller in comparison to scientist and social scientists. In addition, the discoveries in the humanities depend on sources that are not previously brought to the attention of the discipline. Therefore, the benefit of digitizing the most heavily used material, and neglecting those materials least used, is questionable from a humanist’s point of view, while the opposite is questionable from a practical, political, and economical point of view.

In a study of the information seeking habits of master and doctoral students in the humanities, the participants expressed that they appreciated the efficiency and time-saving qualities of electronic information technology (Barrett, 2005, p. 326). Further, following the line of thought put forward by Wiberley and Jones (2000), Barrett claims that a new generation is appearing in the humanities, and that this has lead to a fairly dramatic shift in that electronic information technology is now much more commonly used as a tool (Barrett, 2005, p. 328). In the same study, Barrett demonstrates that the most frequent complaint participants put forward about electronic information technology, was the lack of available primary sources (Barrett, 2005, p. 326).

Collaboration in the humanities is less frequent than in both science and in social science. For instance, co-authorships are found for less than 5% of the publications in the humanities (e.g., Lowe, 2003, pp. 18-19; Dalton & Charnigo, 2004, p. 415). In her review of literature on the information needs and habits of scholars in the humanities, Watson-Bone concludes that the humanist scholar continues to work alone, and that this aspect appears to reflect a fundamental characteristic of humanistic research (Watson-Boone, 1994, p. 212). Watson-Bone maintains Stone’s (1982), Sievert and Sievert’s (1989), and recently Dalton and Charnigo’s (2004, p. 404) emphasis of the solitary nature of the humanities. In contrast to the solitude nature of humanists’ publications, humanists do collaborate in their information seeking activities. According to Green (2000, p. 205) scholars in general and scholars in the humanities, in particular, tend to rely on informal sources in their information seeking process, for instance collaboration

56
with colleagues. This behaviour may be contributed to by the flexible nature of human sources over documentary sources, e.g., as argued by Ginman (1983; 1984) or Byström (1999). Again, such behaviour could be ascribed Zipf’s aforementioned Principle of Least Effort (Zipf, 1972, p. 1).

Several authors have examined the information seeking behaviour of academic students (e.g., Kuhlthau, 1991; Valentine, 1993). As reported by Barrett (2005) the information seeking behaviour of students in the humanities differs from the information seeking behaviour of scholars in the humanities. The students “[...] describe haphazard origins to their research projects, often starting out with only a vague notion of a project’s final form or shape” (Barrett, 2005, p. 328). University students are required to understand a new topic in a short period of time. As the topic is somewhat new to the students, they are more uncertain towards their work tasks. Students themselves are aware that scholars spend less time in the initial stages of research, for instance topic selection and background reading (Barrett, 2005, p. 328). In relation to Kuhlthau’s (1991) model of the Information Search Process, it is expectable that students perform more explorative searches than faculty. In order to cope with these differences, an effective IR system in the humanities should obviously allow users to conduct explorative as well as focused searches.

3.4 Information seeking in the context of still and moving images

Only few previous studies have had a specific focus on the information seeking behaviour of users of moving images. However, due to the visual character, earlier research on the behaviour of users’ seeking still images seems relevant in a moving image context. This section, consequently, includes earlier research from both areas, with special emphasis on investigations of academic users. The earlier research is described in relation to several aspects. Firstly, in Section 3.4.1, we introduce and analyse research on typologies of users’ requests for moving images and still images. Secondly, we outline and discuss a few early studies on searching behaviour in the context of (moving) images, in Section 3.4.2. Thirdly, Section 3.4.3 outlines and discusses previous literature on the work with and use of television broadcasts.

3.4.1 Typologies of requests

Investigation of aspects of users’ needs for audiovisual information objects is a relatively recent topic in the research literature. The first consideration of the topic is Seloff’s cursorily description of the dominance of known item information requests at the film archives at National Aeronautics and Space Administration’s (NASA) Johnson
Space Centre. Seloff (1990, p. 684) reports that only 1/3 of the users’ requests are of a topical nature.

The first comprehensive and systematic investigation is Enser and McGregor’s analysis of requests put forward to the Hulton Deutsch Collection (Enser, 1993; Enser & McGregor, 1993). Enser and McGregor analyse 2,722 genuine requests from users with information requests emerging from their professional “work in commercial enterprises (e.g., book publisher and television producers). They group the requests according to two dichotomous categories: 1) uniqueness; and 2) refinement. A non-unique request contains generic concepts, and a unique request enquires specific persons, places, events etc. Refinement of requests refers to specification of time, location, action, event or technical specification. The precise type of refinement is disregarded in the observation. Table 3.1 gives an example of a request from each of the four categories and the distribution in percentages of the 2,722 request in the four categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>Unique</th>
<th>Non-unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-refined</td>
<td>Lady Diana</td>
<td>Paddle steamer</td>
</tr>
<tr>
<td></td>
<td>42%</td>
<td>6%</td>
</tr>
<tr>
<td>Refined</td>
<td>Lady Diana in silhouette in front of Buckingham palace</td>
<td>Paddle steamer on the Mississippi 1830 – 1850</td>
</tr>
<tr>
<td></td>
<td>27%</td>
<td>25%</td>
</tr>
</tbody>
</table>

Enser and McGregor categorise 69% of the 2,722 requests as unique requests, whether refined or not. 52% of the requests are categorised as refined (unique and non-unique), with time being the most frequent refinement appearing in 34% of the 2,722 requests. The results indicate that when the user puts a request forward to the archive, he or she is already in possession of a mental image of the required picture, or one which would satisfy his or her need. Indeed, a request may even contain a pictorial description in the form of a rough sketch (Enser & McGregor, 1993, p. 26).

In a study of genuine requests made to seven different archives (five archives of still images and two achieves of moving images), Armitage and Enser (1996; 1997) further develop Enser and McGregor’s typology of information requests. Armitage and Enser initially analyse 2,095 requests and identify four broad categories of requests: 1) requests concerned with visual content; 2) identification, attribution, provenance checking; 3) accessibility of image or artists work; and 4) miscellaneous (e.g., insufficiently detailed requests or administrative procedures requests) (Armitage & Enser, 1996, pp. 16-19). 1,749 requests are categorised within the first broad category,
and further analysed to elaborate the original categorisation proposed by Enser and McGregor. The elaboration is two-fold.

Firstly, the categorisation is broadened with two non-topical categories, namely ‘by named artist’ and ‘known item’. Armitage and Enser suggest, that requests of the type ‘by named artist’ are mainly related to people who seek (moving) images from an artistic viewpoint, and hence these requests are mostly present in collections containing (moving) images of art. The requests for ‘known items’ are particularly frequent in the two collections of moving images. Of specific interest for the present study is the requests addressed to the British Film Institute, National Film and Television Archive (BFI-NFTVA), because it resemblance the collection of our study. That is, a significant part of UK terrestrial television broadcasting from 1985 and forth is held in the collection (British Film Institute, 2006), and the user base is quite varied (Terris, 1996, p. 115). Armitage and Enser categorise 3% of the 365 requests addressed to BFI-NFTVA as ‘by named artist’, 40% as ‘known item’, and the remaining 57% are categorised to be of a topical nature. The two categories of ‘by known artist’ and ‘known item’ correspond to Ingwersen and Järvelin’s category of ‘known item’ information need.

Secondly, Armitage and Enser do not find the original categorisation of subject matter into unique and non-unique, with or without refinement, adequate for representing the different aspects of the requests. They further develop the subject categorisation of requests by adopting Shatford’s (1986) matrix for analysis of topical content of images (Shatford’s matrix is further elaborated on in Section 4.2.1). The two categories of unique (specific) and non-unique (generic) subject requests are supplemented with a third category containing abstract concepts. In addition, refinements are consider as facets of who, what, where, and when. Consequently, the 2x2 matrix developed by Enser and McGregor is transformed into a 3x4 matrix. A request may contain several facets, and the request may hence be place in different categories, as illustrated with the examples in Table 3.2.

Armitage and Enser’s investigation verifies Enser and McGregor’s results in that most topical requests are for specific or unique content. This is, in particular, the case for the television archive (BFI-NFTVA) for which 90% of the topical requests are categorised to be specific. The results from the BBC Natural History Unit (BBC-NHU), which is the second collection of moving images, are somewhat different. That is, approximately ¾ of the topical requests to BBC-NHU were categorised as specific, while all the requests were categorised as generic. The two collections are very different in their coverage and user group. BFI-NFTVA is heterogenous in character, with public access, and a varied user base (Terris, 1996; British Film Institute, 2006).
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

In contrast, the BBC-NHU holds a specialist collection of moving images on natural history, and the main user base is from a production company attached to the unit. These differences in collection characteristics and user base may be the reason for the differences in distribution of specific and general requests. Generally, across all seven collections, Armitage and Enser only categorise a few requests to be for abstract content.

Table 3.2: Armitage and Enser’s categorisation of visual information requests. Examples from Armitage and Enser (1997, pp. 290, 292), and results from Armitage and Enser (1996, p. 31), and Sandom and Enser (2002, p. 38).

<table>
<thead>
<tr>
<th>Who</th>
<th>Specific</th>
<th>Generic</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>Individually named person, group, thing</td>
<td>Kind of person or thing</td>
<td>Mythical or fictitious being</td>
</tr>
<tr>
<td>Results</td>
<td>Napoleon at Jaffa, 1979</td>
<td>Writers in Hollywood</td>
<td>Vampires, spirits</td>
</tr>
<tr>
<td></td>
<td>A: 58%; B: 64%; C: 32%</td>
<td>A: 55%; B: 14%; C: 36%</td>
<td>A: 0%; B: 0%; C: 0%</td>
</tr>
<tr>
<td>What</td>
<td>Individually named event, action</td>
<td>Kind of event, action, condition</td>
<td>Emotion or abstraction</td>
</tr>
<tr>
<td>Example</td>
<td>Martyrold of St Lawrence</td>
<td>Chinese foot binding</td>
<td>Depiction of vanity</td>
</tr>
<tr>
<td>Results</td>
<td>A: 2%; B: 9%; C: 9%</td>
<td>A: 30%; B: 15%; C: 27%</td>
<td>A: 1%; B: 0%; C: 1%</td>
</tr>
<tr>
<td>Where</td>
<td>Individually named geographical location</td>
<td>Kind of place: geographical, architectural</td>
<td>Place symbolised</td>
</tr>
<tr>
<td>Example</td>
<td>Writers in Hollywood</td>
<td>Physiological laboratory</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>A: 12%; B: 9%; C: 31%</td>
<td>A: 14%; B: 2%; C: 10%</td>
<td>A: 0%; B: 0%; C: 0%</td>
</tr>
<tr>
<td>When</td>
<td>Linear time: date period</td>
<td>Cyclic time: season, time of day</td>
<td>Emotion, abstraction symbolised by time</td>
</tr>
<tr>
<td>Example</td>
<td>Napoleon at Jaffa, 1979</td>
<td>Toads at night</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>A: 0%; B: 8%; C: 27%</td>
<td>A: 2%; B: 0%; C: 1%</td>
<td>A: 0%; B: 0%; C: 0%</td>
</tr>
<tr>
<td>Overall</td>
<td>A: 72%; B: 90%; C: 100%</td>
<td>A: 100%; B: 31%; C: 74%</td>
<td>A: 2%; B: 0%; C: 2%</td>
</tr>
</tbody>
</table>

Legend: Results are for requests addressed to the following collections:
A: British Broadcasting Corporation – Natural History Unit (BBC-NHU) (Armitage & Enser, 1996, p. 31)
B: British Film Institute – National Film and Television Archive (BFI-NFTVA) (Armitage & Enser, 1996, p. 31)
C: Overall finding for 11 collections: BBC Information and Archives – Television Archive; BBC – Natural History Unit; British Pathe; Film Images (London) Limited; Hunter Film Archive; Imperial War Museum Film and Video Archive; London Transport Museum Film and Video Collection; BFI NFTVA; North West Film Archive; Shell Film and Video Unit; and South East Film and Video Archive (Sandom & Enser, 2002, p. 38)

Armitage and Enser’s study is recently followed up by Enser and Sandom’s (Enser & Sandom, 2001; 2002; Sandom & Enser, 2002) analysis of 1,270 genuine requests put forward to 11 archives of moving images, including the two archives investigated by Armitage and Enser (BFI-NFTVA and BBC-NHU). Overall, they follow the same approach as Armitage and Enser, and their results are included in Table 3.2. Overall,
they find that 10% (122) of the 1,270 requests are of a non-topical character. Table 3.2 shows that the main part of the 1,148 topical requests received at the 11 archives (the ‘C’ results in Table 3.2) are for specific content (1,143 requests), ¾ (852) are for generic content, and only a few (22) are for abstract content. In this way, Enser and Sandom verify the results by Armitage and Enser. However, these numbers fail to disclose great difference between the 11 archives under scrutiny. For instance, for the BFI-NFTVA collection, ‘known item’ requests account for 57% (Sandom & Enser, 2002, p. V. 5), and hence verifies the high share of known item requests found by Armitage and Enser for this collection. The differences between the types of requests in the 11 collections may to a large degree be due to different user and collection characteristics.

Sandom and Enser (2002, p. V. 6) also look into the characteristics of the users of the BFI-NFTVA. They identify 31% of the users as commercial clients, 24% of the users come from educational institutions, 17% are personal clients, and finally 5% of the users are from non-commercial organizations. 23% of the requests made to the collection had no affiliation assigned.

Keister (1994) categorises requests reconstructed from query logs at the National Library of Medicine into two groups: 1) topical requests, in which no specific visual cues are given by the user; and 2) visual requests, in which user constructed perceptions of concrete elements and abstract concepts are describe with words. Keister estimates that the second type comprises one-third to one-half of the requests. She also demonstrates that users do not request images in a consistent manner. Picture professionals think visually and use art or graphics jargon e.g., horizontal colour image. Health professionals’ requests for images are more in the line with the terminology of the medical field, while the academic community “[…] often have precise citations to images” (Keister, 1994, p. 10).

Ørnager investigates the information requests for images at 13 Danish newspaper archives of still images (Ørnager, 1999; Ørnager, 1995). According to the archivists 60% of the request to the 13 archives are for named persons or objects appearing in the image, 20% are for a specific theme, and 20% are so-called complex requests that require in-depth interviews to retrieve relevant images (Ørnager, 1999, p. 108). By observation of the archivist handling user requests, Ørnager verifies the distribution and is able to identify five user types: 1) specific inquirer (narrow requests, because he or she has a specific image in mind), 2) general inquirer (broad requests, because he or she wants to make choice without inference from archive staff), 3) story teller inquirer (who tells about the story, and is open to suggestions), 4) story giver inquirer (who hands over the story to the archive staff, and lets them choose images), and 5) fill in space inquirer.
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

(who need to fill in space and cares about the size of the images, only) (Ornager, 1995, p. 215). Ørnager (1999, pp. 111-114) concludes that her five inquirer categories are comparable to the three general types of information needs (verificative, conscious topical, and muddled topical) identified by Ingwersen (1992, pp. 116-117). The specific inquirer, the story teller inquirer, and the story giver inquirer contains aspects of the verificative as well as the muddled topical information need. The general inquirer is comparable to the conscious topical need, while the fill in space inquirer is comparable to verificative need.

Like Ørnager, Choi and Rasmussen (2003) carry out an investigation of user requests within a domain specific collection of still images, the domain being American history. 38 end-users participated in the study, each contributing with one search request. Choi and Rasmussen apply four categories of topical visual information needs originating from Batley (1988b, pp. 374-375): 1) specific needs (specific persons, events or activity, e.g., Thomas Jefferson); 2) general or nameable needs (can be expressed in key terms, e.g., a ruined castle); 3) general or abstract needs (might involve abstract concepts rather than concrete objects, e.g., a busy street scene); and 4) subjective needs (emotional responses that are dependent on interpretation by the individual, e.g., a scene that illustrates how times have changed). Choi and Rasmussen find that 26% of the 38 requests are specific needs, 61% are general or nameable needs, 8% are general or abstract needs, and 5% are subjective needs (Choi & Rasmussen, 2003, p. 504).

Recently, Hertzum (2003) investigates film related requests addressed to the Deutsche Filminstitut (DIF) collection. He analyses 275 genuine requests made to the archive “[...] for text (e.g., dialogue lists and censorship cards), images (e.g., photos of actors), sound (e.g., the music from a film), video (e.g., video-copy of a film), as well as analyses (e.g., of the religious symbolism in the film Metropolis)” (Hertzum, 2003, p. 169). Based on Meadow’s four generic types of database search (Meadow, 1992, pp. 243-244), Hertzum categorises the request in: 1) known item retrieval; 2) fact retrieval; 3) subject retrieval; and 4) exploratory retrieval. Hertzum finds that 40% of the requests concern known items, 13% concern facts, 32% concern subjects, 10% are exploratory, and 3% could not be placed. Hertzum does not distinguish the specific type of information object requested. Consequently, the results do not specifically demonstrate whether the types of requests for e.g., moving images differ from the types of requests for film-related text. One may for instance maintain that fact retrieval is more often associated with text than with moving images.

Hertzum distinguishes between contextual and focus information in the requests. Contextual information informs on the requester’s underlying work task and what the
Chapter 3: Information seeking behaviour

user is going to use the information for. Whereas focus information specifies what is requested (Hertzum, 2003, p. 174). Hertzum finds that 57% of the analysed requests contain contextual information. Based on the contextual information Hertzum shows that 7% of the users’ requests reflect commercial activities, 30% of the requests are associated with education, 9% of the requests represent private and personal issues, and 10% of the requests are of non-commercial nature. To be expected, contextual information is least common for know item requests and most common in request of a topical nature. That is, in request for known items the sought information object can generally be identified on the basis of formal data, whereas e.g., information about point of view is important in request of a topical nature. Hertzum further classifies the requests according to the following six information need attributes: 1) production-related (e.g., title, production year, actor, book film is based on); 2) screening-related (e.g., cinema, TV channel, date and/or period the film was shown); 3) content-related (identifiable entities appearing in the moving images, e.g., persons, events, objects); 4) subject-related (the interpretation of the moving images e.g., theme, genre, emotional experience); 5) context-related (e.g., reviews, censorship material, magazines, festivals); and 6) other. He demonstrates that production related attributes is far the most frequent attribute in the 275 user requests. Again, Hertzum does not present results for the different types of information objects requested (e.g., text, sound, still image, moving image), and specific results for requests for moving images are, consequently, not produced.

Due to difference in the collections, users, and the different typologies applied in the reviewed studies, the results on categorisation of user requests may be difficult to summarise. However, with inspiration from Jörgensen (1999, p. 306), we propose the categories depicted in Figure 3.5 for summarising the results.

Non-topical requests may be summarised in the category of known items, which is comparable to Ingwersen and Järvelin’s ‘known item’ information need. Non-topical requests are not the main focus of any of the investigations, but when non-topical or known item information needs are differentiated in the analysis, they account for a large portion of the requests. The share of known item requests is pronounced in the archives of moving images, and especially in relation to the BFI-NFTVA, for which the share of known item requests is reported to be between 43% and 57%.
Topical, or content, requests are the main focus in all six investigations discussed above. Topical requests are comparable to Ingwersen and Järvelin’s ‘known topic or content’ information need. The main difference between the typologies is the specificity they apply for classification of information requests, e.g., illustrated in the difference between Enser and McGregor’s 2x2 matrix and Armitage and Enser’s 3x4 matrix. The same specificity may prove difficult to apply in actual indexing of moving images due to the resources needed, but it is appropriate for identifying features that could be relevant when designing an IR system. In summarising the results according to the three topical or content categories depicted in Figure 3.5, the results show that the majority of topical or content requests are for named entities, and that “[…] the demand for uniquely defined and named visual features increases when requests are for moving image items” (Sandom & Enser, 2002, p. 53). This said, the results are dissimilar across the different types of collections investigated, and they indicate that also general topical and subject entities are frequently contained in the user requests. When request for abstract concepts or affective aspects are specifically identified in the investigations, they represent a minority of the requests. The low frequency of requests for abstract concepts or affective aspects may be related to the label effect. That is, due to the fact that the investigated requests originate from operational retrieval system, a user may be reluctant to put forward request that they do not think the IR system can handle. Consequently, they feed the system with the information they consider relevant for the system, to handle their request. This is in line with the observation made by Keister (1994, p. 17), that operating IR systems for images generally can not handle such subjective requests.
The studies reviewed in this section are important, e.g., for gathering knowledge about the problems associated with indexing of moving images, but the typologies do not consider the dynamic devolvement and maturation of a user’s information need. Instead, they focus on a static representation of the need in form of the user’s request. Hence, e.g., muddled information needs may have matured into focused information needs prior to the user’s interaction with an archive. Further, the focus on static requests, might also explain the low share of requests for abstract concepts or affective aspects, found in the reviewed studies. The studies are conducted in non-digitized collections, and they do not involve user interactions. Therefore, they cannot explain user searching behaviour in terms of how users formulate their requests, and the strategies applied for searching moving images. In the following section, we briefly present studies that have focused on the users’ behaviour when searching for (moving) images. We do this by focusing on the tactics users apply when searching for images.

3.4.2 Searching behaviour

Rorvig and colleagues work on the use of a so-called visual thesaurus for retrieval of items in the collection of (moving) images at NASA (Rorvig, Turner & Moncada, 1999). The motivation for development and implementation of a visual thesaurus for the collection was the creation of a controlled vocabulary by which the collection could be indexed and subsequently searched. The authors propose a thesaurus which equivalences “[…] images drawn from the collection with broad categories of terms” (Rorvig, Turner & Moncada, 1999, p. 794). Experimental results indicate that the novelty value of the visual thesaurus makes it a favoured search tool. However, as the novelty value of the visual searching features runs out, and the user becomes more of an expert in searching the collection, the traditional textual search interface is favoured. It seems that a visual searching interface is most useful with new users, and it may hence not be applicable in all situations.

Batley (1988b; 1988a) identifies four search strategies: 1) seeking; 2) focused exploring; 3) open exploring; and 4) wandering. According to Batley, the user will adopt a familiar search strategy and use a variety of activities ranging from non-exploratory to unstructured exploratory. Jörgensen (1999, p. 310) maintains this behaviour to be similar to the behaviour in traditional library environments. Batley investigates the search tactic in relation to different types of requests (specific, general or nameable, general or abstract needs, and subjective needs), and she establishes that browsing increases as the specificity of the search task at hand decreases. That is, for specific requests, users tend to employ keyword searching, while abstract or subjective information needs mostly lead to random browsing (Batley, 1988b, p. 378). Recently,
Hung (2005) follows up on Batley’s results, and she investigates the behaviour of five journalist students while they search for images in order to handle a task put forward by the researcher. Browsing is found to be the key move for image searching. The results indicate a relationship between the type of search task and the search tactic chosen. That is, the searchers employ more tactics and complex search moves when searching for general or abstract images content, compared to searching for nameable entities in images. This is in line with Vakkari’s model of the information search process in task performance (Vakkari, 1999), discussed in Section 3.2.5. Hung concludes that “[...] image retrieval systems should be designed accordingly to support different types of images” (Hung, 2005, p. 343).

Sandom and Enser (2002) investigate the search facilities preferred by 24 persons seeking moving images for professional purposes, by asking them to indicate whether specific search facilities are very useful, of some use, or of no use. The results are summarised in Table 3.3. They show that 79% of the 24 respondents find keywords very useful for searching, while numeric classification system were not considered to be very useful by any of the 24 respondents. Interestingly, 75% of the respondents do not consider key frames either very useful or of some use when searching for moving images.

Table 3.3: Preferred facilities for searching moving images. Adapted from Sandom and Enser (2002, p. 46).

<table>
<thead>
<tr>
<th>Search facility</th>
<th>Very useful</th>
<th>Some use</th>
<th>No use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keywords</td>
<td>79%</td>
<td>0%</td>
<td>21%</td>
</tr>
<tr>
<td>Natural language</td>
<td>58%</td>
<td>29%</td>
<td>13%</td>
</tr>
<tr>
<td>Classification system</td>
<td>0%</td>
<td>38%</td>
<td>63%</td>
</tr>
<tr>
<td>Boolean functionality</td>
<td>71%</td>
<td></td>
<td>29%</td>
</tr>
<tr>
<td>Selection of best match or exact match</td>
<td>50%</td>
<td>21%</td>
<td>29%</td>
</tr>
<tr>
<td>Relevance feedback</td>
<td>25%</td>
<td>21%</td>
<td>54%</td>
</tr>
<tr>
<td>Key frames</td>
<td>0%</td>
<td>25%</td>
<td>75%</td>
</tr>
</tbody>
</table>

Legend: Dark grey percentages are calculated on the basis of the figures given by Sandom and Enser, while light grey percentages are inferred from Sandom and Enser’s figures.

Generally, the search patterns presented above move along a continuum from focussed to explorative or a browsing type of searching. This continuum is comparable to the searching continuum employed in the Information Search Process model developed by Kuhlthau, and discussed in Section 3.2.3. Fidel (1997) relates the searching behaviour to the intended use of the image, and she derives a continuum of use from the Data Pole.
to the Objects Pole. At the Data Pole images are retrieved because of the information they embody, while at the Object Pole they are to be used as object e.g., in an advertisement. She maintains that the intended use of an images is not inherent in the images, but that the same images may be retrieved for different kinds of intended use (Fidel, 1997, pp. 189-190). In relation to the present work, searching behaviour is important in order to gather knowledge as to whether the behaviour could be satisfied in the process of indexing of television broadcasts.

### 3.4.3 Work with and use of television broadcasts

In this section, we outline and discuss literature on work with and use of television broadcasts.

Though empirical investigations on task based information seeking receive an increased attention in the research literature, only few studies have had a focus on the use of moving image archives in genuine work task. Markkula and Sormunen (2006) investigate information needs and searching behaviour of television professionals in the genuine course of television broadcast making. They interview 17 people involved in television production (journalists, image journalists, broadcast assistants, and information specialists) and derive a model of the common work process of television broadcast making. The model is an elaboration of Tan and Müller’s television broadcast production process model (Tan & Müller, 2002, pp. 51-53). The derived model is illustrated in Figure 3.6.

![Figure 3.6: Markkula and Sormunen’s stage model of work processes in television broadcasts making (Markkula & Sormunen, 2006, p. 195).](image-url)
Markkula and Sormunen’s model consist of six stages in television broadcast production: idea generation; planning; shooting; pre-selection; script writing; and editing. In the idea generation stage the person seeks to gather information in order to decide whether or not the programme should be made. In this stage the person gathers background knowledge about similar programmes. Depending on the programme type the person works on textual descriptions of broadcasts or the broadcasts themselves. At the planning stage an outline for the script is put down, widespread information gathering is conducted, and extensive explorative searching is carried out several times. The knowledge on the topic increases, and the information needs are getting focused. The shooting of new images is planned. In the shooting phase the planned shooting is carried out. New needs may arise during shooting, but no searching is carried out in this stage. Naturally, extensive video acquisition appears, due to the shooting of new moving images. In the pre-selection stage clips that might be applied in the final broadcasting are selected. The selected material is watched, and in this viewing process new needs may appear, and hence additional searching is carried out in this stage. In addition, the needs arisen in the preceding shooting stage are sought fulfilled. In the script writing stage the script for the final broadcasting is written based on the material previously selected. No new needs appear in this stage, since the urgent needs for additional material are identified in the previous pre-selection stage. At the editing stage the broadcasting is edited according to the prepared script. In this stage the broadcast as a whole is seen for the first time and new needs may transpire. The information needs in this stage are focused and they can be described in great details.

Markkula and Sormunen emphasise that “[t]he boundaries of the stages are fluid and different work tasks may sometimes be carried out concurrently” (Markkula & Sormunen, 2006, p. 195).

The stage model of television broadcast production is comparable to Kuhlthau’s model of the Information Search Process. In regard to feelings, Markkula and Sormunen verify Kuhlthau’s findings in that the “…subjects perceived a similar uncertainty regarding the goals and outcomes of the task at the early stages of the process as emphasized by the ISP [Information Search Process] model” (Markkula & Sormunen, 2006, p. 198).

In summary, the investigation shows that needs for moving images arise throughout the process of producing television broadcasts, but they gradually become more focused. Though moving images were sought both in the beginning and the end, they were typically acquired in the first three stages and their volume reduced in the last three stages.
Auffret and Bachimont propose a model for archiving television broadcasts in a library context. They focus on the hypermedia access to archived television and radio broadcasts, with specific focus on the French national audiovisual archive (Institut National de l’Audiovisuel). In our presentation and discussion, we focus on television broadcasts, only. The users considered are scholars, journalists, schoolteachers, and students, and it is consequently relevant for the present work. Auffret and Bachimont (1999, p. 62) emphasise five tasks that a user should be able to conduct in a broadcast retrieval system:

1) **Search for broadcasts**, by using efficient IR tools to search and retrieve broadcasts;
2) **Browse broadcasts**, where the user is granted direct access to play and/or skim the broadcasts in a non-linear fashion, using traditional VCR functions (e.g., play, forward, back), and new functions (e.g., key-frames as entry points);
3) **Navigate the broadcasts using metadata structure**, enabling the user to access the broadcasts at different levels of granularity, by using navigational tools such as table of content, as well as providing the user with summaries of the content of broadcast;
4) **Interpret broadcasts in context**, provide the user with contextual information about broadcasts, e.g., including critics’ articles, the right management file, the script, pictures taken during the shooting, the original non-edited shootings; and
5) **Annotate broadcasts**, facilitating the users’ with tools to annotate the broadcasts in their viewing and interpretation of broadcasts. The user should be able to use the annotations for searching and accessing the broadcasts, at a later point in time.

The point of departure for the model for archiving television broadcasts proposed by Auffret and Bachimont (1999, p. 63) is domain knowledge. Knowledge about the users and their behaviour is a prerequisite for the design and construction of an IR system that facilitates the five tasks listed above. The first task is similar to searching tasks in traditional textual IR systems. This requires the broadcasts to be indexed in a manner coherent to the users’ cognitive perceptions of the broadcasts. However, the approaches and tools for indexing is not necessarily the same for different types of information objects, which is one of the motivational factors underlying the present work. The second task is somewhat similar to the task of browsing printed documents, e.g., a
scientific article, but techniques for efficiently browsing moving images are different from the techniques used for browsing text. Techniques for browsing moving images are intensively being researched in the content-based research community (Smeaton, 2004). The techniques for browsing available in a broadcasts system should be based on knowledge about the users work tasks, and their specific need for browsing. The third task requires that an index or table of content is constructed for the archived broadcasts. Such an index should enable the user to access broadcast at a specific level of granularity (see e.g., Section 4.2.2 on the levels of granularity in a television broadcasts context). The level(s) of granularity that is/are of relevance for the users must be learned prior to the construction of the IR system. The contextual information that is relevant for users is again related to the work tasks, and the IR system designer needs to get familiar with the types of work tasks prior to identifying the contextual or external sources that are appropriate in a broadcast retrieval system. Lastly, the users needs for annotating broadcasts is associated to their work tasks, e.g., including the cooperation between users.

Auffret and Bachimont do not explicitly state their methodical approach for identifying the five behavioural tasks presented above, and we do not know how and/or to what extent users have been involved in the identification of such behaviour. Consequently, we are inclined to consider it to be based on the authors’ logical reasoning based on their experiences with users at the French national audiovisual archive. Whether or not the behavioural tasks are based on an empirical study of users, the observations by Auffret and Bachimont are interesting in relation to the present work, and the five behavioural tasks are discussed in relation to our results in Section 7.4, and Section 9.1.4.2.

Auffret and Prié (1999, pp. 322-323) focus on humanistic scholars’ use of moving images. In an analysis of the literature and through personal observations (which are not explained further), they identify four types of work tasks involving moving images:

1) **Pedagogical tools**: moving images act as examples in lectures or conferences;
2) **Testimony of the past**: moving images are used for historical analyses of the past, since broadcasts are considered mirrors of our society;
3) **Work of art**: moving images are considered a testimony of an author’s creativity, of his/her aesthetics, and/or thematically choices;
4) **Personal notes**: observational video recordings are part of the scholars methodical approach, e.g., for a posterior analyses of human behaviour; and
5) **Communication acts:** the moving images are considered as communication vehicles, and they are studied in order to reveal their communicative strategies etc.

Auffret and Prié (1999, pp. 319-320) emphasise that all five work tasks represent so-called active reading of the moving images, as opposed to the passive consumption of television broadcasts for leisure. Active reading concerns the thorough analysis of moving images, which is in line with the humanities' emphasis on human generated artefacts as the primary source of analysis (see e.g., Section 3.3.2). The scholar’s creation of his/her own moving images in the fourth type is not relevant in a television broadcasts contexts. Otherwise, the typology forms an interesting point of departure for identification and discussion of humanistic scholars’ work tasks in the context of television broadcasts.

This ends our section on information seeking behaviour in the context of still and moving images. In the following section, we put emphasis on one element of information seeking behaviour, namely relevance and in particular, the criteria applied when assessing the relevancy of information objects. This is important since it relates to research question 3 on the relevance criteria applied when seeking television broadcast.

### 3.5 Relevance

As introduced in Chapter 1, our work concerns the identification of appropriate access points for retrieval of television broadcasts in order to facilitate effective retrieval of television broadcasts. A central element in this identification is knowledge about the information which information seeker’s makes use of, or desires in order to assess the relevancy of retrieved information objects. By identifying the criteria or metadata elements users apply in order to evaluate whether or not a retrieved television broadcast is relevant for their task at hand, we can provide users with information in relation to such criteria. Consequently, users can make more qualified relevance assessments, and retrieval is hence more effective. Research question 3 is concerned with gathering such knowledge about users’ application of relevance criteria, and therefore the aim of the present section is to outline previous research on relevance criteria. In order to comprehend the complexities inherent in the notion of relevance criteria we find it important to describe our perception of the concept of relevance as such. Consequently, Section 3.5.1 discusses the concept of relevance in LIS. Section 3.5.2 concerns previous empirical investigations focussing on users’ application of relevance criteria
3.5.1 The concept of relevance

Relevance is the central concept in LIS, in the sense that it, implicitly or explicitly, functions as a fundamental element in evaluation of retrieval effectiveness (Schamber, 1994, p. 3). Relevance was considered troublesome at an early stage, but at the same time, it was considered a necessary concept. For instance, Doyle concludes that “[r]elevance is a thought-crutch; with it we may think inaccurately about the retrieval problem, but without it (or something better) we couldn’t think at all” (Doyle, 1963, p. 200). Later Saracevic states that “[…] not only information, but information characterized by its relevance became the key notion in Information Science. And the key headache” (Saracevic, 1996, p. 201).

In their review of the literature on relevance Eisenberg, and Schamber (1988) instigate what could be termed the new wave of relevance research. Concurrently with an increased focus on the concept of relevance, the concept is understood in more complex ways. This development should be seen in relation to a general movement from system-driven approaches towards user-oriented or holistic approaches, e.g., as exemplified by the development of the cognitive viewpoint. In corporation with Nilan, Schamber and Eisenberg follow up on their initial review, and they consider relevance judgments as users’ evaluations of information in relation to their information need situations at a particular point in time (Schamber, Eisenberg & Nilan, 1990, p. 771). Relevance depends on subjective judgments of the relationship between information and information need, and hence the criteria users apply for these judgments are the values, which determine the users’ perception of the quality of the relationship. We agree with Schamber, Eisenberg and Nilan that an understanding of relevance criteria as the reasons underlying relevance judgments is an important element in reaching a more complete and useful understanding of the concept of relevance (Schamber, Eisenberg & Nilan, 1990, p. 771). Further, an understanding of the factors that influence the users’ perception of relevance is important in order to recognise the practical concerns, which might be accounted for in design of IR systems. In summary, Schamber, Eisenberg and Nilan (1990, p. 774) understands relevance to be:

1) A multidimensional cognitive concept whose meaning is largely dependent on users’ perception of information and their own information need situation;
2) A potential dynamic concept that depends on users’ judgments of the quality of the relationship between information and information need at a certain point in time; and

3) A complex but systematic and measurable concept if approached conceptually and operationally from the user’s perspective.

The implicit understanding of relevance is for the concept to be multidimensional, potentially dynamic, and complex. The perception and use of relevance can differ between different relevance assessors, which, when put simple, refers to the multidimensionality of the concept. Relevance is potentially dynamic since a user’s perception of the concept can, but will not necessarily change over time. Taken together, the three conclusions define what is termed situational relevance.

In a recent re-examination of the literature on relevance, Borlund (2003a, p. 914) takes a point of departure in these three conclusions in order to demonstrate that a consistent and compatible perception of the relevance concept has been reached in LIS. The perception acknowledges that relevance can be approached from different points of views. Briefly, relevance is what the user says is relevant to the request, or what the IR system retrieves. The succeeding sub-section briefly outlines the different aspects of the concept of relevance including the multidimensional, potentially dynamic, and situational nature of the concept. This is followed by a section with emphasis on the relevance criteria applied by users, and special attention towards relevance criteria in the context of moving images.

3.5.1.1 Multidimensional and situational relevance

The many definitions and discussions of the concept of relevance proposed over the years have shown that the concept is many faceted. It is this many faceted nature of the concept which Borlund (2003a) comprise into one coherent conception of relevance. This includes classes, types, and degrees of relevance, as well as relevance criteria as complementary framed by Borlund (2000). In this section, we will describe these facets, respectively, though, for the purpose of the thesis, emphasis will be given to relevance criteria in the subsequent section. In addition to this multidimensional nature of the concept, this section discusses the notion of potentially dynamic and situational relevance.

Following Saracevic (1975), Swanson (1986), and Harter (1992), Borlund distinguish two main classes of relevance: “(1) objective or system-based relevance; and (2) subjective or human (user)-based relevance” (Borlund, 2003a, p. 914). Objective relevance is in principle assessed independent from the user. As stated by
Swanson “[t]he issue is not what the requester meant to ask but what the request itself actually said” (Swanson, 1986, p. 392). In contrast, in subjective relevance “[…] whatever the requester says is relevant is taken to be relevant; the requester is the final arbiter” (Swanson, 1986, p. 390). Objective relevance adheres to the system-driven approach to LIS, while subjective relevance adheres to user or cognitive-oriented approaches.

Borlund (2003a, pp. 914-915) further develops Saracevic’s five types or manifestations of relevance and sees them in relation to the two main classes. The five manifestations proposed by Saracevic are: 1) system or algorithmic relevance; 2) topical or subject relevance; 3) cognitive relevance or pertinence; 4) situational relevance or utility; and 5) motivational or affective relevance. Saracevic’s first type of relevance is related to the objective class, while the latter four are associated with the subjective class of relevance. Algorithmic relevance is concerned with a match between the content of the information objects and the topic (or rather content) of the query. Algorithmic relevance is measurable and different models may be applied to measure relevance, e.g., the Vector Space model (Salton & McGill, 1983). For that reason, Borlund propose that algorithmic relevance may be further separated into vector space relevance, probability relevance, or Boolean relevance. Algorithmic relevance is sometimes referred to as topical relevance (e.g., Harter, 1992, p. 602), and is considered to be “[…] the clearest and most persistent definition of relevance” (Eisenberg & Schamber, 1988, p. 164). A human assessor may also interpret the topical relation between a request and information objects intellectually. This is the second type of relevance proposed by Saracevic. To distinguish it from the algorithmic type of relevance we, in line with Borlund, denoted it intellectual topicality. Saracevic’s third type of relevance designates the relation between the user’s interpretation of the information need, and the retrieved information objects, hence referred to as pertinence relevance. Pertinence is closely related to situational relevance, which designates the relation between a user’s perception of the work task and the retrieved information objects. Motivational or affective relevance is not considered a specific type. Instead, this type is considered a characteristic of all of the subjective types of relevance. That is, “[…] the ‘drive’ to want information is not an independent, specific type of relevance, but an inherent characteristic of relevance behaviour in general” (Borlund, 2003a, p. 915). Borlund’s perception of Saracevic’s manifestations is illustrated in Figure 2.3.

While “[…] users’ relevance judgments exist on a continuum of relevance regions from highly relevant, through partially relevant to non-relevant” (Spink, Greisdorf & Bateman, 1998, p. 600), one has to decide on some scale for measuring relevance. We
follow Borlund’s terminology and use degrees of relevance to denote “[…] the rating and indication of the value of relevance of a given assessed information object” (Borlund, 2003a, p. 918). The use of degrees of relevance is independent of the classes or types of relevance employed. The degrees applied, range from everything between a binary scale and a continuous scale. In system-driven evaluation of IR systems, the most common has been the application of binary relevance (relevant or non-relevant). Binary relevance is highly favoured because of its strong tie to the performance measures of recall and precision (and associated measures, e.g., mean average precision). In user or cognitive oriented evaluations, researchers tend to use non-binary relevance (three or more degrees), but in order to apply recall and precision for analysing the data, non-binary relevance assessments are merged into binary relevance. The underlying assumption being that no information is lost by merging the categories (Schamber, 1994, p. 18).

The choice of scale was tested in the early relevance studies (e.g., Cuadra & Katter, 1967; Rees & Schultz, 1967), and in a follow up study by Janes (1993). Janes verifies that users’ assessments tend to fall in the two ends of the assessment scale. Nonetheless, Tang and colleagues (Tang, Shaw & Vevea, 1999) find that the user’s confidence with relevance judgments are highest when a 7-point scale is applied. At the same time they emphasise that no one scale is appropriate under all circumstances. With specific emphasis on partial relevance, Spink, Greisdorf, and Bateman (1998, p. 599) identify differences between the different degrees of relevance, and suggest that partially relevant information object may be important in the early stages of a user’s search process. In this way, relevance is linked to the stages of Kuhlthau’s model of the Information Search Process. This link may be discussed in terms of different levels of relevance assessments. That is, at the early stages the user makes topical relevance judgments, while at the post focus stages ‘pertinent’ information is sought (Borlund, 2003a, p. 920). Maglaughlin and Sonnewald (2002) investigate the relevance criteria applied in three categories of relevance (relevant – partially relevant – non-relevant), and they find that different relevance criteria are applied in the different categories of relevance. Hence, information would be lost if the categories of relevant and partially relevant were to be merged.

The dynamic nature of relevance refers to the fact that the user’s judgment may change over time. This dynamism is associated with the developing nature of the user’s perception of the underlying work task, as well as the developing nature of the user’s perception of the information need, which we have discussed previously in relation to the cognitive viewpoint and the maturation of the information need. The dynamism is also an element in Kuhlthau’s model of the Information Search Process since the user
may potentially learn throughout the search process. Tang and Solomon (1998) empirically show that one person’s assessment of relevance may be different at different points in time, and hence they demonstrate the potentially dynamic nature of relevance. The dynamic relation between relevance and the user’s perception in an information seeking situation is illustrated in Figure 2.3 and related to the notion of situational relevance.

An information object is considered to be situational relevant if it causes a change in the person’s perception of his or her situation (Borlund, 2003a). The change in the person’s knowledge structure may be caused by the topic of the information object or the potential use of the information object. This definition of changing knowledge structures is in line with the cognitive viewpoint conception of information. This is, for instance, related to the fifth central and interrelated dimension of the viewpoint, proposed by Ingwersen and Järvelin (2005, pp. 25-28), namely that information is situational and contextual (see e.g., Section 2.2). However, Wilson does not consider situational relevance to be of a psychological nature, but rather to be of a logical construct. Borlund ascribes this to the fact that Wilson may have tried to avoid the intractability of a subjective viewpoint (Borlund, 2003a, p. 922). To evade this problem, Borlund proposes the application of a simulated work task situation for evaluation of interactive IR systems. The simulated work task situation functions as the trigger of the information need of the user. In Borlund’s definition situational relevance “[…] expresses the relationship between the user’s perception of usefulness of a retrieved information object, and a specific work task situation” (Borlund, 2003a, p. 922). In Figure 2.3 the relationship between situational relevance and the information need development is depicted. Situational relevance of an information object is judged against the user’s perception of the work task. Therefore, if the user’s perception of the work task changes, and his or her cognitive structures change, e.g., due to learning, so will the user’s assessment of the relevancy of an information object change. This effect of a new perception of the situational context on assessment of relevance is depicted with dotted arrows in Figure 2.3. This said, it is important to stress that the new assessment may, but do not necessarily, result in an altered outcome of the assessment. That is, although the premises for conducting a relevance judgment have change, an information object may still be relevant or not for handling a specific work task.

The perception of relevance to be multidimensional, dynamic, and situational fits well within the cognitive viewpoint e.g., the notion of the user’s dynamic and situational perception of the underlying work task is directly related to situational relevance. The following section puts emphasis on the criteria users apply when assessing the relevancy of information objects.
3.5.2 Relevance criteria

Relevance criteria refer to the criteria applied for judging the relevancy of an information object, and hence due to human involvement the criteria may be rather complex for the subjective class of relevance. In the traditional system-oriented approach to IR, based on the objective class of relevance, the sole criterion is the match between the query and the document representation. With more focus on the user, and the consideration of a subjective class of relevance, the relevance criteria become much more complex. Two empirical studies conducted concurrently in the 1960s were prominent in the investigation of criteria related to subjective relevance. These studies by Cuadra and Katter (1967), and Rees and Schultz (1967) show the complexity of criteria applied by users in subjective relevance assessments. A characteristic of the early investigations of relevance criteria is that they do not focus on real users in genuine situations, retrieving information objects from actual collections. This lack of realism in the investigation motivated Schamber, Eisenberg and Nilan (1990, p. 771) to call for more realism in studies on relevance criteria. Subsequently, a new wave of investigations on users’ relevance criteria has seen the light of day, investigations in which realism is a main concern. Below, we briefly discuss the results obtained in some of these investigations (Schamber, 1991b; Park, 1992; Barry, 1993; Wang, 1994), which put emphasis on different information sources as well as different types of users and work contexts.

Schamber focuses on real users’ work tasks concerned with retrieval of weather information in a multimedia environment. The users are professionals working in aviation, electric power utilities, and construction, and the information sources are diverse, including weather information systems, people, newspapers, and television. Schamber employs Dervin’s (1983b) micro-moment time-line interview technique, and interviews 30 professionals about previously occurred information-searching events with the purpose of identifying the relevance criteria the professionals applied. The time-line interview technique relies solely upon the respondents’ recollection of the situation in question. Schamber identifies 10 criteria and groups them in 3 categories, according to what the evaluation seems to focus on: 1) sources (reliability, verifiability, accessibility,); 2) presentation (presentation quality, dynamism, clarity); and 3) information or content (currency, geographic proximity, specificity, accuracy). The investigation validated to a large extend earlier studies in textual contexts (e.g., Taylor, 1986, p. 50), and in addition the criteria were mentioned regardless of the source of information, which indicates that a fixed amount of criteria may be identified across contexts and sources. However, the investigation also identified criteria (e.g.,
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

gеographic proximity) that are related to the specific context under investigation, which suggests that relevance criteria are related to the situational context. Further, it indicates that the importance of the criteria vary across information sources (Schamber, 1991a, p. 131). Schamber discusses the implications of her results for design of IR systems, e.g., the focus on specific sources in different contexts, but implementation of the results in systems design is left to others.

Park investigates the criteria applied by ten academic users in genuine searching situations. Relevance is judged on the basis of bibliographic references, and Park subsequently, identifies four categories within two dimensions, termed citation-based and user-based. The citation-based dimension comprises one category, namely interpretation of a citation or bibliographic record (e.g., title, author name, journal name, and abstract). The user-based dimension is concerned with the user’s experience, perception, and knowledge of the information problem and it comprises three categories: 1) internal (experience) context mirror the user’s interpretation of a bibliographic record based on the user’s own experience in the problem area; 2) external (search) context reflects factors that are particular to the search, including purpose of the search; and 3) problem (content) context which includes the user’s understanding of the problem as well as the motivation for conducting the search in the first place (Park, 1992, pp. 87-91; 1993, pp. 329-330). The two dimensions are interconnected, and the user-based dimension acts as the basis for the interpretations of citations in the citation-based dimension. Park shows that cognitive and psychological changes may occur for the user during the process of assessing relevance. The user’s cognitive structures may change, and the user’s interpretations can consequently be affected by such a change in cognitive structures. With a focus on the citation-based dimension Park indicates the relation between access points in a bibliographic record and relevance criteria, e.g., that publication date is important in relation to the relevance criteria of perceived benefit.

Similar to Park, Barry studies the criteria applied by 18 academic users, each judging 15 bibliographic records and 3 full texts. The bibliographic records and full texts are based on searching a genuine request handed to the researcher prior to the assessments. Barry identifies 23 criteria and classifies the criteria according to seven categories: 1) information content of the document; 2) user’s background/experience; 3) user’s belief and preferences; 4) other information and sources within the environment; 5) sources of the document; 6) document as a physical entity; and 7) the user’s situation (Barry, 1994, p. 157). The overlap between the different studies is identified by Barry as she states that “[…] it does seem apparent that there is a great deal of consensus in terms of the general areas of the information seeking process that have been suggested to include factors that will affect users’ judgments of relevance” (Barry, 1993, p. 47).
Chapter 3: Information seeking behaviour

Barry and Schamber explicitly compare their results and though a certain overlap is identified between the two studies they conclude that different criteria will be found when studying different types of information objects, users, and situational contexts (Barry & Schamber, 1998, p. 234).

In line with Hertzum and colleagues (Hertzum et al., 2002), Borlund divides relevance criteria into efficiency related criteria (e.g., cost and quality) and effectiveness related criteria. Efficiency related criteria are not directly related to the user’s information need, and she consequently maintains that effectiveness related criteria should be the attention of future research (Borlund, 2003a, p. 918).

Wang and Soergel (Wang & Soergel, 1993; 1998; Wang, 1994; 1997a; 1997b) develops a model of the document selection process. In this development, they employ personal interviews and think aloud to investigate 25 academics in agriculture. The model of the document selection process puts emphasis on each component of the process as well as the relation between the components. Like Schamber (1994), Park (1992; 1993), and Barry (1993), Wang and Soergel identify several relevance criteria, and they explicitly translated the identified criteria to access points or so-called document information elements. They identify the importance of different access points, as well as point to the specific access points that are potentially important when handling information needs. The identification of appropriate access points for design of IR systems is comparable to Pejtersen’s (1989) work on the Book House project presented earlier. The primary focus for Wang and Soergel is not as much construction of new access points, but rather development of a model for the document selection process. Therefore, Wang and Soergel focus on scientific literature, and primarily on access points that are already present in bibliographic records. In contrast, Pejtersen’s research is directed toward analysis of deficiencies in present surrogate records, and consequently the development of more effective access points for imaginative literature.

Wang and Soergel relate the identified relevance criteria to data elements in surrogate records. In this way, their results are relevant for the present work, despite the difference in focus. That is, Wang and Soergel focus on retrieval of textual documents from an operating IR system, whereas we focus on retrieval of television broadcasts in a future IR system. The following sub-section discusses research on relevance criteria in the context of moving images.

3.5.2.1 Relevance criteria and moving images

The first study with a primary focus on relevance in the context of moving images is, Enser and Sandom’s aforementioned investigation of users’ behaviour when requesting moving images from 11 operating archives in the UK (Enser & Sandom, 2001; 2002;
Sandom & Enser, 2002). The study has a minor focus on the criteria or metadata elements applied by users when assessing the relevancy of moving images. This part of their study is based on a questionnaire filled in by information seekers, who have a professional need for retrieval of moving images. As a part of the questionnaire, Enser and Sandom compiled a list of 19 data elements, and asked the participants to indicate the usefulness of each element for judging the relevancy of the information objects. Their results are depicted in Table 3.4.

**Table 3.4:** Relevance criteria for moving images identified. Modified from Sandom and Enser (2002, p. 46) by including percentages.

<table>
<thead>
<tr>
<th>Data element</th>
<th>Very useful</th>
<th>Some use</th>
<th>No Use</th>
<th>No response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Copyright information</td>
<td>20</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Summary</td>
<td>20</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Shotlist</td>
<td>19</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Colour/black &amp; white</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Indication of quality</td>
<td>14</td>
<td>7</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Keywords</td>
<td>14</td>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Title</td>
<td>12</td>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Silent/sound</td>
<td>11</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Duration/footage</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Location</td>
<td>9</td>
<td>10</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Medium</td>
<td>9</td>
<td>10</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Format</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Genre</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Language</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Clip</td>
<td>4</td>
<td>13</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Thumbnails/key frames</td>
<td>3</td>
<td>11</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Director/maker</td>
<td>1</td>
<td>10</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Rating</td>
<td>1</td>
<td>6</td>
<td>11</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3.4 shows that date, copyright information, summary and shotlists are the most popular data elements for judging the relevancy of moving images. Interestingly, only a small part of the participants finds the visual elements (thumbnails or clips) very useful. This is in line with the results for the users’ preferred search facilities (see Table 3.3). At the time of the investigation, none of the 11 archives provided thumbnails or clips.
for retrieval, and participants may hence have had difficulties envisioning the usefulness of such visual retrieval facilities.

Recently, Yang specifically focuses on users’ relevance criteria when searching for moving images. In a small pilot study Yang and colleagues investigate four participants who regularly search for video as part of their professional work task, by employing Schamber’s variation of time-line interviews (Yang, Marchionini & Wildemuth, 2004). Similar to previous studies (e.g., Peijtersen, 1989; Wang & Soergel, 1998) they find a relation between users requests and aspects of the information objects searched for. The identified relevance criteria are related to elements in the surrogate records the participants retrieved in the search processes. They divide the identified relevance criteria into 3 types: 1) criteria which relates to textual elements of the surrogate records; 2) criteria which relates to visual elements; and 3) implicit criteria which relates to “[…] users’ own experience or task” (Yang, Marchionini & Wildemuth, 2004, p. 232). As found in the earlier studies ‘topicality’ is the most prominent relevance criterion, but relevance criteria relating to visual elements are also found to be important when searching for video, which is in line with results from studies of relevance criteria in the context of images (e.g., Choi & Rasmussen, 2002). Yang follows up on this pilot study by interviewing 26 participants, spanning a range of occupations from academia, production companies, and libraries. The follow-up study generally verifies the results from the pilot study and in addition, Yang demonstrates that the relevance criteria applied by users is related to the work task at hand. When compared to relevance criteria found in the textual relevance literature, Yang’s results show that the diversity of criteria increase for moving images, and although topicality is found to be the most important relevance criteria, it seem to have less importance than when assessing textual information objects (Yang, 2005, pp. 161-166). Yang also shows that users are interested in access to moving images at the scene or feature levels of granularity (e.g., see our definition of levels of granularity in Section 4.2.2). This is comparable to Enser and Sandom’s results that users perceive shotlists very useful for relevance judgments.

Enser and Sandom, as well as Yang, discuss their results in relation to either human intellectual or automatic indexing of moving images, and they conclude that the most effective retrieval would involve a combination of intellectual and automatic approaches for indexing. The different approaches for indexing moving images are discussed in the following chapter. Before that, the final section provides a summary of the present chapter.
3.6 Summary

Information seeking behaviour is a broad concept, which encompasses a person’s behaviour when seeking information in the course of striving towards the satisfaction of a goal, and hence it includes a range of different topics. The instigation of any purposeful information seeking behaviour is an individual’s aspiration for solving a work task, whether leisure or job related. The individual’s recognition of some inadequacies in order to solve the task, gives rise to a perception of a need for information. The information need may subsequently be presented for intermediaries in relation to a seeking activity. The information need is a cognitive, psychological and personal construct, which is necessarily connected to a situational and environmental context at a particular point in time. This conceptualisation of the information need is in line with the five central features of the cognitive viewpoint discussed in Chapter 2, and it is an important underlying assumption for the present chapter. Both in relation to the general models and the more specific information behaviours discussed in this chapter.

The theorising and modelling of information seeking behaviour by Dervin (e.g., 1983b), Ellis (e.g., 1987), Kuhlthau (e.g., 1991), Byström and Järvelin (e.g., 1995), and Vakkari (e.g., 1999), outlined in this chapter, shows the importance of work tasks as the instigating factor of any information seeking activity. In addition, these general models show that information seeking is highly complex, and that it is related to situational and environmental context, as well as the information sought. Our emphasis is on behaviour of academics in Media Studies when seeking television broadcasts, and consequently the chapter discusses previous work, which emphasises seeking behaviour in the humanities as well as seeking behaviour in relation to (moving) images. The previous work indicates that humanistic researchers are familiar with the research area in which they seek information and that even though they tend to conduct their research alone they do use their colleagues in the process of seeking information. Further, an important aspect of the humanistic researchers is their need for information objects to serve as their objects of analysis, which makes them dependent on good retrieval facilities, and not least retrieval facilities which may convey objects for analysis which have not been under scrutiny in the field.

In the final part of this chapter, we discuss one specific element of information seeking behaviour, namely assessing the relevancy of an information object. Our conception of the concept of relevance is strongly related to our conception of the information need. When the need is a cognitive, psychological and personal construct, interconnected with a specific context at a particular point in time, so becomes assessing
the relevancy of information for fulfilling the need, and hence the underlying work task situation. With focus on the specific context of seeking (moving) images, the visual aspect of the images should be considered when designing moving image retrieval systems.

Knowledge about users’ information seeking behaviour is important because it provides clues as of how users behave when seeking information, and hence how retrieval systems might help users cope with their information need. The first research area, and the three first research questions in the present work are concerned with gathering knowledge about three aspects of users’ behaviour when seeking television broadcasts: 1) information needs; 2) preferred search entries; and 3) relevance criteria, and hence this chapter on previous knowledge about information seeking behaviour is important for our work. Knowledge on the users’ information seeking behaviour is a prerequisite for discussion of access points and the value of information in external sources for indexing of television broadcasts, which is the focus in the fourth and fifth research question, as well as the following chapter. That is, Chapter 4 outlines earlier research on indexing of (moving) images.
Chapter 4: Construction of access points in surrogate records: indexing

The purpose of this chapter is to outline approaches for indexing of moving images. This outline is important as a mean to discuss future construction of appropriate access points for indexing and retrieval of television broadcasts. The emphasis is on the activity of indexing, and on the use of external sources for descriptions of primary information objects. In this way, the chapter is related to the second research area on appropriate access points for indexing and retrieval of television broadcasts, and of relevance to the third research area on construction of surrogate records for television broadcasts.

The aim of indexing is to make information objects available through structured searchable indexes. Indexing is hence the activity, which results in the generation of a representation of an information object in a surrogate record. Indexing systems “[…] represent attempts to extend the organizing capabilities of the human mind to […] artificial (humanly devised) information storage and communication systems. Indexing systems turn information storage and communication systems into information retrieval systems analogous, in widely varying degrees, to the human information retrieval system of the mind” (Anderson, 1985, pp. 287-288). Indexing is related to searching in that it essentially addresses the opposite side of the coin. Users are naturally only capable of utilizing or interacting with information, which is made available for them in the first place. Indexing provides the access points, which enable a user to search, retrieve, and evaluate the relevancy of information objects. Indexing is always, explicitly or implicitly, conducted from a particular point of view, which implies a given behaviour or access by users (e.g., Hjørland & Nielsen, 2001, pp. 249-250; Lancaster, 2003, pp. 8-9; Ingwersen & Järvelin, 2005, p. 36). The problem then is to ensure a correlation between the behaviour implied or rather facilitated in the IR system through indexing, and the users’ behaviour when seeking information. Our work investigates three aspects of seeking behaviour of users (information need, search entry preferences and application of relevance criteria), prior to discussion of approaches for future indexing and retrieval of television broadcasts.

The chapter comprises four sections. The 1st section presents and discusses previous work on traditional textual indexing, and commences with a description of how we define and use the term indexing in the present work. Despite our focus on television broadcasts, the point of departure is textual information objects. The reason
for this point of departure is that indexing predominantly has been approached from a textual context, and we maintain that the basic issues are akin whether the information objects are of a textual or non-textual nature. However, the specific nature of non-textual information objects (e.g., moving images) brings about additional complications in relation to the process of indexing (Borgman, 2000, p. 75). Consequently, Section 4.2 focuses on the specific characteristics of the information objects under scrutiny, namely television broadcasts. The 2nd section includes a review of literature about the manual, concept-based and the automatic, content-based approaches for indexing moving images. The 3rd section discusses earlier work on external sources for indexing of primary information objects. Finally, in the 4th section we summarise the chapter, and conclude the theoretical part of the thesis with a brief summary.

4.1 The process of indexing

In line with Anderson, we define indexing to be the processes and activities of identifying and describing “[…] content and other document characteristics (authors, titles, publishers)” (Anderson, 1985, pp. 288-289). Indexing is concerned with identification of both tangible as well as intellectual characteristics of the information objects in question, and may consequently be divided into descriptive indexing and subject indexing (Soergel, 1985, pp. 63-64; Rafferty & Hidderley, 2005, p. 1).

What we have termed descriptive indexing is concerned with formal data intrinsic to the information object in question. Ingwersen (1992, p. 53) makes an explicit distinction between formal and topical representations. The formal representation includes bibliographic data and is comparable to ‘other document characteristics’ in Anderson’s definition, and what is denoted descriptive indexing. In Ingwersen’s terminology the formal description is concerned with the isness of the information object, and represents common media-dependent formal (bibliographic) data such as author name and affiliation, title, publication date in a traditional textual context (Ingwersen & Järvelin, 2005, p. 271). Examples of isness in a television broadcast context are title, participants, and duration. Due to the formal bibliographic aspect, descriptive indexing is also referred to as descriptive cataloguing in the research literature (e.g., Rafferty & Hidderley, 2005, p. 1).

Subject indexing is the more complex of the two, and the one receiving most attention in the literature (e.g., Hjørland & Nielsen, 2001; Lancaster, 2003). Subject indexing is concerned with creation of representations of content or intellectual characteristics of information objects, or topical representations in Ingwersen’s terminology (Ingwersen, 1992, p. 53). The content or intellectual characteristics are
also designated aboutness, and in the subsequent section, the concept of aboutness is discussed further. Here we will not attempt a philosophical discussion of the concept, but refer to previous literature on the matter (e.g., Wilson, 1968; Maron, 1977; Hutchins, 1978; Langridge, 1989; Hjørland, 1992).

A clear-cut functional division between the two types of indexing (descriptive and subject indexing) does not exist in an electronic environment. For instance, the title of a scientific article may contribute to both descriptive and subject indexing, at the same time, since all words in the title may be searched by the user (Hjørland & Nielsen, 2001, p. 251).

4.1.1 The concept of aboutness

In line with our statement that indexing implies a given user behaviour, the aboutness of an information object can be approached from different perspectives. Inspired by Ingwersen (2002, p. 289) and Ingwersen and Järvelin (2005, pp. 281-291) we distinguish between three types of aboutness:

1) **Author aboutness**: the content as it is;
2) **Indexer aboutness**: the interpretation of content (subject matter) with a purpose; and
3) **User aboutness**: user interpretation of objects.

Author aboutness is concerned with the content of the information objects as it is expressed by the author in the information object. Author aboutness is associated with the content bearing units of the information object, and these units may hence represent the aboutness of an information object. An example is representation of textual documents by single terms or phrases extracted from the documents themselves (Ingwersen, 1992, p. 50). In this way, author aboutness might also be denoted document aboutness. Author aboutness is an immanent aspect of the information object, and it is hence objectively identifiable. Further, the process of indexing may be readily automated by the design and construction of algorithms that identify and extract the content bearing units from the information objects.

Indexer aboutness implies representations to be based on analysis of the content of the information objects. The processing or interpretation of the content bearing units is what distinguishes indexer aboutness from author aboutness. The foundation of the interpretation or processing is the cognitive structures of the indexing device, whether the indexing device conducts human intellectual interpretation or algorithmic computational processing. Automatic computational devices may for instance process
the content bearing units, and extract keywords from a list of controlled vocabulary or a thesaurus (Ingwersen & Järvelin, 2005, p. 384).

User aboutness is the perception of the content of information objects seen from the user’s perspective, and it may be divided into two kinds depending on the perceiver being an indexing device (human or computer) or the users themselves. The former aims at shaping the surrogate records at the potential future use of the information objects. This aim is reached by gathering extensive knowledge about the domain and the users therein. An example of this kind of user aboutness is Pejtersen’s previously mentioned work on fiction indexing in the Book House project (e.g., Pejtersen, 1989). The latter kind of user aboutness is concerned with the user’s perception of the content of retrieved information objects, and it is comparable to the concept of relevance outlined in Section 3.5. We use the term user aboutness in the first sense, only.

The three types of aboutness range from extraction of content from the information objects as such, to interpretation of the users’ perception of the information objects. This is comparable to Soergel’s distinction between entity (or document) oriented indexing and request (or information need) oriented indexing (Soergel, 1985, pp. 226-227). In document oriented indexing the point of departure is the content bearing units of the information object being indexed. It is evidently closely related to author aboutness, and it produces static indexing of the information objects in question. The point of departure in information need oriented indexing is the users’ information needs, e.g., as they are represented in a collection of written request. Information need oriented indexing requires interpretation of the content of the information objects, interpretations that traditionally have been conducted by humans. It is directly related to user aboutness, and it provides a dynamic type of indexing.

Table 4.1 summarises our definition and use of types of aboutness in the present work, with reference to Soergel’s two kinds of indexing, and whether the content bearing units in the information object are analysed or not. The distinction between the three types of aboutness is not as unambiguous as Table 4.1 might indicate.

Table 4.1: Types of aboutness in relation to Soergel’s (1985, pp. 226-227) two kinds of indexing, and whether the information objects are being analysed or not.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document-oriented</td>
<td>Author aboutness</td>
<td>Indexer aboutness</td>
</tr>
<tr>
<td>Information need-oriented</td>
<td>-</td>
<td>User aboutness</td>
</tr>
</tbody>
</table>
For instance, the thesaurus or list of controlled vocabulary used in indexer aboutness may very well be constructed on the basis of knowledge about the domain, its users, and their information needs. From a cognitive perspective, one cannot argue that applying either one of the three conceptions of aboutness would yield the most effective retrieval. Instead, one should always choose the approach that is most advantageous for the particular collection and user group in question. Indeed, since the three types of aboutness are not mutually exclusive, a combination of different types might be advantageous from a cognitive polyrepresentative perspective.

The types of aboutness appropriate for indexing of television broadcasts in the context of academics in Media Studies is briefly touched on in Chapter 0.

In the following section, we briefly discuss different types of information content, and how they affect the process of indexing.

4.1.2 Indexing of different types of information content

Lancaster (2003, pp. 200ff) distinguishes between substantive and imaginative works. Substantive works are non-fiction, e.g., textbooks, scientific journal articles, and television news. Substantive works have an educational and enlightening nature and their prime function is to provide information that pertains to a given problem. In contrast, imaginative works are fiction, e.g., poetry, drama, and television drama serials, which general aim is to entertain. The research literature primarily focuses on indexing of substantive information objects, though imaginative information objects are receiving increased attention (e.g., Pejtersen, 1979; Elias & Cawkell, 1997; Nielsen, 1997; Albrechtsen, Pejtersen & Cleal, 2002).

Lancaster’s distinction between types of content is comparable to linguistics’ perception of levels of interpretative freedom associated with texts. Barthes (1990) differentiates between readerly and writerly texts, and similarly, Eco (1979) differentiates between closed and opened texts. Barthes and Eco share the perception that different types of texts induce different levels of interpretative freedom. Writerly or closed texts tend towards one objective interpretation, and hence none or little freedom in interpretation, whereas readerly or opened texts imply large freedom in interpretation, and hence a wide variety of perceptions. Both Barthes and Eco sketch the two extremes along a continuum of interpretative freedom, and genuine texts will naturally take a continuous range of levels between the two extremes.

With reference to Lancaster’s distinction between substantive and imaginative information objects, substantive information objects are placed at the writerly or closed end of the spectrum, while imaginative information objects are placed at the readerly or opened end. That is, imaginative works tend to impose more interpretation from the
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

Indexer than substantive works. At this point, it is important to note that we do not attempt to express an eternized deterministic law, but rather sketch a general conception with several exceptions. For instance, Wiberley (1983) finds that the imprecision in humanists vocabulary (authors as well as searchers) complicates subject access in the humanities. When compared to writings in science, humanities writings tend towards freedom in subjective interpretation, though still being substantive in nature.

Barthes (1990) and Eco (1979) are both linguistics and they explicitly focus on textual information objects. However, their perception of levels of interpretative freedom is applicable to non-textual information objects as well, and it is indirectly implied in Fidel's (1997) distinction between Data pole and Object pole (mentioned in Section 3.4.2), and Greisdorf and O'Connor's (2002) distinction between high and low levels of comprehension in images.

The old saying 'an image is worth a thousand words' indicates the complex nature of images in indexing and retrieval contexts. According to Santini and Jain (1997, pp. 1-2), the complexity of (moving) images is due to the fact that (moving) images contain raw unstructured data, only. In contrast, the authors logically structure the raw data in textual documents (e.g., the ASCII code characters). Therefore, unlike a textual context, the most basic activity in indexing of (moving) images is concerned with processing of primitive features, e.g., the distribution of colours in (moving) images. One could speculate whether the low-level of consistencies of image inter-indexing (e.g., Markey, 1984; Turner, 1990; 1994; Enser, 1995), compared to the inter-indexer consistencies for texts (e.g., Hooper, 1965; Lancaster, 2003) are connected to the complexity of images, and the level of individual interpretation needed for processing such information objects. This said, the level of interpretation needed in processing (moving) images differs naturally from information object to information object. An example is the interpretation needed for processing a news feature on the Iraq war (where most content information is provided in the spoken words) compared to the interpretation of the motion picture 'The Name of the Rose'.

The level of interpretive freedom is important because it indirectly implies the level of subjective interpretation required in the process of indexing. Similar to the fact that different types of genres of academic texts (see e.g., Swales, 1990; Bazerman, 1994) entails different levels of interpretative freedom, different types of television broadcasts entails different levels of interpretation. This naturally affects the approaches for indexing which are applicable in a given context, and this is briefly discussed in Chapter 9.

Moving images differs from most other information objects, in that, they consist of different modalities or channels (Snoek & Worrning, 2005). That is, visual, audio, and
textual channels are simultaneously employed to comprise a whole. The multimodal, spatial and temporal nature of moving images adds a further level to the complexity of these information objects (Yang, 2005, p. 14). This is noticeable in cognitive film theories (e.g., Grodal, 1997; Ponech, 1997). For instance exemplified in Grodal’s (1997) conception of stages of cognitive and emotional perception of visual fiction. Different approaches and processing devices can be employed to analyse these channels, and a range of different methods and purposes for processing each channel exists. The approaches for indexing (moving) images are further outlined in the following section.

4.2 Approaches for indexing of moving images

Cawkell (1992) and Persson (2005) present each a citation analysis exploring the field of image indexing. Both analyses clearly distinguish between two groups of researchers which are identical to the two basic approaches to image indexing identified in the review literature (e.g., Rasmussen, 1997; Eakins & Graham, 1999). These two approaches are the manual or concept-based approach, and the automatic or content-based approach. The former has its roots in LIS, while the latter steams from technological developments in Computer Science. This section briefly outlines both of these approaches for indexing and retrieval of (moving) images.

The section comprises two sub-sections. Section 4.2.1 discusses the concept-based approach for indexing, and Section 4.2.2 reviews literature on the content-based approach. Approaches for indexing of moving images are considered a further development of approaches for indexing of still images (Smeaton, 2004, p. 374), and consequently the section focuses on still as well as moving images.

4.2.1 The concept-based approach

The concept-based approach is derived from practitioners, and represents the traditional approach for indexing and retrieval in operating (moving) image archives (e.g., Seloff, 1990; Enser, 1993; Green & Klasén, 1993; Terris, 1996). The concept-based approach is characterised by manual intellectual interpretation of the content of (moving) images. Representations of (moving) images almost solely contain textual surrogate records in the concept-based approach. However, manual interpretations may also result in visual representations, e.g., by applying representative pictures instead of representative keywords as exemplified by Rorvig and colleagues (Rorvig, Turner & Moncada, 1999). With reference to the types of aboutness discussed above, the typical approaches by practitioners are comparable to indexer or user aboutness, e.g., by applying thesauri or
other kinds of controlled vocabulary (e.g., TELCLAS (Evans, 1987) for indexing of television broadcasts). Another example is the so-called democratic indexing suggested by Brown and colleagues (Brown et al., 1996), and later applied by Hidderley and Rafferty (1997), which hand the process of indexing to the ones who use the information objects (see e.g., Library Thing for an operational example of democratic indexing).

Means for interpreting the meaning of an image is an important part of the concept-based approach, and it has been widely discussed in the literature (e.g. Rasmussen, 1997). Panofsky (1972, pp. 5-8) studies Renaissance art and suggests that art images comprises three strata or levels of perception: 1) pre-iconographical; 2) iconographical; and 3) iconological. The three levels signify an increasing degree of abstraction in the human perception of the images, and hence increased freedom in interpretation. At the pre-iconographic stratum, the factual or expressional meaning of the images is perceived, including objects, actions, and expressions. Panofsky’s example is a man lifting his hat with a friendly smile. Pre-iconographic perception is also denoted primary or natural subject matter. At the iconographical stratum, the lifting of the hat is perceived as a greeting. Greeting by lifting ones hat is peculiar to the western civilization and knowledge about a culture and its traditions are hence a prerequisite for the iconographic stratum of perception. Iconographic perception is also denoted secondary or conventional subject matter. At the iconological stratum, previous knowledge is applied in order to interpret the perceptions from the previous strata. For instance, based on ones knowledge about classes and clothing in western civilization in the 1890s, the greeting man is believed to belong to the upper class of society. Iconological perception is also denoted intrinsic meaning or content. In summary, the pre-iconographic stratum signifies description, the iconographic stratum signifies analysis, and the iconological stratum signifies interpretation.

Markey (1983), and later Shatford (Shatford, 1986; Shatford Layne, 1994) extends Panofsky’s theory on perception of art images for indexing in the broader context of images as such. Shatford (1986, p. 45) disregards Panofsky’s iconological stratum since it requires indexing to be too subjective in nature, and she suggests that the meaning of an image is comprised by its ofness and its aboutness. Ofness denotes the factual and concrete aspects of the image, in contrast to ‘aboutness’, which denotes the abstract semantics of the image. In order to avoid misunderstandings, we apply the term abstract to denote Shatford’s conception of aboutness in the remainder of our work. Shatford further divides ofness into specific and generic ofness. Specific ofness refers to the particular, factual or concrete aspects depicted in the image (e.g., Tower

http://www.librarything.com/
Bridge), while generic ofness refers to the common or broad aspects depicted (e.g., a bridge). Shatford combines her three aspects of perception of images (specific ofness, generic ofness, and abstract) with Ranganathan’s (1962) facets of who, what, where, and when, and she develops a 3x4 matrix, which identifies the kinds of descriptions that are relevant when indexing images. The matrix is latter applied by several researchers for analysis of (moving) image archives (e.g., Armitage & Enser, 1996; 1997; Sandom & Enser, 2002; Westman & Oittinen, 2006) and it is depicted in Table 3.2. The distinction between specific, generic and abstract content is introduced for still images, however, we see no problem in employing it to moving images, as shown by the application of the 3x4 matrix in different contexts (e.g., Sandom & Enser, 2002). With specific focus on television broadcasts, Green and Klasén (1993) describe the indexing process at an operating television archive in Sweden, and despite the terms are never used by Green and Klasén, a distinction between specific, generic, and, to a lesser extend, abstract content is employed.

The freedom of interpretation in the description of abstract content is far beyond the level of interpretation needed for description of ofness content, which implies that more intellectual and subjective efforts are needed in the process of indexing.

Shatford Layne (1994, pp. 584-585) discusses the attributes of images and distinguish four main categories: 1) biographical attributes; 2) subject attributes; 3) exemplified attributes; and 4) relationship attributes. Biographical attributes describe the formal aspects of an image. Subject attributes describe what is depicted in the image. Exemplified attributes describe the type of image in question of indexing, e.g., a picture, a drawing, or a poster. Relational attributes describe the relationship between images, e.g., an early sketch and the final drawing. Subject attributes are related to subject indexing, while the remaining three attributes are related to isness and descriptive indexing.

The concept-based approach is characterised by manual and intellectual interpretation of (moving) images and manual construction of surrogate records, and it adheres to a large extend to a subjective conception of the concept of aboutness (indexer as well as user aboutness). In contrast, the content-based approach substitutes human intellectual efforts with computational processing of digital inputs, in order to automatically construct surrogate records for (moving) images, and it adheres to an objective conception of aboutness, as expressed in author aboutness, and to some extend expressed in indexer aboutness. In the following section, we briefly review literature on the content-based approach for indexing and retrieval of (moving) images, by outlining and comparing different approaches for categorisation of attributes, and different types of processing tasks.
4.2.2 The content-based approach

According to Rui, Huang and Chang (1999) digital image retrieval can be traced back to the 1970s, and with the 1990s increased attention from the computer vision community, the field expanded remarkably.

Bimbo (1999, p. 2) distinguishes between three kinds of attributes for representation of (moving) images: 1) content-independent; 2) content-dependent; and 3) content-descriptive. Content-independent attributes are formal data (e.g. the author’s name, the date, and the format), which represent the isness of a (moving) image, and is hence related to descriptive indexing. Content-dependent attributes describe the content as the content-bearing units in the (moving) image express it. It is related to subject indexing, and more precisely to a conception of aboutness which is similar to author aboutness. Content-descriptive attributes require interpretation of the connection between entities in the (moving) images and entities outside the image and content-descriptive attributes related to a conception of aboutness, which is similar to indexer aboutness.

In similar vein, Eakins and Graham (1999) derive three levels of retrieval with increasing complexity. The first level involves retrieval based on the most basic features of (moving) images (e.g., colour, texture, shape, and motion), and this is comparable to Bimbo’s content-dependent attributes. The second level comprises retrieval based on derived features. These are features, which are analysed with reference to some generic or specific features of ofness in Shatford’s terminology, e.g., a car. The third level comprises retrieval based on abstract attributes, and it is comparable to Shatford’s abstract level of perception.

Hanjalic’s (2004, p. 8) distinction between a cognitive level and an affective level of analysis of content of moving images is regarded a further division of Bimbo’s content-descriptive level of interpretation. The cognitive level comprises the factual or concrete aspects of the (moving) image, and it is similar to Shatford’s notion of ofness. The affective level of interpretation aims at identification of emotions and moods and it is to some extend akin to Shatford’s abstract level of interpretation.

Throughout this chapter, we have provided different researchers’ perspectives on the kinds of attributes that are relevant when constructing surrogate records for (moving) images. Table 4.2 provides a comparison and summary of the different terminology applied, as well as the terminology we apply in the remainder of our work. Further, Table 4.2 relates the different kinds of attributes to the process of indexing, the concept of aboutness, and the level of interpretative freedom. The interpretation required for subject indexing rises as we move from left to right. That is, primitive
features induce a low level of freedom in interpretation, whereas abstract attributes induce more freedom in interpretation, and hence require a more complex high-level interpretation by the indexing device.

Table 4.2: Comparison of terminology for kinds of attributes for indexing and retrieval of moving images, and the relation to indexing and types of aboutness.

<table>
<thead>
<tr>
<th>Our terminology</th>
<th>Isness attributes</th>
<th>Primitive attributes</th>
<th>Ofness attributes</th>
<th>Abstract attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shatford (1986)</td>
<td>-</td>
<td>-</td>
<td>Specific and generic ofness</td>
<td>Abstract (”aboutness”)</td>
</tr>
<tr>
<td>Shatford Layne (1994)</td>
<td>Biographical, exemplified, and relationship attributes</td>
<td>-</td>
<td>Subject attributes</td>
<td></td>
</tr>
<tr>
<td>Bimbo (1999)</td>
<td>Content-independent</td>
<td>Content-dependent</td>
<td>Content-descriptive</td>
<td></td>
</tr>
<tr>
<td>Eakins and Graham (1999)</td>
<td>-</td>
<td>Primitive</td>
<td>Derived</td>
<td>Abstract</td>
</tr>
<tr>
<td>Hanjalic (2004)</td>
<td>-</td>
<td>-</td>
<td>Cognitive</td>
<td>Affective</td>
</tr>
<tr>
<td>Indexing</td>
<td>Descriptive indexing</td>
<td></td>
<td>Subject indexing</td>
<td></td>
</tr>
<tr>
<td>Aboutness</td>
<td>Isness</td>
<td>Author aboutness</td>
<td>Indexer aboutness</td>
<td></td>
</tr>
<tr>
<td>Interpretative freedom</td>
<td>-</td>
<td>Low-level</td>
<td></td>
<td>High-level</td>
</tr>
</tbody>
</table>

According to Wang and co-authors (Wang, Liu & Huang, 2000) the process of computational analysis of moving images comprises three basic tasks: 1) segmentation; 2) classification (also denoted indexing); and 3) summarisation. The three tasks are briefly explained below.

The process of segmentation is concerned with decomposing broadcasts into structural units. Inspired by previous categorisations of structural units (Yeo & Yeung, 1997, p. 46; Bimbo, 1999, pp. 9-10; Petkovic & Jonker, 2004, p. 37), we consider five structural units in a television broadcast context. The five units are depicted in Figure 4.1.
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

Starting from the bottom, a frame is the fundamental unit represented by the single image. A shot is a series of uninterrupted frames, e.g., a continuous sequence of frames between transitions e.g., cuts. A feature is a sequential collection of shots, which are unified by a semantic meaning, e.g., a dialogue scene in a motion picture, or the coverage of a specific event or story in a news broadcast. The identification of semantic units is not based on a perpetual set of laws, but based on the semantic situation and context. Features are also denoted scenes or stories in the literature. A programme is the logical unit, which comprises scenes or stories, e.g., a motion picture or the news. In this way, we consider a programme in a similar fashion to television broadcasters, and as an example, programmes are the units described in the television schedules. The term broadcast is defined as the continuous sequence of frames at any level of granularity. This ranges from a few frames to a programme as a whole, and it may even comprise a succession of several programmes, e.g., by comprising all broadcasts from one day. For moving images, in general the term clip is applied to denote this structural unit, which comprises frames at any level of granularity. These different levels of granularity are here introduced with reference to indexing of television broadcasts or moving images. However, the level of granularity is also important in relation to our
empirical investigation of the users’ behaviour when seeking for television broadcasts, since it is important to be informed about the level of granularity of the broadcasts, which users need and/or expect to find. Consequently, questionnaire participants are requested to indicate the level of granularity of needed television broadcasts, and the issue is also addressed in the nine interviews we have conducted (see e.g., Section 6.3.1 about our methodical approach).

The fundamental segmentation focuses on identification of image sequences or shots and it is termed ‘shot boundary detection’. That is, the automatic identification of the transition from one uninterrupted camera shot to another uninterrupted camera shot. This is comparable to decomposing a textual document into sentences. The transition from one shot to another can be either abrupt as with a cut or it can be gradual as with a fade, a wipe, or a dissolve (Hanjalic, 2004). According to Brunelli, Mitch and Modena (1999, p. 90) and Smeaton (2004, p. 384), algorithms detect cut transitions fairly well, whereas gradual transitions prove more difficult to detect. Shot boundary detection depends on processing of primitive features of a set of images, or regions in images. With the detection of shot boundaries, high-level automatic video parsing can group the shots into semantic segments or story bounds, e.g., episodes in a movie, reports in a news programme, or scenes in a sit-com (Hanjalic, 2004), or what we denote features. The detection of story boundaries relies heavily on prior information, e.g., knowledge about the programme structure for news broadcasts. O’Connor and colleagues (2001) develops an algorithm for story boundary detection in television broadcasted news and initial tests show promising results. It has proven more difficult to detect story boundaries for less structured television broadcasted programmes (Smeaton, 2004, p. 384). This again is related to the level of interpretative freedom in different types of information objects, as briefly discussed in Section 4.1.2. While the approaches described above primarily focus on the visual channel, the auditory channel or the textual channel could also be employed in the segmentation process (e.g. Li et al., 2001).

The task of classification aims firstly at identification and extraction of primitive attributes. These attributes are for instance represented by histograms (e.g. RGB-colours) (Rui, Huang & Chang, 1999, p. 41). Secondly, it aims at the construction of an index by mapping the machine processed values of the primitive attributes to higher-level human semantic processing (ofness and abstract attributes). This mapping is also denoted bridging the semantic gap (e.g., Enser, 2000; Bachimont, 2001; Hanjalic, 2004), and it is essentially concerned with the difference between the computational values of primitive attributes, and human interpretation of content conveyed in (moving) images. The mapping is typically achieved by developing a representative
model of the higher-level attribute being sought (e.g., an object, event, or a feeling). The development of the representative model is based on values of primitive attributes (colour, texture, shape, motion, etc.). Construction of the representative model can be done in several ways, e.g., by the researcher/indexer, by processing user feedback (e.g., in relevance feedback sessions), or by a combination of the two. Hence, the representational model is potentially dynamic. Based on the model at a specific point in time an algorithm tries to find similar values in other broadcasts, and when a match (or similarity above a predefined threshold) is found, the broadcast is indexed with the appropriate attribute. The focus of such algorithms have for instance been to detect and describe camera motion (e.g., Lehane, O'Connor & Murphy, 2004), detect and recognise people, (e.g., Li et al., 2001; Mohan, Papageorgiou & Poggio, 2001), detect objects (e.g., Zhang & Kuo, 1999; Mohan, Papageorgiou & Poggio, 2001), detect settings (e.g., Szummer & Picard, 1998; Vailaya, Jain & Zhang, 1998; Zhang & Kuo, 1999; Christel, Olligschlaeger & Huang, 2000), or detect affective moods (e.g., Corridoni, Bimbo & Vicario, 1998; Hanjalic, 2004). The predominant part of the research efforts have been directed towards ofness attributes. The mapping of abstract attributes have been undertaken by only few researchers such as Corridoni, Bimbo and Vicario (1998) or Hanjalic (2004).

According to Snoek and Worring (2005) most former research focuses on one channel (predominantly the visual channel), and is hence unimodal. Primarily, due to increased computer capacity, recent research tends to employ and integrate more channels in the analysis of the information objects, and can hence be categorized as multimodal (Snoek & Worring, 2005). Multimodal approaches show significantly better results (e.g., Bimbo, 1999; Smeaton, 2004), however, unimodal approaches are crucial for development and improvement of algorithms for analysing each channel.

The task of summarisation is concerned with building comprehensive abstracts or summaries of the moving images. There are two kinds of approaches for construction of such abstracts: 1) summarisation by extraction of key frames; and 2) skimming which reduces a broadcast into a trailer-like segment (Bailer et al., 2005, p. 67). The goal of summarisation is to enable a user to browse through a collection of retrieved information objects. Summarisation is important for interactive IR, however the processes involve a range of discussion on how to select and extract relevant frames, which will not be discussed here. Instead the reader is referred to the work by Hanjalic (2004) and Bailer and colleagues (Bailer et al., 2005).

This knowledge about different approaches for indexing of moving images, which we have discussed in this section, is important for our discussions on the applicability of external and other sources for construction of the identified appropriate access points in
Chapter 9. In the following section, we outline and discuss how external sources previously have been applied for generation of primary source representations.

4.3 External sources for indexing of primary information objects

Our fifth research question is concerned with the extent to which external and other sources can be applied for indexing of television broadcasts. Therefore, it is appropriate to outline and discuss previous work, which has used external sources as a supplement for creation of representations of primary information objects. These approaches can be divided into two groups with regard to whether or not they exploit embedded referential connections of the primary information objects in order to identify the external sources, which are subsequently analysed.

Several researchers have previously focused on external sources as a supplement for indexing of primary information objects. Among these are Small (1978), who builds on Garfield’s (1970) notion that citations in scientific articles are analogous to subject headings. Small exploits the referential connection of citations, analyses the context of several citations to highly cited papers, and identifies so-called ‘standard concept symbols’ for the highly cited articles. These standard concept symbols represent the citing researchers’ view of the content of the cited article, and they can hence be applied as descriptive keywords for the cited article. In this way, they are comparable to the access points constructed in democratic indexing. Building on Small’s (1978) notion of standard concept symbols, Schneider (Schneider & Borlund, 2002; Schneider, 2004; 2006) develops an approach for thesaurus construction and maintenance by exploring the context of citations. The text surrounding the citation is collected from a number of research articles (external sources), and through various analyses, a standard concept symbol is derived and ascribed to the representation of the cited primary information object, and applied in thesaurus construction.

The referential connections of hyperlinks on the Internet are exploited in a similar fashion by several researchers. As early as 1993, Dunlop and van Rijsbergen (1993) propose a model for retrieval of non-textual information objects on the Internet which exploits the referential connection of hyperlinks. The assumption behind the approach is that the text in, or in close proximity of, the hyperlinks usually describes the content of the information object they link to. Hyperlink anchored text as well as the text surrounding the hyperlinks will typically describe the content of the ‘cited’ web page. For non-textual information objects, hyperlink texts provide textual external sources for indexing of the primary non-textual information object. Hyperlink anchor text, text from URL’s, and text in the file name have subsequently been employed by several
researchers for indexing of images (e.g., Harmandas, Sanderson & Dunlop, 1997; Goodrum & Spink, 2001; Upstill, Nagappan & Craswell, 2001; Walker, 2002). In a similar fashion, Brin & Page (1998) use the hyperlinks between web pages in the construction of the ‘Page Rank’ algorithm, which is an important element of the Google search engine. Further, the basic idea is presumably applied in operational video retrieval systems on the Internet (e.g., Yahoo’s video search\(^5\)).

In order to organise musical genres or styles Knees and co-authors (Knees, Pampalk & Widmer, 2005) analyse the term occurrence in external sources in the form of web pages. The primary information object in this case is the single terms in a taxonomy of musical genres. This approach does not use any referential connection between information objects, but instead it relies on retrieval of external sources by searching several web search engines.

The MediAssist project undertaken by O’Hare and team members (e.g., O’Hare et al., 2005b) creates and exploits what might be termed external sources in order to organise and retrieve personal images. The external sources comprise information from weather stations based on the date and time stamp for taking a photo, as well as the geographical location captured by a GPS-device. Based on geographical location and the time the photo was taken, the photo is annotated with information about the weather. Further, by processing the data several other annotations can be made to each photo, e.g., season, daytime or night. These external sources (weather stations) do not have a referential connection to the primary information objects (photos), instead the external source requires specific data to be captured and linked to the primary information object, in order to extract relevant information (GPS-location, and time stamp). A test shows that the exploitation of such external sources for automatically annotation of images makes retrieval more effective (O’Hare et al., 2005a).

Few researchers have had specific emphasis on external sources in a television broadcast or motion picture context. Yap and colleagues (Yap, Simpson-Young & Srinivasan, 1996) describe a prototype application for indexing and retrieval of documentaries in film archives. Secondary textual information objects created in the documentary production process (e.g., screenplays, transcripts or shot lists), are synchronised with the digitised moving images. The main aim for synchronising the external sources with the moving images is to allow for searching and retrieval of the moving images by free text searching of the external sources. With emphasis on television broadcasts at the French national audiovisual archive, Lespinasse and Bachimont (2001) use the term ‘peritext’ to denote what we have termed external sources (e.g., television schedules or programme manuscripts), and they propose

“peritext” is employed for indexing of television broadcasts. In this way, peritext is not to be understood in rhetorician Genette’s (1991; 1997) original conception of the term, since external sources are more comparable to epitext in Genette’s terminology.

The approaches outlined above show that several researchers have exploited available external sources for improving access to primary information objects. The external sources are used in order to derive descriptions of primary information objects, which otherwise would be unfeasible to construct. The focus in these approaches is given to the possibilities for applying or extracting information from the external sources. For instance, the results by Yap and colleagues (Yap, Simpson-Young & Srinivasan, 1996) focussed on free text searching, however whether free text searching is the most appropriate mean for the users’ access to documentaries is never investigated. By comparison, our primary focus is on the users and aspects of their information seeking behaviour prior to considerations as to the applicability of external sources. Hereby, we gather knowledge about the access points and features, which are appropriate for the users, and how automatic processing of external sources could be applied for indexing of the primary information objects. The point is that knowledge about the users’ information needs, preferred search entries, and applied relevance criteria is important in order to provide users with effective IR systems. This said, we do acknowledge the importance in gathering knowledge about possible approaches and features, as a prerequisite for mapping such approaches and features to knowledge about the actual as well as potential users’ information seeking behaviour.

This concludes our section on earlier approaches for exploitation of external sources for indexing and retrieval of primary information objects. The following section provides summary statements.

4.4 Summary

Technological advances have had a tremendous impact on the information objects used in society, and consequently, they have affected information objects made available in libraries. The traditional library collection of books has been extended with audiovisual information objects, including television broadcasts. In the same manner as with other information objects, an index is required in order to retrieve these broadcasts. Indexes, regardless of the media has traditionally been constructed by describing the information object in question manually (Hertzum, 2003, p. 169; Smeaton, 2004, p. 380). However, the process of manual indexing of television broadcasts is very resource demanding, and 1 hour of footage can take as long as 30 hours to index manually (Enser, 2000, p. 202). Due to the technological evolution, information objects can now be made available in
digital form, including television broadcasts. This enables computers to analyse the information objects automatically, constructing an index by automatic means. Automatic indexing allows for a reduction in the resources needed for indexing television broadcasts.

The two approaches for indexing and retrieval of (moving) images are denoted the concept-based and the content-based approach. Terminologically, these denotations are somewhat misleading, since analysis of the content of the (moving) images is not restricted to the content-based approach. For instance, Panofsky’s use of the term of content relates to his third stratum of perception, which requires intellectual interpretation. In contrast, content-based indexing primarily focuses on the primitive content attributes of the (moving) images, or the content as it is. The difference between the two approaches can be described as human intellectual interpretation of high-level concepts, versus computational processing of low-level content features. In comparison to De Mey’s four stages of processing (see Section 2.2.2), the concept-based approach is capable of indexing (moving) images at the cognitive and epistemic stage of processing, whereas the content-based approach is restricted to the lower stages of processing. Further, the primary focus on primitive content features implies that the content-based approach conducts processing at De Mey’s monadic stage. Generally speaking content-based indexing does not attempt to express the abstract attributes of (moving) images (Rasmussen, 1997, p. 183; Bimbo, 1999, p. 12; Chen & Rasmussen, 1999, p. 293). This refers to the so-called semantic gap between computational processing and intellectual interpretation.

With reference to the three types of aboutness discussed in Section 4.1.1, the content-based approach’s focus on primitive features relates it to author aboutness, e.g., as seen in Snoek and Worring’s (2005) explicit focus on the author’s perception of the moving images. This said, the higher-level processing of ofness and to a lesser extent abstract attributes relates the content-based approach to indexer aboutness. The concept-based approach’s focus on intellectual interpretation of the (moving) images relates it to user and indexer aboutness.

Research in the field of automatic indexing of (moving) images is still in its infancy, and researchers (e.g. Bachimont, 2001, p. 1) argue that automatic analysis tools can only analyse the physical descriptors of the television broadcasted programmes, and achieving conceptual representations will necessitate a manual analysis. Nevertheless, content description by automatic means shows promising results (e.g. Smeaton, 2004; Snoek & Worring, 2005), and it is recognized that automatic approaches for indexing television programmes can complement traditional manual approaches (e.g., Bimbo, 1999, p. 12; Rui, Huang & Chang, 1999, pp. 54-55; Smeaton, 2004, p. 382).
Content-based approaches for indexing of (moving) images require the information objects to be digitized prior to computational processing, whereas intellectual interpretation, which is the main element of concept-based approaches, can be conducted with digitized as well as analogous (moving) images. The automatic exploitation of external sources for indexing can be conducted whether the primary information object is of a digital or analogous character. That is, the important element is for the external sources to be of a digital nature. In this way, digital external sources might be applied for indexing of digital as well as analogous primary information objects. Further, another important element for automatic exploitation of external sources for indexing is identification of the connection between a primary information object and an external source. Different clues in the external sources can be applied for identification of these connections, e.g., as seen in citation or hyperlink analysis. Our work does not aim at identification of such referential cues. Instead, we take one step back and investigate to what extent the textual external sources contain data, which can be exploited in the indexing of television broadcasts (digital as well as analogue). Future research can focus on automatic identification and extraction of such exploitable information.

In the introductory section of this chapter, we stated that indexing implies a given searching and retrieval behaviour from users. It is consequently important to gather knowledge about aspects of the users’ information seeking behaviour (e.g., information need characteristics, search entry preferences, and application of relevance criteria), in order to identify the behaviour that should be supported in a given information seeking context. Knowledge about different indexing approaches is important in order to be able to discuss how different approaches lend support to different kinds of information seeking behaviour. Our point of departure in our empirical study is to gather knowledge about three aspects of user’ information seeking behaviour, namely characteristics of information needs, preferred search entries, and applied relevance criteria, in the context of television broadcasts, in order to identify appropriate access point in surrogate records. The knowledge gathered in this chapter is important in order for us to discuss the approaches for indexing that are appropriate in order to support the construction of the identified appropriate access point. The applicability of the different approaches is briefly touched upon in Chapter 9.

This concludes the theoretical part of the thesis. Chapter 2 explains how the cognitive viewpoint constitutes our epistemological background, and how the viewpoint integrates research in information seeking and IR. The chapter places the present work within the cognitive viewpoint approach to LIS, and most notably the emphasis on a holistic approach. With reference to the discussion above, the cognitive viewpoint does
not imply the existence of any overall best practice. Instead, the viewpoint emphasises that we need to gather knowledge about different approaches (e.g., for indexing), and knowledge about the situational context. Hereby, we are able to identify the best practice (e.g., the combination of indexing approaches) for the contextual situation at hand. Chapter 3 and the present chapter focus on each of their part of this investigation. Chapter 3 focus on previous research in the area of information seeking behaviour and how this behaviour is related to the underlying work task and the concept of the information need. Further, the chapter outlines characteristics of humanists’ information seeking behaviour. The knowledge gathered in Chapter 3 is important as a prerequisite for our investigation of three aspects of users’ information seeking behaviour (information need, preferred search entries, and applied relevance criteria), and the knowledge is an important precondition for our methodical approach. The present chapter provides knowledge about indexing approaches and as discussed above, this knowledge is important in order to discuss the value of information in external sources and hence the applicability of the information in such external sources for future construction of the identified access points in television broadcast surrogate records.

In the following five chapters, the empirical study is outlined and discussed, prior to our conclusion in Chapter 10. The following Chapter 5 explicates the context of our empirical study as well as outlines our identification and selection of a user target group for our empirical study. Chapter 6 explains our methodical considerations for the empirical study, and Chapters 7, 0, and 9 reports on the results of our empirical investigation.
5 Empirical context

This chapter introduces the context of our empirical studies. The purpose of these introductions is to provide background information for our methodical disposals, outlined in Chapter 6, as well as our analysis and results reported on in Chapters 7, 0, and 9.

Firstly, the chapter outlines the historical and contemporary situation for television broadcasting in Denmark. Secondly, the chapter outlines the present situation for accessing television broadcasts in Denmark. The chapter gives a thorough description of the Danish national collection of television broadcasts at the State and University Library, as well as a short introduction to alternatives for searching and retrieval of Danish television broadcasts. Thirdly, the chapter outlines and discusses our selection of a user target group for the empirical investigations. The selection is based partly on investigation of the present users of the Danish national archive for television broadcasts, and partly on practical circumstances. The investigation of users of the archive is given priority, and hence carefully described and discussed in the chapter. The user target group selected is scholars and students in the field of Media Studies, and hence the chapter gives a brief introduction and overview of the academic field of Media Studies.

5.1 Television broadcasting in Denmark

This section briefly outlines the historical and contemporary settings for television broadcasting in Denmark. The outline focuses on production and broadcasting of Danish television, and, where relevant, institutional and broadcasting infrastructural elements are included. The purpose of this sketch of the historical development of Danish broadcasting is for the reader to better understand the work tasks and information needs of our respondents, which are reported in Chapter 7.

The Danish television culture is regarded as one of the most stable television cultures in Scandinavia, with DR as the supreme broadcasting institution until the second national channel (TV2) commenced broadcasting in October 1988 (Bondebjerg, 1994, p. 45). In line with Søndergaard (2006, p. 26), we divide Danish television broadcasting into three phases: 1) the formative phase (1951-1964); 2) the monopoly phase (1964-1988); and 3) the competitive phase (1988-).
The formative phase covers the period from 1951 to 1964. 1951 is the beginning of Danish television as the Danish authorities (Radiorådet) decides to broadcast television in a try-out period of three years. Television broadcasting is considered a national affair, and it is established as a public monopoly with public service liabilities (Søndergaard, 2006, p. 29). Production and broadcasting is undertaken by the Danish State Broadcasting Service (Statsradiofonien), later the Danish Broadcasting Corporation (Danmarks Radio) or in short DR. The formative aspects of the phase concerns the broadcasting station (e.g., technical equipment), the establishment of the broadcasting infrastructure (e.g., broadcasting towers), as well as the distribution of television sets, exemplified with the development of television licenses from 0 in 1951 to nearly 1,000,000 in 1964 (Søndergaard, 2006, p. 30). An important element in the dissemination of television is the coverage of broadcasting, and it was not until the broadcasting tower at the island of Bornholm in 1960, that the whole country could receive the transmitted television (Ahm, 1972, p. 171). The broadcasting in the first years comprises 1 hour of programmes, three days a week, which rose to nearly 30 hours a week at the end of the formative phase in 1964.

The monopoly phase starts with the segregation of the organisation of radio and television production in 1964, and ends with the abolishment of the broadcasting monopoly in 1988. DR is the sole national television channel during the 24 years period. Towards the end of the phase a few small local television channels are established e.g., TV Syd in 1983. The weekly broadcasting doubles from 30 to 60 hours in the period. One of the main technical developments in the monopoly phase is the transmission of television in colours starting in 1967 (Ahm, 1972, p. 234). According to Hjarvard (2006a), DR is to be considered a cultural institution in the monopoly phase, in contrast to the media company it turns into in the competitive phase.

The competitive phase starts with the establishment of TV2 as the second Danish television station. The Danish Parliament puts the legislation for a second national terrestrial broadcasting channel through in 1986, and the legislation is carried out with the first broadcasting in October 1988. TV2 is partly financed by licence fees, and partly by commercial revenues. The already existing regional channel TV SYD becomes part of TV2|Danmark in 1988, and seven other regional channels are gradually established within the TV2|Danmark broadcasting corporation, these are: TV2 /Fyn, TV 2 /Nord, and TV/Midt-Vest in 1989; TV2 / Østjylland and TV 2/Bornholm in 1990; TV2 Øst in 1991; and TV 2/Lorry in 1994. Prior to that, TV3 had started broadcasting to the Danish audience from England in 1987, and the abolishment of the monopoly meant that several channels were to follow, e.g., DR2 in 1996 (the main DR channel is at the same time re-launched as DR1); TV2 Zulu and TV Danmark (later Kanal 4) in
2000; TV2 Charlie in 2004; TV2 Film in 2005; and most recently TV2 News in 2006 and TV2 Sport in 2007. In 2003, the regional stations are economically and organisationally detached from TV2|Danmark. The new channels on the marked means more competition, and the audience share becomes one of the strongest parameters for measuring the penetration rate of each channel. The competitive phase gives rise to the extensive use of audience ratings for measuring channel performance, available through the TV-Meter measuring system (tns Gallup, 2007), which is established in 1992. Further, the competition means an explosive rise in the weekly hours of broadcasting e.g., 138 hours in 2005 from DR1 alone (see e.g., the development in hours of yearly broadcasting in Appendix 1).

In recent years, the development of the broadcasting corporations into media corporations with specialised broadcasting channels (e.g., TV2 Sport), Internet broadcasted television, and Web 2.0 portals containing the users own video clips, may be the beginning of a new phase in Danish television. A phase, which might be denoted the atomised and personalised audience phase. However, such analyses are outside the scope of this thesis, and we leave it to the judgment of posterity to arrive at such conclusions.

This ends our brief outline of the historical development of Danish television broadcasting. This outline is an important prerequisite for understanding users’ needs for television broadcasts, and hence it is a precondition for fully comprehension of our results. This takes us to the present situation for fulfilling needs for television broadcasts by accessing archived broadcasts.

5.2 Access to Danish television broadcasts

The purpose of this section is to give the reader an understanding of the present situation for accessing archived broadcasts transmitted via Danish television channels. The section comprises two sub-sections. In the first sub-section, the focus is given to the collection of television broadcasts at the State and University Library, since it is the national institution that has the task to preserve and give access to the Danish cultural television heritage. In the second sub-section, we cursorily sketch other alternatives for accessing archived television broadcasts in Denmark.

5.2.1 The collection of television broadcasts at the State and University Library

This section extends the previous introduction to the Danish national collection of television broadcasts given in our introductory chapter (Section 1.1.1). The prime focus of this further introduction is on the content of, and access to the collection.
The Danish national collection of television broadcasts is established at the State and University Library in 1987 with the sole function of preservation of the cultural heritage (Ministeriet for Kulturelle Anliggender, 1987). The archive immediately began the collection of the transmitted broadcasts, and over the years, the archive has also been able to collect broadcasts, that were transmitted prior to 1987. This includes for instance all broadcasting from TV Syd from its establishment in 1983 and onwards, and news programmes from DR1 from 1981 and onwards. The collection further comprises all broadcasting from TV2, DR2, the eight regional channels, TV2 Zulu, TV2 Charlie, and TV2 News, from each channel’s first transmission, as well as the Danish broadcasts from TV3 transmitted from 1998 and onwards. The coverage of the archive is depicted in Table 5.1. Further, the fourth column in Table 5.1 contains the hours of broadcasting from each channel and the fifth column contains the distribution in percentages.

Table 5.1 shows that the main part of the collection comprises broadcasts from the two national terrestrial channels DR1 and TV2 (54.7%). In addition, the collection comprises broadcasts from local television stations from one week each year, included in the ‘Other’ category in Table 5.1. Further, the collection contains all broadcasts from the sports channel: TV S also known as the ‘Elkjær channel’. The channel commenced transmission on 1st of March 1997, but due to financial problems, broadcasting was terminated by New Year, 1998. This illustrates one of the main arguments for having the archive in the first place, namely that television stations perish, while libraries persist. In total, the collection comprises 350,537 hours of broadcasting, of which approximately 90% are store in an analogical format, mainly on VHS-tapes.

The legislation for establishing the archive in 1987 is not followed by a Copyright Act for television broadcasts until 1995. Consequently, the archive is not given a legislative basis for collecting or giving access to the broadcasts in its early years, and the archive depends on the television stations kindness with respect to handing over their broadcasts on a freely basis. This means that broadcasts become part of the archive approximately three months after transmission (Kirring, 2001). The missing copyright legislation also means that no users can be granted access to the collection in the early years. Even the stations that freely hands over their broadcasts cannot be granted access to the very same broadcasts, once these broadcasts are part of the collection.

Television broadcasts are for the first time covered in the Copyright Deposit Act of 2005, which means that the archive is given the legislative basis for collecting television broadcasts. In this way, the archive is not dependent on the broadcasts being handed in freely by each station. Instead, the archive has the legislative basis for recording the transmission digitally. Therefore, from January 2006 and onwards, the
broadcasts in the archive are recorded and stored digitally. Digitalisation has so far not included the archive in any retrospective way.

Table 5.1: Broadcasts in the Danish national collection of television broadcasts distributed on station, and channel, and with indication of coverage in hours and percentages.

<table>
<thead>
<tr>
<th>Station</th>
<th>Channel</th>
<th>In collection</th>
<th>Hours of broadcasting(^1)</th>
<th>Percentages of collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR</td>
<td>Early broadcasts</td>
<td>1970-1986</td>
<td>9,000</td>
<td>2.6%</td>
</tr>
<tr>
<td></td>
<td>DR1</td>
<td>1987-2006</td>
<td>93,510</td>
<td>26.7%</td>
</tr>
<tr>
<td></td>
<td>DR2</td>
<td>1996-2006</td>
<td>34,087</td>
<td>9.7%</td>
</tr>
<tr>
<td></td>
<td>DR in total</td>
<td></td>
<td>136,596</td>
<td>39.0%</td>
</tr>
<tr>
<td>TV2/Danmark</td>
<td>TV2</td>
<td>1988-2006</td>
<td>98,004</td>
<td>28.0%</td>
</tr>
<tr>
<td></td>
<td>TV2 Zulu</td>
<td>2000-2006</td>
<td>35,658</td>
<td>10.2%</td>
</tr>
<tr>
<td></td>
<td>TV2 Charlie</td>
<td>2004-2006</td>
<td>8,791</td>
<td>2.5%</td>
</tr>
<tr>
<td></td>
<td>TV 2 News</td>
<td>2006</td>
<td>684</td>
<td>0.2%</td>
</tr>
<tr>
<td></td>
<td>TV2 in total</td>
<td></td>
<td>143,137</td>
<td>40.8%</td>
</tr>
<tr>
<td>Regional stations</td>
<td>TV Syd(^2)</td>
<td>(1983)1987-2006</td>
<td>6,809</td>
<td>1.9%</td>
</tr>
<tr>
<td></td>
<td>TV 2/Fyn</td>
<td>1989-2006</td>
<td>4,520</td>
<td>1.3%</td>
</tr>
<tr>
<td></td>
<td>TV 2/Nord</td>
<td>1989-2006</td>
<td>7,091</td>
<td>2.0%</td>
</tr>
<tr>
<td></td>
<td>TV/Midt-Vest</td>
<td>1989-2006</td>
<td>9,115</td>
<td>2.6%</td>
</tr>
<tr>
<td></td>
<td>TV2 / Østjylland</td>
<td>1990-2006</td>
<td>5,520</td>
<td>1.6%</td>
</tr>
<tr>
<td></td>
<td>TV 2/Bornholm</td>
<td>1990-2006</td>
<td>5,508</td>
<td>1.6%</td>
</tr>
<tr>
<td></td>
<td>TV2 Øst</td>
<td>1991-2006</td>
<td>7,817</td>
<td>2.2%</td>
</tr>
<tr>
<td></td>
<td>TV 2/Lorry</td>
<td>1994-2006</td>
<td>5,318</td>
<td>1.5%</td>
</tr>
<tr>
<td></td>
<td>Regional stations in total</td>
<td></td>
<td>51,699</td>
<td>14.7%</td>
</tr>
<tr>
<td>TV3</td>
<td>TV3 (Danish broadcasts, only)</td>
<td>1998-2006</td>
<td>7,106</td>
<td>2.0%</td>
</tr>
<tr>
<td>TV Danmark (Kanal 4)</td>
<td>TV Danmark (Kanal 4)(^3)</td>
<td>(one week)</td>
<td>267</td>
<td>0.1%</td>
</tr>
<tr>
<td>dk4</td>
<td>dk4</td>
<td></td>
<td>9,292</td>
<td>2.7%</td>
</tr>
<tr>
<td>TVS</td>
<td>TVS</td>
<td>1997</td>
<td>2,440</td>
<td>0.7%</td>
</tr>
<tr>
<td>Other (e.g., local)</td>
<td>Other(^3,4)</td>
<td>1987-2006</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>In total</td>
<td></td>
<td></td>
<td>350,537</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Legend: 1): The hours of broadcasting in the collection is calculated from the broadcasting figures stated by each station, and the figures are consequently subject to some uncertainty. For further details and indication of sources, see Appendix 1.
2): TV Syd’s broadcasts from 1983-1986 are covered by the State and University Library, but due to practical reasons these broadcasts are counted as part of the ‘Early broadcasts’
3): One week of broadcasting is collected to represent each years broadcasting.
4): No figures or estimates are available.

Due to the digitalisation, the broadcasts are available one or two days after transmission in contrast to the three months prior to 2006. However, the digitalisation has not
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

affected the user’s searching and retrieval access to the collection in any way. Users still need to state the name of the broadcasting channel, the title of the broadcasted programme, the exact date and time of broadcasting, and administrative information in order to retrieve broadcasts from the collection (see e.g., the requisition form in Appendix 2), and no search facilities are available for the users. Because search facilities are not available, users must retrieve the formal data elsewhere, e.g., by relying on memory, or by consulting printed television schedules. The printed television schedules are filed chronologically for each channel similar to the arrangement of the physical collection of broadcasts at the State and University Library. The user is to know by necessity exactly which programmes he or she is looking for, and the user must provide all the formal information the librarian needs in order to treat the request unambiguously. The requirement to provide administrative information is to ensure that users intend to use the programmes in accordance with the legislation, and to insure that the users can be granted access to the requested programme. The users must have a very clear perception of their information need prior to interacting with the intermediary, and the librarians can handle information needs concerned with location of known items, only.

The collection exists, but is difficult to access. Digitalisation has not improved the collections accessibility simply because no electronic search facilities are (developed and) provided. In this respect, alternative ways of getting access to Danish television broadcasts are interesting. These are briefly introduced in the following section.

5.2.2 Other alternatives for searching and accessing Danish television broadcasts

From a historical perspective, the most important alternative to the collection at the State and University Library is the broadcasting archive at DR. DR is the sole transmitting channel prior to 1987, and hence their archive comprises a unique collection of historical broadcasts. This said, a portion of these broadcasts have been lost for posterity, either because they were never stored or because they were later deleted e.g., in order to save storage space. This is mainly due to the technical production aspects and/or the economical resources needed for storage of the broadcasts, and the amount of broadcasts lost for posterity is naturally bigger the older the broadcasts gets. The registration of broadcasts at DR has been conducted in several separate places and into separate databases. Recently, these registrations have been combined into one overall archive. However, the archive and databases at DR are constructed with internal media production in mind, and not for external use (Hjarvard,
2006a, p. 22), and hence a combined overall database of transmitted broadcasts has never been composed.

Due to the absent search and retrieval facilities at the Danish national collection of television broadcasts, any searchable index of television broadcasts is important. From a historical perspective this include the many different databases at DR, another important source is the index of Danish broadcasts from 1951-1964 constructed by Danish media researchers denoted ‘the Registrant’ (Hjarvard & Jespersen, 2006). The databases at DR are production files based on the stored clips of moving images, whether these have been broadcasted or not. In contrast, the Registrant is concerned with the broadcasts that have been transmitted, whether these broadcasts are stored or not. Essentially, the Registrant is an expanded and searchable version of the printed television schedules. The Registrant comprises information about every programme broadcasted in Danish television (DR) from the beginning in 1951 until 1964, and it is the most thorough index for the historical part of Danish television. The Registrant does not hold any information about availability or location of broadcasts, or even whether or not the broadcasts still exist. Such archival information must be retrieved from one of DR’s databases.

For the more recent part of the broadcasts from Danish television stations, the TV-Meter system comprises the most comprehensive index. The TV-Meter system is managed by the analysis institute tns Gallup, and it comprises data of audience behaviour (e.g., the channels and programmes consumed by whom), as well as each of the approximately 1,000 participant’s personal opinions (e.g., politically) and specific consumption behaviour (e.g., the purchase of conventional and organic milk products). The main purpose of the TV-Meter system is to conduct complex audience reception and consumer analysis, and the users of the TV-Meter system are television stations, commercial and/or media companies, and a few scientific research institutions (Hjarvard, 2001, pp. 13-14). The TV-Meter system holds detailed information on broadcasts from several channels (e.g., DR1, DR2, TV2, and TV2 Zulu), from 1992 and onwards. The TV-Meter system is different from the Registrant in the granularity of broadcasts included, as well as the type of information it contains about each broadcast. The TV-Meter system contains database entries for programmes as well as the trailers, commercials, spots, etc. in the interim period of programmes. Further, the system contains a variety of access points for each entry, e.g., country, language, production company/department, duration, original title, target group, and keywords, reported by each television station. Hence, the TV-Meter system is applicable as a tool for searching of television broadcasts, transmitted later than 1992. However, as with the Registrant, the TV-Meter system does not contain any information about access to or
location of the broadcasts. In addition, since the system is developed for rather complex reception and consumer analyses, and searching for television broadcast is only a by-product, it can be very difficult to use by anyone, but the expert user. The Registrant and the TV-Meter system both aim at full coverage of transmitted broadcasts.

Other institutions have focused on archiving television broadcasts e.g., Roskilde University Library or the County Centres (Amtscentralerne, now Centre for Undervisningsmidler – [Centres for Educational Service]). Three main aspects differentiate such indexes or catalogues from the Registrant and the TV-Meter system, namely: 1) selected broadcasts are included, only; 2) human intellectual effort is applied for indexing of the selected broadcasts; and 3) all the indexed broadcasts are stored and accessible in local archives.

In addition, some private initiatives are initiated in order to construct indexes of television broadcasts. The most ambitious of these initiatives is ‘Dansk film og TV’ – [Danish movies and television], which is an index of Danish movies and television broadcasts. According to the owners of the website, the index contains descriptions of approximately 1,600 television broadcasts (e.g., see http://danskfilmogtv.dk/).

In the aftermath of the digital storage and retrieval at the State and University Library, the library is constructing a searchable archive of the digital part of the collection, with direct electronic access to the archived television broadcasts. The index is based on electronic television schedules, and due to technical as well as administrative issues, the database is available for the in-house users, only, but the archive is working on a public accessible version of the database.

In brief, the Danish national collection of television broadcasts at the State and University Library is the main collection of television broadcasts, which are accessible for research purposes in Denmark. However, no search facilities are provided for the collection, and the collection is consequently not searchable. Instead other databases are available to search for transmitted broadcasts (e.g., the Registrant and the TV-Meter system), but they are either concerned with old broadcasts (1951-1964), or new broadcasts (1992 and onwards), only. Further, these registers focus on transmitted broadcasts, only and they do not hold any information about the availability of television broadcasts. From a historical perspective the State and University Library holds a fraction of broadcasts transmitted prior to 1987, only. For access to such historical broadcasts, users are referred to the broadcasting station, namely DR. That is, if these broadcasts are stored at all. This ends our section on access to television broadcasts in Denmark. In the following section, we focus on our selection of user target group for the empirical investigations.
5.3 User target group

This section outlines our selection of user target group. Knowledge about the typical users is an important prerequisite for selection of a user target group for our empirical investigations. In order to understand who the typical users of television broadcasts are, and why they request television programmes, we analyse genuine requests for television broadcasts made to the State and University Library. We report on an investigation of the genuine requests in the first sub-section. Subsequently, we outline our rationale behind the selection of Media Studies as the user target group for our main empirical investigations. Finally, we introduce the field of Media Studies for the reader. The introduction serves as background knowledge that can be used in relation to our analysis and discussion of the empirical data in Chapters 7, 0, and 9. This is in line with our purpose for introducing the historical development of Danish television in Section 5.1.

5.3.1 Users of television programmes from the State and University Library

The data set analysed in this section consists of genuine requests for television broadcasts. The State and University Library generously allowed us to analyse genuine request for television broadcasts already received at the library. The analysis of these requests constitutes our point of departure for identification of an appropriate user target group for the empirical investigation of aspects of information seeking behaviour in the context of television broadcasts. Analysis of the requests is previously reported on and discussed in Kirkegaard and Borlund (2006).

The data set consists of 1,688 genuine requests electronically submitted to the State and University Library in the normal course of its activities. They constitute all the electronic requests received in 2004. An electronic requisition form is introduced in 2003 (e.g., see Appendix 2), and the traditional paper version of the requisition form is being phased out during 2004. Consequently, the paper version is applied by users concurrently with the electronic version. Requests on paper comprised 128 requests, or 7% of the requests received in 2004. Due to the direct availability of the information in the electronic requisition forms, the low usage of the paper version, and uncertainty with respect to the legislation for research access to the paper version, we decided to concentrate our investigation on the electronic version, only. We see no problem in omitting the paper request from our investigation of the users who request broadcasts from the archive, since they represent a small proportion of the requests, only. Further, a brief analysis does not indicate that the paper requests differ noticeably from the electronic requests.
The 1,688 information requests originate from 340 individual users and contain valuable details besides information about the requested television programme. From the requisition forms we know that the gender of the 340 users are distributed with 60% being women and 40% men, and they submits on average 5.0 requests each, with a minimum of 1 and a maximum of 178 requests from one individual user. The majority of the users, requests a few television programmes in 2004, only. 80% of the users’ requests 5 or less television programmes, and approximately half of the users (48%) put forward only one request to the library. Hence, the distribution of requests is rather skewed as shown in Figure 5.1, and documented with the standard deviation being 12.8.

![Figure 5.1: Number of requests from individual users.](image)

Analysis of the requisition forms reveals that 90% of the users are affiliated with an educational or research institution (e.g., university or college of education). While 5% of the users are affiliated to non-commercial establishments (e.g., library), compared to 3% of the users who are affiliated to commercial organizations (e.g., advertising agency). 2% of the users have personal affiliations, and for three users (1%) we are unable to determine their affiliations. The distribution of the users’ affiliations is depicted in Table 5.2.

The greater part of the users from the category of ‘Education or research’ are affiliated with institutions at an academic level, as academic users account for 67% of all individual users, and 80% of all requests, termed academic users, and academic
requests, respectively. 16% of the 229 users who are affiliated with academic institutions are scholars, whereas 84% are students. On average a scholar request more television programmes than a student (respectively 9.1 television programmes pr. scholar compared to 5.3 television programmes pr. student). Accordingly, Table 5.2 shows that the scholars account for 25% of the academic requests whereas students account for 75%.

Table 5.2: User affiliations.

<table>
<thead>
<tr>
<th>Users</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education or research</td>
<td>90%</td>
</tr>
<tr>
<td>Academic</td>
<td>92%</td>
</tr>
<tr>
<td>Faculty</td>
<td>80%</td>
</tr>
<tr>
<td>Student</td>
<td>75%</td>
</tr>
<tr>
<td>Non-academic</td>
<td>23%</td>
</tr>
<tr>
<td>Non-commercial</td>
<td>12%</td>
</tr>
<tr>
<td>Commercial</td>
<td>3%</td>
</tr>
<tr>
<td>Personal</td>
<td>3%</td>
</tr>
<tr>
<td>Unassigned</td>
<td>1%</td>
</tr>
<tr>
<td>Total</td>
<td>101%</td>
</tr>
</tbody>
</table>

The users of the collection are required to state their purpose for requesting a programme from the collection, and such information is hence found in 99% of the requests. The quality and comprehensiveness of these purposes for requesting television programmes is somewhat erratic. That is, several of the requisition forms contain minimal information e.g., ‘research’, ‘for studies’, ‘university project’, while other contains long description of the specific work task to be undertaken. However, we are able to identify some characteristics across the requests from academic users. For ¾ of the academic users (171 of the 229 academic users) the requisition forms contain enough information to identify the academic departments that the users are affiliated to. By comparing the information with the organisational arrangement of departments in faculties from the universities own websites, we identified the branch of science for each department. The predominant part (134 users – 59%) of the 229 academic users is affiliated with departments from the humanities, and a smaller part (24 users – 10%) is affiliated with social science departments. Seven users (3%) are affiliated with science departments, four (2%) are from theology departments, two users
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

(1%) come from health science departments, and for 25% of the users we were unable to assign specific departmental affiliation. The distribution is depicted in Figure 5.2

![Figure 5.2: Distribution of academic users’ affiliation to branches of science.](image)

A detailed examination of the requisition forms in relation to users’ affiliation indicates that most users are concerned with Media Studies. Other frequently occurring areas of research include Journalism, Communication, Cultural Studies, and Political Science.

The television programmes from the Danish national collection are predominantly requested by academic users to be used as research objects, which is not surprising since humanists generally take particular interest in primary sources (e.g., Watson-Boone, 1994; Barrett, 2005). Our analysis of the requisition form shows that the main part of users is affiliated academic institutions, and mainly departments within faculties of the humanities. Further, our analysis shows that both students and scholars request broadcasts from the collection. This information is an important element in the selection of a user target group for our empirical investigation of aspects of information seeking behaviour in the context of television broadcasts, and in the following section, we outline how we make use of this information for selection of a user target group.

### 5.3.2 Selection of user target group

In order to follow the recommendation to make in-depth studies of a well-defined category of persons put forward by Wilson (1981, p. 11), we need to identify and select such a well-defined group. With reference to the rules for accessing television broadcasts from the State and University Library (Kirring, 2006; Statsbiblioteket, 2006), the scientific area of specialisation is selected as the main definitional element in our
Chapter 5: Empirical context

selection of a user target group. Based on the investigation reported above, the field of Media Studies was chosen as the area of specialisation. This selection is based on several intermingled considerations. First of all, the field of Media Studies covers a tangible unit of users, and Media Studies is a humanistic field (see Section 5.3.3 below). Secondly, the investigation shows that scholars and students in Media Studies comprise a somewhat large amount of users. Due to several reasons, television broadcasts are only applied by a somewhat limited amount of scholars and students, and in our choice of user target group we were attentive towards selecting a user target group that could be expected to comprise a fairly amount users with information needs for television broadcast. This is important in order to ensure the methodical soundness of the main investigations, which are outlined in Chapter 6. Based on our investigation reported above and the nature of the field (outlined in Section 5.3.3) we expected Media Studies to comprise a fair amount of users of television broadcasts.

Further, we were able to arrange a two month research visit at one of the departments of Media Studies in Denmark. This practical factor is an important element in our methodical approach for investigating the users’ information needs, search entry preferences, and relevance criteria application, and it is further outlined in the following Chapter 6.

In brief, we define our user target group as scholars and students associated with one of the two main university departments, which offer Media Studies educations in Denmark: Media Studies, Aarhus University, and Film and Media Studies, Copenhagen University. Therefore, it is relevant to introduce the field of Media Studies in Denmark, which is the focus of the following section.

5.3.3 Introduction to the field of Media Studies

This introduction to the field of Media Studies serves as a further introduction to the selected user target group, as well as background knowledge in connection to our analysis presented in Chapter 7. The purpose of this introduction is to provide a brief overview of the history, content, and methods in Danish media research, as well as the organisation of the field. Consequently, our review is not comprehensive, nor does it provide a detailed description of the different specialisations at particular departments, or go into details about description and/or evaluation of specific research results. The introduction focuses on two Danish universities that offer bachelor and Master’s educational programmes in Media Studies in 2006. These are Aarhus University and Copenhagen University. At both universities, the field is placed within the faculty of the humanities, which, with reference to the investigation reported above, is one of the reasons for our choice of the field as user target group.
Our introduction is based on several sources. First of all, we consulted curricula and syllabuses from each Media Studies education provided by the two universities, introductory information from the universities web-sites, and the official education guide from the Danish Ministry of Education. These sources are chosen because they represent the universities own expression of the content of their research and educational programmes. In addition, we consulted literature from the field of Media Studies, and particularly a review of Danish media research by two prominent Danish media researchers (Hjarvard & Søndergaard, 1998), as well as a handbook of methodological approaches to media research (Jensen, 2002), and a selection of 20 student papers (see a listing of the student papers in Appendix 3).

According to Hjarvard and Søndergaard (1998, p. 265) the post-war expansion of the mass media provided an inducement for development of media research in Denmark. In the late 1960s and the early 1970s systematic and continuous research in mass media began to develop in the literature and film departments of Danish universities, and media research became institutionalised in the 1980s within the faculties of the humanities. Today, Media Studies have developed into one of the most popular Danish university educations, reflected in the high entry grade averages required by student candidates (e.g., Tilmeldingssekretariatet, 2006, 2007).

Media Studies is the study of mass media in their historical, cultural and social context. This includes audiovisual media (e.g., Bondebjerg’s (1993; 2006a) focus on television broadcasts), traditional media (e.g., Poulsen’s (1999) investigation of newspapers), as well as new media (e.g., Finnemann’s (2005) focus on the Internet). The field comprises all types of communication (e.g., art, entertainment, news, commercials, and interactive communication). The focus is contemporary as well as historical (e.g., Jensen, 1996-2003; Hjarvard, 2006b), and Media Studies is predominantly concerned with Danish mass media. Foreign mass media (for instance, Scandinavian, European or American) are also included in some of the comparative research (e.g., Bondebjerg, 2006b).

Hjarvard and Søndergaard state that Danish media research has a primary focus on television broadcasts (see e.g., Agger, 2005; Frandsen & Bruun, 2007), and the media departments at Copenhagen University and Aarhus University are emphasised to give specific attention to television broadcasts (Hjarvard & Søndergaard, 1998, p. 268). This knowledge contributes to our selection of user target group. The focus in Media Studies can be ascribed three main traditions: 1) reception studies; 2) aesthetic analysis; and 3) media structures and the contemporary development of media systems. Empirical reception studies are mainly qualitative studies of the audience’s experiences of the media. The aforementioned cognitive film theories (e.g., Grodal, 1997) is an
example of a theoretical approach within the audiences responses to visual media. The reception analyst often uses computer software for analysis of reception data (e.g., data from tns Gallup’s TV-Meter). This supports Wiberley and Jones’ statement that the low usage of technology in the humanities is due to a limited value of technology, but when technology is valuable for advancing research, it is naturally also used in the humanities (Wiberley & Jones, 2000, p. 426). Aesthetic analysis is for instance concerned with the investigation and definition of different media genres (e.g., Bondebjerg, 1993). Within the tradition of media structures and the contemporary development of media systems, examples are the focus on local electronic media, or consequences of the late abolishment of the Danish television monopoly.

Media Studies is concerned with the preparation, mediation and use of mass media, and though placed within the humanities the field comprises inter-disciplinary elements, and qualitative as well as quantitative methods are applied in the field. The field applies traditional text and language analytical tools for analysing the content of mass media, alongside with psychological and sociological tools for analysing viewers’, listeners’, and/or readers’ reception and use of media, and cultural and social science theoretical tools for analysing cultural developments and trends, and the media’s place and function in our society (Jensen, 2002).

This ends this chapter’s outline of the empirical context of the present work. The chapter provides background knowledge for the succeeding four chapters, Chapter 6 concerned with our methodical approach, and Chapters 7, 0, and 9 which reports the results of our empirical investigation. Knowledge about the present users of television broadcasts at the State and University Library is a rationale behind our selection of scholars and students in Media Studies as the user target group for the empirical investigations. We do acknowledge that the user’s possibility for access to the collection of television broadcasts affects whether or not the user addresses a request to the library. Therefore, neither the amount nor the types of information requests is seen as an indication of the information needs which could be fulfilled in a future broadcast retrieval system. Consequently, construction of future effective electronic access to television broadcasts should not purely be based on analysis of current information requests, current users and/or current use. This is one of the issues addressed in the following chapter on our methodical approach for the main empirical investigations.
6 Methodical approach

The overall goal of the present work is to gather knowledge about users’ information needs, search entry preferences, and application of relevance criteria in the context of television broadcasts, in order to develop and construct effective broadcast retrieval systems. We aim at reaching this goal by conducting explorative investigations of these three aspects of information seeking behaviour for a group of users in the contexts of television broadcasts. In the preceding Chapter 5, we outline the context for the present work, and we report on our selection of a user target group for our empirical investigations. The empirical investigations are conducted by applying several different data collection methods, and the aim of the present chapter is to describe and explain our methodical approach, as well as to place our approach within a broader methodological frame.

This chapter comprises six sections. In the 1st section, Section 6.1 we describe our methodological background. Subsequently, four sections describe our methodical considerations regarding the exploration of aspects of users’ information seeking behaviour in a television broadcast context. In the second section, Section 6.2 we describe our approach for gathering knowledge about the domain in question. Section 6.3 outlines the methodical considerations regarding data collection and data analysis in our web questionnaire survey, while the 4th section, Section 6.4 regards our methodical reflections and choices in respect to in-depth follow-up interviews. Section 6.5 briefly outlines the group discussion of preliminary results with scholars. Finally, Section 6.6 provides an overview of the methodical approaches, including a visual illustration of the time of execution of the different elements of our methodical approach.

6.1 Methodological background

Traditionally, user issues in LIS have been studied from a reductionistic system-oriented point of view, but from the 1980s more holistic user-oriented approaches have been introduced to the field (Ingwersen & Järvelin, 2005, p. 87). As part of the introduction of holistic user-oriented approaches, Wilson states that knowledge obtained from real users is necessary in order to design effective IR systems, and that “[q]ualitative research seems particularly appropriate to the study of the needs underlying information-seeking behaviour” (Wilson, 1981, p. 11). Recently, Ingwersen and Järvelin follow this line of thought and “[…] consider qualitative methodology a
necessity in order to further develop the understanding of cognitive actors in the IS&R [Information Seeking & Retrieval] process” (Ingwersen & Järvelin, 2005, p. 109). Further, Ingwersen and Järvelin emphasise the benefits of employing a triangulation of methods in order to convey more aspects of the phenomenon in question, and to increase the reliability and validity of the investigation (Ingwersen & Järvelin, 2005, p. 93). The grounded theory methodology and the triangulation of methods fits well within the holistic cognitive framework for information seeking and IR, which is the overall theoretical approach for the present work, as outlined and discussed in Chapter 2.

In the present work, we make use of triangulation of methods, since we use quantitative and qualitative methods in our data collection as well as our data analysis. To be specific, we employ a semi-structured web questionnaire survey, open-ended interviews and group discussion for data collection in our investigation of aspects of information seeking behaviour. In our data analysis, we use statistical analysis, and inductive and dynamic coding. This triangulated research approach is described in more details in the following sub-section.

6.1.1 Triangulation of research strategies

Structured quantitative empirical surveys are the predominant research strategy in LIS, which it has been throughout the history of the field (e.g., Järvelin & Vakkari, 1990; Powell, 1999; Julien & Duggan, 2000; Wang, 2001). The predominance of survey methods is ascribed its flexibility and ease of use. The structured data collection supports the collection of a vast amount of data, in addition to easing the subsequent data analysis (e.g., statistical analysis). However, survey methods are also accused for only conveying parts of the phenomenon under investigation (Järvelin & Vakkari, 1990, p. 409), and consequently, we also employ a qualitative research strategy in the present work.

Inductive grounded theory (Glaser & Strauss, 1967) is characterised as the qualitative methodology (Ingwersen & Järvelin, 2005, p. 88), and it is used in several studies of user behaviour in LIS (e.g., Schamber, 1991b; Ellis, 1993; Spink, 1997; Pharo, 2002). Grounded theory has emphasis on inductive generation of theories, models and/or hypothesis from raw empirical data, rather than the deductive test of already established theories, models and/or hypothesis.

Seldén (2005) discusses the use of grounded theory, and he points out that rigorously following the line of grounded theory produces some problems. For instance, the fact that researchers should be attentive towards prevention of imposing a priori categories gives a problem (Seldén, 2005, p. 117). How can a researcher
commence a research project with a *tabula rasa* (no a priori categories)?, and how can a researcher avoid using a priori categories if he or she is familiar with similar research? We believe the answer is that the researcher should be attentive towards the problems of a priori categories, but that the researcher naturally needs to be familiar with previous research in the area. In this way, the researcher does not have *tabula rasa*, but instead aims a being inspired by previous literature in order to analyse the data. We follow the recommendation by Strauss and Corbin (1998, p. 44) and we use the examples in the literature “[…] to stimulate our thinking about properties or dimensions that we can use to examine the data in front of us” (Strauss & Corbin, 1998, p. 44). Familiarity with the relevant literature is important in order to enhance the researcher’s sensitivity to nuances in the data, but at the same time, it may block the researcher’s creativity, which is an important element of the methodology. Therefore, the researcher should be attentive towards the concepts or categories emerging from the analysis, and make sure they are really emerging from the data and not from the literature (Strauss & Corbin, 1998, p. 49). Further, the emerging concepts should be compared to those in the literature in order to pinpoint similarities and difference, and in this way contribute to the development of the field. Consequently, we do not follow Seldén’s line of reasoning.

The research strategy in the present work is based on the grounded theory methodology. An element of grounded theory is not to follow methodical procedures rigidly, but to use the research procedures as inspirational sources, keeping the overall aim of the methodology and not least the research project in sight (Strauss & Corbin, 1998, p. 46). Triangulation is not an explicit part of grounded theory, but it is in line with the thoughts underlying the methodology, since an important element of the methodology is to understand a phenomenon from different perspectives by applying different techniques (Strauss & Corbin, 1998, p. 44).

Our methodological background is grounded theory, but in our methodical approach we are inspired by several qualitative research techniques, including anthropological inspired participatory observation (e.g., Bogdan & Taylor, 1975), critical incident interviews (e.g., Flanagan, 1954), naturalistic research (e.g., Lincoln & Guba, 1985; Mellon, 1990), and sense making (e.g., Dervin, 1983b). The purpose of the present empirical research is exploratory in nature. Different methods reveal different aspects of the empirically investigated reality. Since no method can completely reveal all relevant features of an empirical reality, Denzin (1970, p. 13) recommends a triangulation of methods to increase the validity of the results. Triangulation is applicable for reaching a broader perspective of the phenomenon under scrutiny, and we find it especially relevant in relation to research of an exploratory
character. We apply what Denzin (1989, p. 244) denotes the between-methods or across-methods triangulation with the underlying principle being that the flaws of one methodical approach are the strengths of another methodical approach. By using a triangulation of methods, we focus on the strengths and at the same time limit the deficiencies of each method.

Several methods can be applied for gathering of empirical data within grounded theory. Wilson argues that all data collection methods are “[…] ultimately substitutes for the fundamental method of observation” (Wilson, 1990, p. 5). Because direct observation generally is unfeasible in studies concerned with users’ inner thoughts, indirect observation through self-reflecting techniques can usually be used to replace direct observation. Several indirect observational methods within the inductive grounded theoretical methodology can be applied to study information seeking behaviour. The methods employed in the present work are the focal point of the succeeding four sections, commencing with a focus on our informal data collection.

6.2 Domain knowledge: informal information gathering

An important element in our methodical considerations has been not to decide upon specific approaches or methods until we had gained substantial knowledge about television broadcasting and archiving (including indexing), the users and use of archived television broadcasts, and the field of Media Studies. Therefore, the first task undertaken is to gain insight into these three aspects through information gathering which largely has been informal in nature. The aim of the informal information gathering is not to produce empirical data as such. Rather, the aim is first and foremost to gather necessary intellectual knowledge for the identification of a user target group, selection of a sample, and analysis of the collected empirical data. The purpose of this section is to provide the reader with insight about the informal information gathering undertaken with respect to these three areas. This section is a continuation of our description of the empirical contextual settings, and there are some overlaps between the present section and Chapter 5. Our methodical approaches as well as some of the knowledge gained in the three areas are presented in the following three sub-sections, respectively.

6.2.1 Television broadcasting and archiving

Insight on the historical and contemporary broadcasting of television in Denmark is gathered from the literature (e.g., Ahm, 1972; Hjarvard, 2006b), and outlined in Section 5.1.
In Section 5.2 we describe the main archives for Danish television broadcasts. Insight on these archives has been gained in somewhat subtle terms. Very few archives of television broadcasts are publicly available, and most of the knowledge on the existence of the different archives is acquired through our increasing knowledge about the field of Media Studies. More specifically, the interviews with academics in Media Studies (see Section 6.4), a two month research visit at a Danish University Media Studies department (see Section 6.2.3), and three visits at television stations are the main sources for gaining knowledge about television archiving in Denmark. In addition, we worked at the State and University Library approximately once every fortnight. In this way, we are able to get an idea of the routines at the archive, and gain insight to archiving of television broadcasts. Being at the State and University Library is also relevant for gaining insight into the State and University Library’s collection of television broadcasts (Section 5.2.1). Further, we gained knowledge about the need for and use of television broadcasts presented in Section 6.2.2, below (for a methodological discussion of this kind of information gathering, see Section 6.2.3).

In order to increase our insight into indexing of television in operating archives, we arranged three visits to television broadcast archives, since these archives hold invaluable knowledge on the subject. At two occasions, we visited DR. The first visit is in the initiation of the research project (20th of April 2005), at which point methodical considerations are our prime focus. The second visit is towards the last part of the project (25th of January 2007), at the time when the empirical data is collected and we have conducted initial data analysis. In-between (13th of June 2006) a visit is arranged at Radio Telefís Éireann (RTÉ), which is Ireland’s national radio and television broadcaster, and the Irish counterpart to DR. At the three visits, the data models and indexing practices at the stations broadcast archives are discussed. Further, we observed indexing in practice and we are handed examples of indexing from each station. DR is considered to be one of the leading television archives in Europe and RTÉ is inspired by DR in their archival approach. Investigation of the indexing practices at television stations is not the purpose of the present work, however these visits give important insight on how this type of information objects are handled in the professional community. That is, the community that has been prominent in dealing with television broadcasts. It is important to acknowledge that the purpose of the archives at television stations is to comprise production archives and not research archives (Hjarvard, 2006a, p. 22). The difference in purpose between production units and researchers will naturally affect the construction of the archive, and these contextual and situation aspects are motivational aspects underlying the present work.
Prior to the present work, the State and University Library conducted four interviews with researchers about their use of and work with moving images. The four interviews are analysed in order to gain initial insight as to the researchers need for and use of television broadcasts, and they are analysed and discussed in the following section.

6.2.2 The need for and work with television broadcasts

The four interviews originate from a project concerned with the use of commercials in scholarly research, and they are considered a supplement to our analysis of genuine requests presented in Section 5.3. We are neither involved in the design nor in the conduction of the interviews, but nevertheless we find the interviews useful in respect to the present work. Unfortunately, the interview guide used in these interviews has been lost, and we are consequently unable to produce the guide in an appendix. However, in a report which analyses the interviews, Lund (2004) states that the interview design is largely inspired by Holtzblatt and Jones’ (1991) contextual inquiry interview technique. This technique helps people to crystallize and to articulate their work experiences (Holtzblatt & Jones, 1991, p. 178), and it is a technique that is closely related to so-called participatory design in the human-computer interaction community (e.g., Bødker, 1996). In order to avoid any terminological confusion with the nine interviews conducted by ourselves, we denote these four interviews the State and University Library interviews, or simply the SUL interviews.

The SUL interviews compromise two hours and 50 minutes of audio. The transcription of these four SUL interviews is conducted in the same manner as our nine interviews with academics, and the approach is described in Section 6.4.3. The purpose of the SUL interviews is to examine and gain insight as to how scholars conduct research using moving images as their primary research objects (see e.g., Lund, 2004; Kirkegaard & Borlund, 2006). Despite the focus on commercials in the State and University Library project, the SUL interviews focus on moving images in general, and television broadcasts in particular. Therefore, the SUL interviews are very useful as a supplement to the information gathering mentioned above. We are aware that the number of interviewees is small in relation to generalisation of the results, but we consider the SUL interviews from an explorative perspective. That is, we use the SUL interviews to get an idea of the issues to be aware of in the present deeper investigations of users’ cognitive perceptions of television broadcasts.

The interviewees’ use of moving images is mainly oriented towards television broadcasts, but the scholars also use moving images that was not broadcasted on television (e.g., cinema commercials or video recordings of doctor-patient dialogues).
In our analysis of the SUL interviews, we focus on the scholars’ statements concerning television broadcasts, only. Two of the four scholars are affiliated Aarhus University, one is affiliated Copenhagen University, and one is affiliated University of Southern Denmark. Three of the four interviewees are affiliated departments within the humanities while one is affiliated a social science department. Two scholars are affiliated Media Studies departments, one is affiliated a Department of Comparative Literature, and one is affiliated a Department of Marketing and Management. This distribution is comparable to the distribution of users requesting programmes from the collection in 2004, as reported in Section 5.3.1, and by Kirkegaard and Borlund (2006).

All four interviewees have a focus on broadcasted television, though naturally their focus is somewhat different from each other (e.g., the rhetoric in debate programmes, visual effects in television commercials and product marketing, the role of television in our culture, or reality TV and its connection to reality). Similarly, the number of broadcasts the interviewees include in their academic work tasks range from one special selected prototypical example to the analysis of more than 100 broadcasts. We are able to identify a connection between the interviewees’ research interests and the content of the Danish national collection of television broadcasts. However, none of the four interviewees requests programmes from the collection, though the interviewees’ research interests indicate that the collection seems a natural place for them to request broadcasts for their academic work tasks. The interviewees did not explain why they do not use the collection, and the interviewer did not prompt for the information. All four interviewees describe other channels for collecting their primary data, e.g., home recording of television broadcasts, ‘small’ collections at their department, or collaboration between colleagues. An example is given in the following quote:

**INTERVIEWER:** “Would improved access mean anything for the way you work? Would you use more broadcasts?”

**INTERVIEWEE:** “I would use a lot more examples, a lot more diversified examples. Now I use the same examples a lot, namely the examples I can record on my own VCR”.

*Quote 1*

This is comparable to the ‘normal’ information seeking behaviour for scholars, as discussed in Section 3.3.2. However, we assume the deficiencies of electronic access to

---

*Whenever we reproduce quotes from the interviews we present an English translation of the quote. The original Danish matching part of the SUL interview quote is reproduced in Appendix 9. See Section 6.4.3 about transcription prior to reproducing and translating the quotes.*
the collection at the State and University Library affect whether users make the effort to consult the collection in the first place.

The four interviewees state that they often have a very clear idea of the television broadcasts they need. With reference to Kuhlthau’s model of the Information Search Process (e.g., Kuhlthau, 2004) (e.g., see Section 3.2.3), the four scholars’ information needs for moving images are comparable to the post-focus formulation stages of the model. The four scholars are experts in their field of research. As work tasks generally are perceived as less complex for experts than for novices (Byström, 1999), experts rapidly generate a focus, and information seeking is consequently structured and directed towards the user’s perception of the work task (Kuhlthau, 1991). Therefore, the information requests from scholars to the Danish national collection of television broadcasts could be expected to be structured and directed towards the user’s perception of his or her work task.

The four scholars all describe a need for easier access to television broadcasts. The scholars understand easier access in several ways e.g.:

1) Access to digital copies of television broadcasts;
2) Access to television broadcasts from their own work station;
3) Instant access to television broadcasts; and
4) Electronic access to the collection of television broadcasts in the form of search entries.

One interviewee states that a less good example is preferable for analysis if it is easy to access, rather than to wait for the perfect example.

“To me, and it is not the same for all scholars, because some have a theoretical starting point, which I also do in a few courses, but usually I have a starting point in the primary data objects. That is, in my work with Nordic thrillers for instance, the first step is to find e.g., which Swedish thriller it is possible for me to obtain a video copy of, and a copy that I am allowed to use in my teaching. Then, I see that the one I can actually obtain is number four on my ‘wish list’, which is often the case. This situation is completely different from research in a textual setting, because here we always use the optimal primary data. After obtaining a copy of a broadcast I identify relevant methodical theories and supplemental texts surrounding the broadcasts. The primary data object decides, and this is due to the terms of access to television broadcasts”.

Quote 2

The interviewee exemplifies it further:
"That is one of my returning problems, that I am compiled to be content with mediocre examples. I know the optimal examples exist, but in the way the system works today it is not possible for me to spend the time needed to identify and retrieve these optimal examples, and often it is not even a practical possibility to obtain a copy of the broadcast prior to commencing my teaching”.

Quote 3

That is, the interviewee selects the television broadcasts that, with least effort, are at disposal when it suits the interviewee. A heavy workload is given as the rationale behind the selection. A similar attention toward effective usage of time is found in the other SUL interviews, which is in line with Wiberley and Jones’ (2000) report on humanist scholars’ information behaviour. They find that humanist scholars are very aware of effective usage of their time. For instance, the expected time saved by using a specific technology is considered in relation to the time needed in order to master the technology (see e.g., Section 3.3.2).

In one of the four SUL interviews, the scholar mentions that the problem of retrieving television broadcasts is harder the older the broadcasts are. This should be seen in light of the deficiencies of lack of electronic access to the collection, which favours requests of recent broadcasts at the expense of requests of older broadcasts, as mentioned in Section 5.2.1.

The knowledge gained in these interviews is important in order to gain insight about the use of television broadcasts in academic work tasks, and along with knowledge about the domain in question (Media Studies) the knowledge is useful for collecting and analysing our empirical data. This gives emphasis on the need for a broadcast retrieval system. In the following section, we outline our approach for gaining knowledge about the field of Media Studies.

6.2.3 The field of Media Studies

Due to the fact that we as researchers of the present study are previously unfamiliar with the domain in question, several steps are taken in order to establish general knowledge about the domain, and in particularly knowledge about the situation at the two university departments in question. Such knowledge is somewhat difficult to pinpoint, and it has consequently been collected in several informal ways. This is a continuation of the gathering of knowledge about the field of Media Studies, which is outlined in Section 5.3.3. This knowledge is collected by familiarising with the literature in the field of Media Studies in Denmark (e.g., Henriksen, 1995; Bruun, Frandsen &
Søndergaard, 2000; Hjarvard, 2001; 2003; 2005; Jensen, 2002; Jerslev, 2002; Grodal, Larsen & Laursen, 2004; Agger, 2005), as well as studying curricula and syllabuses from the two departments. Further, we browse through a selection of 20 student papers to gain insight in the field from the students’ perspective (see Appendix 3 for a list of the 20 student papers).

To gain a broader understanding of the field of Media Studies a two month research visit at the Department of Information and Media Studies at Aarhus University took place. This visit gives invaluable insight into the field of Media Studies. Here we are able to observe and participate in the life at the department, including scientific discussions. Our research visit is comparable to participatory observation in Bogdan and Taylor’s (1975) terminology, since we believe we are considered as neutral figures, with no alliances that might harm the investigation. In this way, we as the participant observer are meet with free and open exchange of information. Participatory observation is a research approach, which allows the investigator to become part of the community in order to understand what is not usually made explicit. Kvale (1996, p. 104) states that participant observation, including informal interviews, is appropriate when the researcher is interested in implicit meanings, tacit knowledge and/or taken-for-granted assumptions of a group or a culture.

Our role in this investigation can be described, as similar though not analogous to what Lave and Wenger (1991) from a learning perspective denote the legitimate peripheral participant. This role is powerful since it enables us to be embodied in the practice and learn the tacit knowledge of the community, which does not necessarily get articulated but has to be absorbed (Jordan, 1994; 1996). In this role, the observer is allowed to ask almost any question, and participate in meetings, lunches, and other story telling sessions. Since research in the humanities in general can be characterised as a cognitive and personal affair conducted inside private offices, the work practices are rather difficult to understand for an outsider. This approach gave us an opportunity to gather knowledge about the field of Media Studies, which would otherwise be out of our reach, but it did of course not reveal these cognitive and personal aspects.

An underlying purpose of the research visit (see Section 6.2.3) is to build up an interest for the research project, and to allow us to become familiar with the respondents. Hereby, increasing the respondents’ motivation for participation, and decreasing the risk of misinformation and perceptual distortions in the collected data, and increasing the credibility of our work (Foster, 2005, p. 2). These motivational aspects are also relevant in relation to our data collection using a web questionnaire. This is touched upon in the following section on our methodical considerations regarding data collection and data analysis in the questionnaire survey.
6.3 Semi-structured web questionnaire survey

The aim of this section is to initiate the reader in our considerations regarding the use of a semi-structured questionnaire for collecting data about the three aspects of information seeking behaviour under investigation in the present work. The questionnaire is depicted in Appendix 4. As mentioned in Section 6.1 the questionnaire is one of several empirical methods we employ. The purpose of the questionnaire is to generate initial knowledge about the users’ information needs, preferred search entries and application of relevance criteria. In this way, the questionnaire is related to the first research area on aspects of users’ information seeking behaviour in relation to television broadcasts.

One of the advantages of using a web questionnaire is its ability to be distributed to a large amount of respondents with relatively few resources. Since we wish to reach as wide a variety of the user target group as possible, we find a web questionnaire survey applicable. Employing an electronic web questionnaire is further advantageous for the subsequent processing of the collected data. That is, the questionnaire is programmed as an Active Server Page (ASP), and constructed to automatically add the respondent’s answer to a Microsoft Access database. The database is formatted into SPSS and Excel readable format and the responses are swiftly available for subsequent statistical analysis.

Lack of control of respondents is one of the main limitations of web surveys, e.g., by having multiple responses from the same respondent, or responses from persons outside the intended population (Steffensen, 2004). By contacting respondents personally by e-mail, inviting them to indicate personal contact information, and crosschecking responses, we believe this methodical problem is of a tolerable magnitude.

6.3.1 Design and construction of the questionnaire

In our design and construction of the questionnaire we are inspired by the methodical recommendation put forward by Buckingham and Saunders (2004), and Frankfort-Nachmias and Nachmias (1996). The questionnaire consists of eight sections (see Appendix 4), which we describe in the following.

The questionnaire is constructed with a mixture of closed and open questions. Closed or pre-coded questions are advantageous for several reasons. It is easier for the respondent to select an item, than describing it with his or her own words. It saves time in the later formatting and analysis phase, since respondents are only dealing with predefined categories. Lastly, it ensures that the researcher gets the type of data which are
intended (Buckingham & Saunders, 2004, p. 74). The limit of pre-coded questions is that respondents are not allowed to go outside the pre-defined categories. One might argue that the use of pre-defined categories is problematic in an exploratory behavioural study, since it restricts the range of possible answers. Indeed, this is one of the arguments against the traditional reductionistic survey approach applied in LIS (Ingwersen & Järvelin, 2005, p. 87), and since we are claiming an exploratory nature in the present investigation, the argument against pre-coded categories seems even more relevant. However, the pre-coded categories have been chosen purposefully, in order to inform the respondents of the types of categories which could possible be made available in a future broadcast retrieval system. This is important since the possible features in a future broadcast retrieval system might be intangible for most of the respondents. Further, each section of the questionnaire contains a ‘free text’ field or open questions, in which the respondents are invited to list what they believe to be missing categories, and the 8th section invites additional comments from respondents. Furthermore, the use of follow-up interviews and group discussion adds an additional dimension to the exploratory nature of our investigation (see Section 6.4).

In Chapter 5 we describe our rationale behind the selection of academics in Media Studies as the user target group. It should be emphasised that this group is not considered as one heterogenic mass. That is, in relation to research on the information seeking behaviour of university students and faculty (e.g., Barrett, 2005) the 1st section of the questionnaire is designed to enable us to distinguish between students and faculty. In addition, we are able to distinguish between respondents from the different departments, in order to be able to draw conclusions on the groups separately (faculty and students), and identify differences between the different departments.

The 2nd section invites the respondents to describe a work task, which includes a need for obtaining television broadcasts. Secondly, the section requests the respondent to specify whether the needed broadcasts were obtained or not. In case of a positive reply, the respondent is asked to describe how and where the broadcasts were identified and obtained.

The 3rd section of the questionnaire concerns the types of television broadcasts respondents typically would need access to. A common Danish categorised typology for television broadcasts is not available, but in connection to a 2001 investigation regarding the selection of television programmes for digitization, a so-called genre typology of television and radio programmes is constructed (Bork et al., 2002; Hielmcrone, 2002). The typology is a spin off of the investigation, and representatives from DR, TV2, the State and University Library, and Media Studies scholars construct it in a joint effort. In addition, the typology incorporates ongoing work by the European
Broadcasting Union (EBU) on a future European collaborative system for categorisation of television broadcasts. The fact that the typology is reached in collaboration between the main players in the field is considered a strength of the typology, and we are consequently inspired by the typology in our selection of categories. From the typology’s list of 8 main categories and 79 sub-categories, we comprise a list of 17 categories, containing all main-categories and a selection of the sub-categories where they are considered relevant. The respondent can tick off as many of the 17 types listed, as he or she finds relevant. The categories are listed alphabetically.

In the 4th and 5th section, the questionnaire comprises a list of 20 search entries and a list of 24 relevance criteria. The two lists are nearly identical, except for description of image sequences, summary, broadcast clips and thumbnails added to the list of relevance criteria. The lists are comprised by examining questionnaires employed in previous studies (Sandom & Enser, 2002; Hertzum, 2003), knowledge about the state of the art in video retrieval (as outlined in Section 4.2.2), and features available in operational systems. The operational systems we are inspired by are the Open video Archive (Interaction Design Laboratory, 2006), BBC’s Open News Archive (BBC, 2006), and the British Television and Radio Index for Learning and Teaching (TRILT) (British Universities Film & Video Council, 2006).

Television broadcasts are transmitted in a non-stop stream each day (see e.g., an example of a weekly television schedule in Appendix 5), and in cases with 24 hours broadcasting the transmission has no beginning or ending. In order to provide access to the transmitted broadcasts, the transmission has to be broken down into tangible indexable units. In a traditional library context such tangible units are greatly simplified the physical information objects, e.g., a book, a CD, or a floppy disk. In the fourth section of the questionnaire we provide four level of granularity of such indexable units (serials, programmes, features, and sequences), and the respondents are asked to mark the level(s) of granularity which they find important for searching and retrieval of broadcasts. This knowledge is important in order to gather knowledge about the level of granularity of the indexable units in a future broadcast retrieval system. In the television context, we have identified four levels of granularity of such indexable units: serials, programmes, features, and sequences. The three latter are derived from the literature as described in Section 4.2.2, and depicted in Figure 4.1. The serial level of granularity is inspired by the access to a series of documents in a traditional library context (e.g., providing access to each book in the Library and Information Science Text Series). In contrast to the three other level of granularity, the serials level does not concern broadcasts transmitted in direct proximity of each other. Indeed, the individual broadcasts in a serial are usually transmitted on a daily or weekly basis. Similar to the
third section, respondents can tick off as many of the listed categories as they find relevant.

The 6th section enquires about the respondent’s previous experience with television broadcasts from the collection at the State and University Library. This section is relevant for learning about the respondent’s familiarity with the present ‘system’ for obtaining broadcasts, as well as the distribution of present and potential users of the future broadcast retrieval system.

In the 7th section, we collect contact information, which is relevant for getting in touch with potential interviewees. Providing these personal information is voluntarily, which is emphasised in the questionnaire. In this way, we gather essential information for selecting interviewees and at the same time allows for anonymity in responses, which is important in order to increase the respond rate.

The final section, section 8, collects any comments the respondent may have regarding the research project. This section is included in order to collect data, which we might otherwise have missed.

In addition, upon submitting the filled in questionnaire, the respondent is directed to a page, which thanks him or her for taking the time to answer the questions. This page is depicted with the questionnaire in Appendix 4. Prior to launching the survey, we ran a pilot test, which is described in the following section.

6.3.2 Piloting

In order to learn whether the questionnaire survey is properly designed and constructed with respect to our research questions, we conducted a pilot study at the Department of Communication and Psychology, Aalborg University, and the Department of Communication, Journalism, and Computer Science, Roskilde University (now Department of Communication, Business, and Information Technologies). From a methodical perspective, it is not possible to pilot test with the user target group of the main investigation. The departments are chosen because the fields of Communication and Journalism are closely related to the field of Media Studies, and because our analysis of genuine requisition forms received at the State and University Library shows that several users are connected to these fields of study (see Section 5.3.1). By studying the departments web pages, including project description, and lists of publications from individual researchers, we identify 21 scholars and students whom we consider to be relevant to include in the pilot. Each individual academic receives a personal e-mail, requesting his or her help in piloting the questionnaire (see the e-mail invitation in Appendix 6). The e-mail includes a link to the questionnaire, and the respondents are given approximately two working weeks to respond. After the initial deadline, a
friendly reminder is distributed, and the respondents are given an additional working week to respond. Ten people respond to our invitation, and based on the answers and comments from these academics, a few changes are made to the original questionnaire prior to conducting the survey with academics in Media Studies. Our considerations regarding sample selection for the main survey is outlined in the succeeding section.

6.3.3 Collecting the data: respondents and invitations to participate

The selected user target group is academics in Media Studies, which we outline in Section 5.3. The analysis of the genuine requisition forms discussed in Section 5.3.1 discloses the present users, and their use of the Danish national collection of television broadcasts. We do believe that the collection contains broadcasts, which potentially could be relevant to a larger group of users than the very committed ones who put up with the present lack of search entries. Consequently, we do not want our study of aspects of users’ information seeking behaviour to be narrowed to the present users of the State and University Library. Instead, we intend the study to include present as well as potential users with real information needs. In our analysis of the four SUL interviews reported in Section 6.2.2, we found that none of the four scholars retrieved their television broadcasts from the State and University Library. This illustrates the importance of broadening the sample of the web questionnaire survey to present as well as potential users of television broadcasts.

One of the strong points of a web questionnaire survey is the ability to reach a large group of people with relatively few resources. This advantage is one of our main reasons for the methodical choice of using web questionnaires. We contacted the two departments of Media Studies, which comprise our user target group and requested permission to e-mail an invitation to participate in our questionnaire survey to all scholars and students at the departments. The departments were very co-operative and they either provided us with mailing lists for all scholars and students or distributed our invitation via such mailing lists.

The e-mail invitations along with friendly reminders are found in Appendix 7. The invitations contain a short presentation of the aim of the project and the questionnaire, the deadline for answering the questionnaire, as well as contact information to the main responsible researcher. As recommended by Frankfort-Nachmias and Nachmias (1996, p. 231) the invitation is kept in an altruistic style to increase the respond rate. The respondents were given approximately one working week to respond. Approximately two days after the deadline a friendly reminder was distributed via the mailing lists and a new deadline is set which gives the respondents additionally four days to respond. In total, respondents had approximately two working
weeks to respond, and responses that were received up until one week after the last deadline are also included in our data.

Unfortunately, the survey experienced a few technical problems. First and foremost, the sign ‘ is not allowed in the free text fields of the ASP questionnaire. This problem was discovered immediately after distributing the e-mail invitations, and instantly a message instructing respondents not to use the specific sign is inserted above all free text fields in the questionnaire, as well as in the following friendly reminders. Further, the server holding the questionnaire was not available in approximate three hours during one day of data collection. Such technical problems will naturally have an affect on the responses received. However, examination of log files shows that error messages appeared in 15 instances, only. Consequently, we do not believe the technical problems have had a substantial effect on our data collection. The subsequent section provides more details about the responses we received in our survey.

6.3.4 Respond rates

Table 6.1 depicts the distribution of responses received and their distribution on department and academic status. We received 108 responses to the invitation distributed to 1,785 e-mails, which equals an overall respond rate of 6.1%. At first glance a non-respond rate of 93.9% might be considered problematic, and induce a bias, which affects the validity of the study. That is, generally speaking non-responses are considered problematic since non-respondents might differ considerably from respondents (Frankfort-Nachmias & Nachmias, 1996, p. 232). However, due to several reasons we do not consider the respond rate a problem for the present work.

Table 6.1: Respond rate for web questionnaire survey, distributed in accordandence to academic department and academic status.

<table>
<thead>
<tr>
<th>Department</th>
<th>Responses</th>
<th>Population (≈)</th>
<th>Response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media Studies, Aarhus University</td>
<td>52</td>
<td>466</td>
<td>11.2%</td>
</tr>
<tr>
<td>Scholars</td>
<td>13</td>
<td>33</td>
<td>39.4%</td>
</tr>
<tr>
<td>Students</td>
<td>39</td>
<td>433</td>
<td>9.0%</td>
</tr>
<tr>
<td>Film and Media Studies, Copenhagen University</td>
<td>56</td>
<td>1,066</td>
<td>5.3%</td>
</tr>
<tr>
<td>Scholars</td>
<td>9</td>
<td>65</td>
<td>13.9%</td>
</tr>
<tr>
<td>Students</td>
<td>47</td>
<td>1,001</td>
<td>4.7%</td>
</tr>
<tr>
<td>Scholars</td>
<td>22</td>
<td>98</td>
<td>22.5%</td>
</tr>
<tr>
<td>Students</td>
<td>86</td>
<td>1,434</td>
<td>6.0%</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>1,785</td>
<td>6.1%</td>
</tr>
</tbody>
</table>
First of all, the respond rate is far from a stringent figure. The figure builds on the number of e-mail addresses that are contained in each mailing list. This figure is reported on by each department’s mailing list administrator, and the administrators indicate that the figures are equivocal. For instance, one person can be registered with several e-mail addresses, which especially students take advantages of. If a person is associated to more than one class that person’s e-mail(s) will appear several times in the mailing list, and hence the same individual is again represented more than once. Secondly, the e-mail invitations are distributed to any person (scholar or student) with affiliation to Media Studies, which for instance includes exchange students who might not be able to understand the Danish invitation.

Frankfort-Nachmias & Nachmias (1996, p. 231) state that one way to increase the respond rate is to select a sample of motivated respondents, who have a personal interest in, or familiarity with, the topic in question. By using the mailing lists to distribute our invitation to participate, we do not narrow our population to motivated respondents, but include all academics at the two departments in question. In this way, the survey is distributed to all potential users in the target group, with respect to the aim of the survey, only, but at the cost of a decrease in respond rate. This is for instance seen in the e-mails received, and personal contacts made from respondents, expressing that they find the research important, but that they have no experience in use of, or need for television broadcasts, and consequently, they have not answered the questionnaire. The difference between the respondent and non-respondent might be the non-respondents lack of familiarity with a need for or experience with television broadcasts in relation to an academic work task.

Despite the equivocal nature of the respond rate, it does indicate some difference between departments as well as between the respondents’ academic status. The respond rate at the department at Aarhus University has the highest level and in particular the rate for scholars. The web questionnaire survey is instigated approximately two weeks into our two months research visit at the department (see e.g., Section 6.6), and our presence no doubt affects the number of responses from the scholars. The higher respond rate from scholars compared to students can be explained by the fact that scholars may be more liable to consider the benefits of research, since research is their living. In addition, since they have been involved in more projects they are more likely to have experience with or a need for television broadcasts in relation to their work. Further, students are more likely to be represented several times in the mailing lists, which hence affect the respond rate for this group in a decreasing fashion.
Hence, we do not find the low respond rate to be problematic since the purpose of the web questionnaire survey is to collect data from a wide variety of present as well as potential future user of television broadcasts. In this respect, a low respond rate is to be expected. With reference to the 229 individual academic users who requested broadcasts from the Danish national collection of television broadcasts in 2004 (see Section 5.3.1, and Table 5.2), we consider the 108 responses to our questionnaire to be a fair amount of responses. The 108 responses we received are analysed by applying statistics as well as qualitative analysis. The various analyses we apply are described in the following section.

6.3.5 Analysing the collected data

The web questionnaire survey generates quantitative as well as qualitative data. The main part of the questionnaire generates quantitative data, which we analyse using the statistical software SPSS 15.0, and XLSTAT 2007.5. The quantitative data gathered is dichotomous variables at a categorical level of measurement. That is, each listed category is a separate variable, which may assume the value 1, if ticked off, and 0 if not ticked off. For instance, the 17 listed types of television broadcasts are considered to be 17 individual variables either ticked off (1) or not (0). In order to explore our data we use Multiple Correspondence Analysis (MCA), which is a tool for visualisation of links between the categories of more than two variables. MCA is a method that allows us to study the association between three or more categorical variables. MCA is to categorical variables what Principal Component Analysis is to quantitative variables, and the method can obtain visual maps where it is possible to observe the distances between the categories of the categorical variables and between the observations or respondents (Greenacre & Blasius, 2006). We use MCA to explore the data for relationships between variables. For variables were MCA indicates that relationships might exist, we employ tests of association to determine whether a statistical significant association does exist between the variables. We primarily employ the chi² test to identify the associations since the test “[…] is an inferential statistical test that is used to examine relationships between two variables with nominal or ordinal data” (Vaughan, 2001, p. 75). Thus, the chi² test is appropriate for testing for association between the variables in the collected data. Due to inaccuracies in the use of the chi² test with small expected frequencies (Cramer, 1998, p. 301), we follow Siegel and Castellan’s (1988, pp. 103-111) recommendation to use Fisher’s exact test if 20% or more of the cells have an expected count less than 5. As follows, we use chi² and Fisher’s exact test concurrently to test for statistical association between the data variables. In both cases the significance level α is set to 0.05 and for each statistical calculation we examine
whether the obtained p-value is equal to or less than the pre-set significance level. If so, we reject the null-hypothesis (stating that there is no statistically significant association between the two variables), and conclude that there is a statistically significant association between the two variables being tested. If the p-value is higher than $\alpha$, we fail to reject the null-hypothesis, and we are consequently unable to conclude that a statistically significant association exist (Cramer, 1998, pp. 298-306).

The qualitative data is analysed in two stages. Firstly, they are analysed superficially in order to identify potential interviewees. This analysis focuses mainly on the 47 respondents that have provided contact information in the 7th section of the questionnaire. The second stage of qualitative analysis is done in greater details. For this analysis, we use the ATLAS.ti 5.2 qualitative data analysis software, and we apply the coding scheme developed in our analysis of the nine in-depth interviews (see Section 6.4.3). Our methodical considerations regarding these follow-up interviews are described in the subsequent section, including our approach for analysing and coding the interviews.

6.4 Open-ended interviews

The web questionnaire survey is followed up with nine open-ended interviews. The purpose of these interviews is to gather in-depth knowledge about aspects of the respondents’ behaviour when seeking television broadcasts. In this way, the interviews contribute with additional knowledge in relation to the first research area. The interviewees all responded to the web questionnaire prior to the interview, and the interviews are guided by the content of the filled in questionnaire. The interviews are important because if one wishes to collect data on cognitive perceptions it is important to allow respondents to use their own wording. At the same time, it is important to have an overall structure to ensure that the same ground is covered when collecting data from different respondents. Consequently, an interview guide that assures an overall structure and at the same time enables great variability is constructed. The interview guide is depicted in Appendix 8. Several interview techniques can be applied to guide collection of such in-depth data from respondents. Examples are the critical incident technique (Flanagan, 1954), the micro-moment time-line interview technique (Dervin, 1983b), and the explicitation interview technique (Urquhart et al., 2003). Common to these interview techniques is their focus on specific previous events in order to trigger the respondent to recall and explain his or her cognitive perceptions at the time of the event.
When respondents are asked to recall and explain their cognitive perceptions in relation to a specific incident, a retrospective bias might appear. The retrospective bias appears as the respondent’s limited ability to recall the incident can result in "[...] retrospective ‘aggregate’ responses that reflect faulty reconstruction of the phenomena of interest" (Bolger, Davis & Rafaeli, 2003, p. 585). By minimising the gap between the occurrence of the information need and reporting of the need, the probability of a retrospective bias is limited. Therefore, the in-depth interviews sought to focus on the most recent incident where the interviewee needed television programmes. However, in some cases older incidents are covered in the interviews in order to gather information on more complex information needs (Foster, 2005). This is not least relevant since the instances that involve needs for television broadcasts are so few that some retrospective aggregation is unavoidable. From the literature on humanists’ information seeking behaviour (see Section 3.3), the SUL interviews, and the questionnaire responses we know that the need for television broadcasts to serve as empirical data is far from a daily occurrence. As a consequence, we exercised great care in selecting interviewees. The selection of interviewees is elaborated in Section 6.4.2. Following Lincoln and Guba (1985) the familiarity with the field and the respondents as outlined in Section 6.2.3 increases the credibility of the present work.

The section consists of three sub-sections. Firstly, we outline the design and construction of the interview guide. Secondly, we present the practical data collection, including selection of interviewees, and finally, we outline our approach for analysing and coding the collected interview data. In line with the methodology background of our work, we use inductive content analysis (e.g., Allen & Reser, 1990; Schamber, 2000; Neuendorf, 2002), and open and axial coding (Strauss & Corbin, 1998) in order to analyse the transcribed interviews. The linear presentation of these elements is not an accurate representation of the actual workflow procedure of our work. The elements interact and are conducted concurrently but the division into three sections is a relevant mean in order to present our work in a structured and organised fashion.

6.4.1 The interview guide

In the interviews, we gave attention to the work tasks and work processes involved when using television programmes in order to gather knowledge about the user’s information needs in the context of television programmes.

Construction of the interview guide is based on Dervin’s (e.g., 1983b) micro-moment time-line interview technique, Schamber’s (e.g., 1991b) variation of the technique, and Kvale’s (1996) guide for conducting qualitative research interviews. Schamber emphasises that the micro-moment time-line interview technique is
Chapter 6: Methodical approach

appropriate for collecting data on the user’s cognitive perceptions, where “[…] the researcher is particularly challenged to choose techniques for collecting evidence of these slippery, abstract phenomena in the most appropriate way possible: as expressed by the perceivers themselves” (Schamber, 2000, p. 734). Variations of Dervin’s time-line interview technique has proven applicable for uncovering such cognitive perceptions in several examples (e.g., Schamber, 1991b; Yang, 2005).

The interview guide consists of six parts: 1) introduction; 2) focus on a specific situation or incident; 3) sources for searching and retrieving television broadcasts; 4) idealised searching and retrieval; 5) typicality of described situations; and 6) closure. Each part is outlined in the following. An example of the interview guide is depicted in Appendix 8.

In the introduction, part 1, the interviewee is informed about the purpose of the interview, the estimated duration of the interview, and the agenda for the interview. The two former are also included in our invitation to participate in the interview, which is depicted in Appendix 9, and shortly introduced in the following Section 6.4.2.

The 2\textsuperscript{nd} part of the interview guide focuses on uncovering specific situations or incidents in which television broadcasts are important for the interviewee, including past and present situations. Present situations also covers research projects in the melting pot. The interviewee is asked to give a short outline of resent situations in which the need for television broadcasts has come up. This is in line with the micro-moment time-line interview technique, as described in details by Dervin (1983b) and by Schamber (1991b). Based on this outline, we focus the attention on one specific situation, or a critical incident in Flanagan’s (1954) terminology. For this specific situation, the interviewee is requested to describe the workflow process, including the underlying aim, and the specific needs for television broadcasts. In the interviewer’s selection of the critical incident, recency is an important parameter. The choice of recent situations is taken in order to minimise the retrospective bias discussed in Section 6.4. This said, attention is also given to some of the non-recent situations if the interviewer from the outlined description assessed that including such older situations might bring forth uncovered information. In this way, the interviews follow Foster’s (2005, p. 2) recommendation to focus on the most recent situation, but also to allow respondents to include situations from their whole experience, in order to cover as much of the interviewees’ experiences as possible.

The 3\textsuperscript{rd} part focuses on the explicit sources the interviewee applies for covering his or her need. The interviewee is asked to outline these sources, describe pros and cons about the sources, and to describe how and why the sources are selected. Further, the interviewee is asked about his or her knowledge about the needed broadcasts prior to
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

interacting with the IR system, and why the specific broadcasts sought are considered relevant for the specific information need. The interviewer’s increasing knowledge about the field of Media Studies and especially the sources applied for searching television broadcasts is an important element for probing the interviewee about the sources. The interviewer also probe for information about broadcasts, which the interviewee was seeking but did not manage to retrieve.

The 4th part of the interview is concerned with the interviewee’s perception of the idealised searching and retrieval of broadcasts in the specific situation in question. In this part, we emphasise that the interviewee should try not to think about physical restraints, because we experience that the interviewees had a hard time talking about an intangible and to some extend unimaginable broadcast retrieval system. In this part, the interviewer takes care in discussing both searching and relevance assessment of television broadcasts.

The 5th part concerns the typicality, and importance of the situation in question. The interviewee is asked to evaluate to what extent the situation is representative for other situations, which the interviewee have experienced. If the interviewee mentions other distinctive types of situations, we are able to cover such situations as well. This is for instance relevant in the distinction between situations developed out of a research or teaching context. This part also uncovers how important the interviewee thinks the searching and retrieval of television broadcasts is for handling the situation in question. This inclusion of several situations might be a violation to the critical incident thoughts of the time-line interview technique, but it allows us to collect data on information needs, which might otherwise not have been covered in the interviews.

The 6th and final part of the interview is the closure of the interview. The interviewer thanks the interviewee for his or her participation, outlines the work procedure for the interview (e.g., that a transcript will be produced and that the interviewee will receive the transcript for review and acceptance prior to publication), and asks for a referral to persons whom would be relevant to interview. The latter question generates information which is relevant for selecting interview respondents (see e.g., Quote 4 in the following section). In addition, being able to include a referral in the invitation to participate in an interview session is considered to have a positive effect on the response. This chain referral sampling approach is described in Section 6.4.2. Further, the interviewee is asked whether he or she has any comments to add.

Based on our knowledge about earlier research, an initial analysis of the questionnaire responses, and our increasing domain knowledge, the outlined overall interview guide is adjusted the specific interview respondent. This ongoing development of the data collection tool is in line with our methodological basis in
grounded theory, since the methodology states that data collection tools are developed throughout the data gathering process. That is, Glaser and Strauss’ principle of theoretical sampling states that data collection, coding, and analysing is conducted jointly (Glaser & Strauss, 1967, p. 45). This principle is further discussed in Section 6.4.3 on analysing and coding the collected qualitative data. The adjustments made to the interview guide are few and are either concerned with personalized questions, based on the knowledge we have gained from the questionnaire filled in by the interviewee, and/or our increased knowledge about the field of Media Studies. An example is our increased focus on the use of the TV-Meter system for searching television broadcasts. We increased this focus in the interviews in line with an increased knowledge about the importance of TV-Meter in the Media Studies educations, and the use of this tool for searching television broadcasts.

Our outline of the interview guide makes the nine interviews appear linear, and similar in execution. This is however not the case. The guide gives an overall structure for the interviews and makes sure that each of the six parts is covered, but the execution of each interview is unstructured and is to a large degree conducted on the interviewee’s grounds. This means that we ensured that each of the six parts of the interview guide are covered in each interview, but apart from the introduction, and the closure the parts are covered in different order for each interviewee, and sometimes the interviewer comes back to a part already discussed, e.g., by focusing on another situation outlined by the interviewee in the beginning of the interview. The interview does not follow discrete, formal procedures, instead, much is left for improvisation and intuition on the interviewer’s behalf. Therefore, in order to avoid the interviewer is leafing through papers throughout the interview session the interview guide was kept on a single A4 page. We conduct all interviews as well as the subsequent analyses.

Through several of the interviews, we discuss some of the interviewee’s own examples of registration of television broadcasts. This includes examples from private as well as more collaborative collections of television broadcasts. Where possible, we collect copies of such examples subsequent to the interview.

The rationales for the selection of the interviewees are outlined in the following section.

6.4.2 Collecting the data: respondents

The number of interviews to conduct is not fixed, but based on the notion of saturation (e.g., Strauss & Corbin, 1998; Guest, Bunce & Johnson, 2006). Earlier research (e.g., Park, 1993; Barry, 1994; Schamber, 2000) shows that ten respondents are enough to generate reliable results in explorative studies of users’ cognitive perceptions, and a
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

general yardstick of ten interviews was estimated beforehand. The interviewees are sampled purposively, by chain referral sampling, and by theoretical sampling. The strength of purposive sampling lies in selecting participants who hold rich information on the issues being studied, because "[…] purposive sampling is based on the assumption that one wants to discover, understand, gain insight; therefore one needs to select a sample from which one can learn the most" (Merriam, 1988, p. 48), and purposive sampling is especially relevant for exploratory studies.

A chain referral sampling like Goodman’s (1961) snowball sampling technique, utilise the respondents’ knowledge about the field in order to identify respondents who are relevant to include in an investigation. Respondents are simply asked to indicate other respondents, and the sample is growing like a rolling snowball, hence the name. Snowball sampling inherits several biases (e.g., the fact that people with many acquaintances are more likely to be sampled), but as stated by Erickson (1979) using several incentives can solve such biases. Such an incentive is e.g., the purposive sampling based on our knowledge about the field of Media Studies.

Another incentive is the grounded theoretical concept of theoretical sampling (Strauss & Corbin, 1998). Theoretical sampling refers to sampling of respondents along the way as knowledge about the phenomenon in question is gained. When conducting theoretical sampling, specific sampling techniques are applied, e.g., purposive sampling or as denoted by Strauss and Corbin (1998, pp. 211-212) discriminate sampling.

The criteria for our purposive sampling are constructed in line with gaining knowledge about the field, as well as conducting, and analysing the preceding questionnaires and interviews. An example of knowledge learned from a previous interview is the following quote:

“There are researchers who have a genre specific starting point, and some of us have had a content starting point, and some a target audience starting point, and actually also in terms of content. Then [named researcher] who has been focusing on the form, because entertainment for a large part contains a communicative intention, which is found in many formats”.

Quote 4

Apart from describing different approaches for studying television broadcasts, the interviewee provides a referral to another researcher, and depicts this researcher in relation to the different approaches to the field.

Again, this ongoing learning and adjustment of the data collection goes hand in hand with the grounded theoretical notion of theoretical sampling or “[…] sampling on the basis of concepts and by varying the situations to maximize differences” (Strauss &
Corbin, 1998, p. 82). Due to the explorative nature of the present work, we sought to include interviewees as representatives of the different needs for television broadcasts in the user target group. We focus on six different criteria for our purposive sampling. These criteria are listed in Table 6.2, and concerns: 1) methodical approach; 2) types of programmes; 3) status of work task; 4) academic status; 5) academic department; and 6) the interviewee’s experience in the field.

Table 6.2: Six criteria for purposive sampling

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodical approach</td>
<td>E.g., reception analytical approach or historical approach</td>
</tr>
<tr>
<td>Types of programmes</td>
<td>E.g., documentaries, news or sport</td>
</tr>
<tr>
<td>Status of work task</td>
<td>E.g., ended, ongoing, or commencing research project</td>
</tr>
<tr>
<td>Academic status</td>
<td>E.g., full professor, associate professor, PhD student, or student</td>
</tr>
<tr>
<td>Academic department</td>
<td>Media Studies, Aarhus University or Film and Media Studies, Copenhagen University</td>
</tr>
<tr>
<td>Experience in Media Studies</td>
<td>E.g., educated in Media Studies or newcomer to the field</td>
</tr>
</tbody>
</table>

The knowledge learned about each of the criteria stem from the genuine requisition forms, the SUL interviews, our increasing domain knowledge and knowledge about archiving television broadcasts, the web questionnaire survey, and the preceding interviews.

The interviewees receive an invitation to participate in a follow-up interview. These invitations are comprised differently depending on the situation. We used personal invitations by dropping by a researcher’s office, as well as e-mail invitations. Appendix 9 shows an example of the email invitation we used, and despite difference in the approach for the invitation, all respondents is promised anonymity, and that no quotes are published from the interview without the acceptance from the interviewee. We followed the recommendations by Lincoln and Guba (1985), and interviewed all participants in their normal work context. The interviews are conducted at the participants own office, except for one interview which due to practical circumstances is conducted in the office at our disposal during the two month research visit at Aarhus University. After the interview, each interviewee is thanked with a bottle of wine as a gesture for their participation.

The specific number of respondents to include is based on the point of saturation. That is, the point when no new information seem to emerge from the collected data, or
more correctly that the ‘new’ does not add much more explanation to the phenomenon being investigated (Glaser & Strauss, 1967, pp. 61-62; Strauss & Corbin, 1998, pp. 212-215). We conduct nine interviews at which point we assessed that we had reached the point of theoretical saturation. It is important to stress that the number of interviewees is based on our aim, which is to use interviews to broaden the knowledge gained in the questionnaire survey on aspects of the users’ information seeking behaviour. After the ninth interview, we consider that the behaviour is well covered in terms of its properties and dimensions, demonstrating diversity in the collected data.

The interviews comprise approximately 8 hours of speech, ranging from approximately 35 minutes to 1 hour and 40 minutes, respectively. In one interview, we experience technical problems with the recording device, which means that approximately two minutes are lost from that interview. Immediately after the interview session, we comprised extensive notes for the lost part of the interview, and we do not consider this technical problem a methodical issue.

The interviews are conducted in a period of two months commencing on the 21st of March 2006, which is in the third week of the research visit at the Department of Information and Media Studies, Aarhus University (see Section 6.6 for an overview of the methodical process). Seven interviewees are affiliated Aarhus University, while two are affiliated Copenhagen University. The overweight of interviewees from Aarhus is due to our research visit, their knowledge about our research, and their motivation to participate. The interviewees represent a wide variety of methodical approaches, and types of broadcasts investigated. One interviewee is a Master’s student, two are PhD students, two are assistant professors, two are associate professors, and two are full professors. The experience in Media Studies (including student years) range from four to nineteen years, and the interviewees represent academics educated in Media Studies as well as academics who are ‘newcomers’ to the field. The educational background of the interviewees is apart from Media Studies, mainly Comparative Literature, which is in line with the history of the field (see Section 5.3.3). The critical incidences discussed with the interviewees comprise newly commenced work tasks, ongoing work tasks, as well as ended work tasks, and it includes research, teaching and studying tasks. In order to maintain the promised anonymity of the interviewees, we do not provide further identification of each interviewee, but we provide the contextual information if considered relevant, e.g., when reproducing quotes from the interviews.

An ethical and methodical problem is experienced in the course of the interviews, namely the restraints of the legislation. With reference to Quote 2 in Section 6.2.2 it is obvious that the legislation puts several restraints on the use of television broadcasts, which are not present in a traditional textual environment. The legislative restraints are
evident impediments in relation to the job description of media scholars. For instance, the exemplifying use of television broadcasts is apparent in a course about the historical development of a specific television genre, as illustrated in the following quote:

“It is evident that if you are to teach the media, you need to let the students see and hear the broadcasts, if not, the subject taught will be strange and abstract for the students. The examples in ones teaching is largely based on the copies of broadcasts held in ones network, which is an amateurish base for research-based education”.

Quote 5

An illustrative counterpart from a textual context is an academic course on English literature with specific focus on Shakespeare’s authorship, but without the inclusion of any examples of Shakespeare’s writings in the course. Whether the legislation is followed rigorously in such cases is outside the scope of this work. However, if the interviewees have the legislation in mind and therefore, e.g., avoid the discussion of some sources for obtaining their information, it affects our data collection. Generally, the interviewee trust that agreed upon parts of the interview is published, only, and we consequently believe that this ethical aspect does not hamper the interview sessions. Nonetheless, in at least one interview the interviewee explicitly changed her wordings due to the legislative bindings.

The interviews are transcribed, coded, and otherwise analysed in order to derive knowledge about aspects of the interviewees’ information seeking behaviour in a television broadcast context. Our considerations regarding the analytical methods are outlined and discussed in the next section.

6.4.3 Transcribing, coding and analysing the collected data

We transcribed the four SUL interviews and our nine follow-up interviews. The transcriptions are verbatim word-to-word transcriptions, with no indication of emphasis in intonations or emotional expressions (Kvale, 1996, p. 170). All transcriptions are done by us, and by use of Transana 2.05 for transcription of audio and video. One of the benefits of the Transana software is its use of time-codes. By using time-codes in the transcript, the textual transcription is related to a specific point in the audio file, and it is consequently easier to include audio files in subsequent analyses. Transcripts are artificial constructions from oral to written communication. Transcriptions are not textual copies of the oral communication, but a translation from one ‘language’ to another, and naturally they are a useful tool for the given purpose, namely analysis of
the interview and mediation of the interviewee’s perception (Kvale, 1996, p. 166). Time-codes allows one to quickly identify the one minute of speech, in question, in a long interview. In this way, the interview and the transcription are used in a complementary way in our analysis. The transcription is the first part of the analysis of the interview and it is commenced within a few days after each interview. The knowledge gained from this initial analysis is then applied in the subsequent data collection, in the way we have described above.

After transcription, the transcripts are handed to the interviewees for review, and the interviewees are urged to respond with clarifying comments, corrections, as well as whether or not quotes from the transcript can be published. Along with other member checking facilities (e.g., group discussion of preliminary results), it increases the credibility of the present investigation. The transcripts comprise in total 162 single lined Word pages, and range from 8 to 23 pages for each interview.

The verbatim oral language reproduced in our word-to-word transcriptions, including *hmm*’s and the like, is not applicable for reproduction in the writing style of the present work. Consequently, apart from translating quotes from Danish to English, each quote is edited into a written mode of discourse. However, as emphasised in the hermeneutic tradition of translators: *traduire traittoria* (translators are traitors) (Kvale, 1996, p. 166), and therefore Appendix 10 contains the Danish verbatim counterpart for each quote. In this way, the reader knowledgeable in the Danish language can verify the validity of our translations and editorial amendments.

Upon transcription of the final interview, the Transana Text files are spell checked in Word, and converted into RTF (Rich Text Format), in order to be imported into the ATLAS.ti 5.2 software. The ATLAS.ti software is based on the grounded theory approach, and it has earlier been used to analyse interviews in relation to user behaviour in LIS (e.g., Wilson, 2004; Foster, 2005; Hyldegård, 2006). The software is very useful for ad hoc meaning generating analysis in Kvale’s (1996, pp. 203-204) terminology. Inductive content analysis is such an ad hoc meaning generating type of analysis, which suits well the grounded theoretical approach because it ”[…] derives theory from data rather than verifies existing theory” (Schamber, 2000, p. 735). Inductive content analysis has been applied by several researchers for data analysis (e.g., Allen & Reser, 1990; Schamber, 2000; Neuendorf, 2002). We develop categories or codes inductively from the transcripts being analysed. An important element of the grounded theoretical methodology is that “[…] analysis is not a structured, static, or rigid process. Rather, it is a free-flowing and creative one in which analysts move quickly back and forth between types of coding, using analytical techniques and procedures freely and in response to the analytic task before analysts” (Strauss & Corbin, 1998, p. 58). Our
coding is conducted in several iterations. In-depth interviews are appropriate for collecting data on (potential) users’ cognitive perceptions, but the approach is also very labour intensive for the researcher. This is primarily due to the extensive qualitative analysis needed in order to identify categories in the collected data.

The coding process is commenced with a microanalysis of the transcripts from our nine interviews (Strauss & Corbin, 1998, pp. 57-71). All transcripts are read thoroughly line-by-line, in order to be able to “[…] uncover new concepts and novel relationships and to systematically develop categories in terms of their properties and dimensions” (Strauss & Corbin, 1998, p. 71). The data is mined so to speak. This first reading involves open coding in which concepts are identified and their properties and dimensions are discovered in the data. In this microanalysis of the transcripts, the code scheme is slowly being developed, and due to its developing nature several codes are applied for the same phenomenon, however through the iterative analysis process such double codes are merged into distinct codes. At the same time, we split other codes into several (new) codes.

After coding all transcripts once, a second iteration is commenced. This second iteration includes open coding as well as axial coding. Axial coding denotes the process of relating codes or categories to sub-categories, as well as linking and relating categories. The constructed codes are compared, new codes developed, and relationships established. Similar iterations are conducted until we reach a point where we did not add any new codes or relationships, the point of saturation (Strauss & Corbin, 1998, p. 136). In this way we followed the recommendations by Glaser and Strauss to conduct a constant comparative analysis as a mean to identify categories in the data (Glaser & Strauss, 1967, pp. 101-115). This approach for coding allows the researcher to be systematic and creative at the same time, and to derive what anthropologists denote thick descriptions from the data (Geertz, 1973).

During our iterative analysis of the nine interviews, we construct our coding scheme. This scheme is later applied as the basis in our analysis of the four SUL interviews and the qualitative parts of the web questionnaires. That is, the coding scheme is not further developed but merely applied in this final analysis of the SUL interviews and the questionnaires.

In line with the characteristics of naturalistic research (e.g., Lincoln & Guba, 1985; Mellon, 1990; Foster, 2005), the investigation does not claim generalisability. It discover the common characteristics of the test participants’ information needs, search entries preferences, and applied relevance criteria, in a descriptive manner. Instead of generalisation, the present work ensures rich description to make sure that the research theme is transferable and open for further developments.
One of our ways for ensuring validity of our results is to involve representatives from the population in our analysis and discussion of preliminary results. This involvement is done through a group discussion session, which is the focal point of the upcoming section.

6.5 Group discussion with scholars

Five weeks into the two months research visit we arranged a one-hour group discussion session with scholars in Media Studies about the preliminary findings of our investigation. The session lasted one hour and comprised 11 scholars. Five interviews had been conducted at the time of the group discussion session, four of the interviewees were present at the session, and a fifth participant was interviewed the following day. Due to practical circumstances, we were unable to record the session, and instead we conducted heavy note taking during and immediately after the session.

The session started with a 25 minutes presentation of the project and the preliminary results from our side (Power Point slides from the presentation are reproduced in Appendix 11). In the remaining 35 minutes of the session the participants discussed, and commented on the presentation. The comments and discussions were both directed towards the presenting researcher, and internally among the participating scholars.

The aim of the group discussion is to broaden our analysis. Due to the fact that the group session participants are all scholars, they possess the knowledge to analyse and discuss our results from an academic perspective. In addition, they have first hand knowledge about the field in question. Another aspect of the group discussion is that it allows for member checking of our results. Such member checking is important to increase the credibility of the present work (Foster, 2005, p. 3). Neuendorf (2002, p. 103) recommends the inclusion of a professional committee as a mean to generate the most reliable coding scheme for the content analysis.

To provide a closure of the present chapter, the following section presents a brief overview of the methodical process.

6.6 Methodical overview: timeline for data collection

Our investigation does not follow one methodical approach, but we employ a triangulation of methodical approaches. Further, the methods are not carried out in a stringent waterfall-like fashion. Instead, we apply a fluid and skilful application of the methods, which is recommended by Straus and Corbin (1998, p. 46), and hence it
Chapter 6: Methodical approach

reflects our methodological foundation in grounded theory (Glaser & Strauss, 1967; Strauss & Corbin, 1998).

This section provides the reader with a visual, illustrative overview of the methodical process in terms of time of execution of the different parts. Table 6.3 depicts the overall timeline including the empirical data originating from the State and University Library (genuine requisition forms and the four SUL interviews). During spring 2006, the empirical data collection is at its highest, and this is depicted in detail in Table 6.4.

Table 6.3: General overview of the data collection process.

<table>
<thead>
<tr>
<th>Year</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Season</td>
<td>Spring</td>
<td>Fall</td>
<td>Spring</td>
<td>Fall</td>
</tr>
<tr>
<td>Genuine requisition forms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge about Media Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four SUL interviews</td>
<td></td>
<td></td>
<td>Received 6/10</td>
<td></td>
</tr>
<tr>
<td>Research visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web questionnaire Survey</td>
<td></td>
<td></td>
<td>See Table 6.4.</td>
<td></td>
</tr>
<tr>
<td>Nine interviews</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group discussion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.4 depicts the overlap between the different data collection techniques applied in the spring 2006. The table depicts the 12 weeks in which most of our data collection took place, and it also depicts the exact date for conducting the interviews and the group discussion. Two interviews are conducted on May 2nd and the table consequently holds eight interview dates, only.

Table 6.3 and Table 6.4 show how the different parts of our data collection process are intermingled. This intermingling is partly caused by practical considerations and restraints, and partly founded in theoretical considerations. All in all, the data collection process, as well as the subsequent data analysis is in line with the overall grounded theory methodology.
Table 6.4: Overview of the most active part of the data collection process, which occurred in the spring 2006.

<table>
<thead>
<tr>
<th>Month</th>
<th>March</th>
<th>April</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Day</td>
<td>1.</td>
<td>13.</td>
<td>21.</td>
</tr>
<tr>
<td></td>
<td>24.</td>
<td>27.</td>
<td>30.</td>
</tr>
<tr>
<td></td>
<td>5.</td>
<td>6.</td>
<td>30.</td>
</tr>
<tr>
<td></td>
<td>2.</td>
<td>15.</td>
<td></td>
</tr>
</tbody>
</table>

Research visit

Web questionnaire

Survey

Nine interviews

Group discussion

This ends our methodical chapter. In the following Chapters 7, 0 and 9, we present the results of our empirical investigation. Each chapter reports on one of the three research areas. Chapter 7 concentrates on the first research area on aspects of users’ information seeking behaviour in relation to television broadcasts, and the chapter answers research questions 1, 2, and 3. Chapter 0 draws on the results obtained in Chapter 7 about aspects of seeking behaviour, and focuses on the appropriate access points for indexing and retrieval of television broadcasts, which is the second research area. In this way, the chapter answer research question 4. Finally, the focal point of Chapter 9 is research area 3 on construction of surrogate records for television broadcasts, and the chapter answers research question 5 on the value and applicability of external and other sources for indexing of television broadcasts. In addition, Chapter 9 summarises the empirical results prior to presentation of our final conclusions in Chapter 10.
7 Research area 1: aspects of information seeking behaviour

This chapter reports on our investigation of aspects of users’ information seeking behaviour in relation to television broadcasts, which is the first research area. Firstly, the chapter concentrates on our results concerning scholars’ and students’ work tasks and information needs in a television broadcast context. Secondly, we focus on the search entries preferred by the test participants, and thirdly, we devote our attention to the relevance criteria, which are applied when seeking television broadcasts. In this way, this chapter comprises answers to the first three research questions, which state:

1. What characterises a given group of users’ information needs in the context of television broadcasts?
2. Which search entries are preferred for searching of television broadcasts, and why and when are they preferred?
3. Which relevance criteria are applied when evaluating television broadcasts, and why and when are they applied?

The chapter concludes with a discussion and summarisation of our results in relation to research area 1.

The qualitative and quantitative data have contributed concurrently to our analyses, and in the presentation of our results, the different types of data consequently intermingle. This said, the different methods for data collection and analysis have had different purposes in mind. Accordingly, the data informs on different aspects of the three research questions, and the different types of data are consequently analysed to a different extent for answering each research question. In the following, we use the term respondent to denote a questionnaire participant, interviewee to denote an interview participant, and test participants to denote questionnaire as well as interview participants. Further, the terms academic and user are applied to denote a generalisation beyond the group of participants investigated in the present work.

7.1 Research question 1: what characterises a given group of users’ information needs in the context of television broadcasts?

This section comprises four sub-sections focusing on different characteristics of users’ information needs, and in total, these sections answer research question 1. Information
needs are somewhat intangible entities as they are internal cognitive phenomenon, e.g., according to De Mey’s stages of processing (De Mey, 1980) and Taylor’s information need stages (Taylor, 1968). The intangibility is one of the main reasons for our choice of the in-depth interviews as a data collecting technique, due to the interviewer’s opportunity to prompt the interviewee for additional information on specific aspects. Consequently, the interviews are the main contribution for gathering knowledge about the characteristics of academic’s information needs in the context of television broadcasts.

The intangibility of information needs is dealt with by focusing on different elements or aspects of the needs. The first Section 7.1.1 focuses on the work tasks underlying the test participants’ needs for television broadcasts. In Section 7.1.2 we concentrate on the nine interviewees’ information needs. Section 7.1.3 reports on the results obtained from the questionnaires, and finally in Section 7.1.4 characteristics of the test participants’ information needs in a television broadcast context are summarised and discussed in relation to previous research on information seeking introduced in Chapter 3.

7.1.1 The work tasks underlying the nine interviewees’ television needs

In the interviews, we ask the interviewees to list recent situations that have induced a need for television broadcasts. Each of the nine interviewees mentions one or more situations that have prompted a need for television broadcasts. From this listing, we select one specific situation or critical incident to be the primary focus for the remainder of that interview. In the following, we briefly describe the primary work task or critical incident in the nine interviews. The order of our descriptions reflects the order by which the interviews were conducted, and we attach the labels (A-I) to denote the interviews and the interviewees in the remaining of the present work. In the subsequent section, the work tasks are further elaborated as we identify the needs for television broadcasts for each of the nine interviewees. However, in the present section, we focus on the aims and motivations underlying each of the nine interviewee’s work task. The purpose of this presentation of work tasks is to show the differences in foci and perspectives found in the present data, which form the basis for outlining the associated information needs. In addition, outlining the work tasks and the associated needs enables the reader to become familiar with the pre-requisites for the characteristics of the test participants’ information needs, and hence better understand the conclusions drawn in the present work.

Interviewee A is a PhD student, and the work task is a PhD project concerned with the analysis of programme format versions. In brief, a format version is an
internationally licensed programme which is produced and broadcasted in national versions, e.g., ‘Pop-stars’ or ‘Jeopardy’. In LIS terms this is comparable to different expressions of the same work (see IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998). At the time of our interview, interviewee A is half through the three-year PhD project, and the interviewee has collected the television broadcasts that are to be analysed. The motivation for the interviewee is to gain better insight about, and to contribute to, the research on the role of format versions in an increasingly globalized media picture. The work task includes a comparative text and genre analysis of Danish and Australian format versions, and a historical quantitative analysis of the share of format version programmes in Danish television schedules.

The work task in the second interview is also a PhD project, and interviewee B is approximately at the same stage of the project as interviewee A. The motivation for interviewee B is to learn about the media usage of children in the age of 1½ to 3 years. Interviewee B focuses on the children’s usage of moving images as well as the associated merchandises (lunch boxes, bed linen, etc.). In this way, the interviewee’s aim is to add to our knowledge about children as media consumers.

Interviewee C’s primary work task is a Master’s dissertation, and in the dissertational work the interviewee conducts a comparative analyses of two private corporations in order to learn about the relationship between modern management and corporate communication. The two corporations in question are selected based on their diversity in management and communication. In the research project, the interviewee focuses on corporate communication via television. The work task is a six-month research project, and at the time of the interview, interviewee B is in the midst of identification and collection of the television broadcasts which are to form the basis of the comparative analysis.

In the fourth interview, interviewee D is about to begin a new research project on a definition of the genre of traveller guides. Interviewee D is an assistant professor, the research project had just been approved and, at the time of the interview, the primary work on the project is scheduled to start in one year’s time. Nevertheless, the interviewee is already considering how television broadcasts are identified and collected for the project. The research project is concerned with the development of a definition of the genre of traveller guide programmes in the competitive phase of Danish television (from 1988 and forth). A sub-task in the research project is to conduct a comparative so-called aesthetical-analytical study of Danish and foreign produced traveller guide programmes. The genre of traveller guides is not entrenched within the research community, and the aim of the interviewee is to contribute with an understanding and ultimately establishment of the genre in a scientific perspective.
The main work task being discussed with interviewee E is a completed research project concerned with entertainment programmes in a historical perspective. The interviewee’s motivation is to remedy that the development of entertainment programmes is an unexplored research area nationally as well as internationally. Unlike, interviewee D’s work task, the work task is not as much to define the genre. Rather, the work task is concerned with description of the historical development of the genre, including the space the genre takes up in the television schedules.

Interviewee F’s work task is similarly concerned with a historical perspective, but interviewee F is not focused on television as the sole media. Instead, the interviewee’s aim is to describe the historical development of Danish sports journalism in general. The interviewee includes the printed press, radio, and television, but with considerable focus on mediation via television. The development of Danish sports journalism is an unexplored topic and the interviewee’s motivation is to contribute to our knowledge in this area of research. The interviewee is an associate professor and the work task is a finished research project.

In the seventh interview, the interviewee is a full professor, and the research project in question is the description of Danish television history. The aim behind the research is to produce a coherent description of Danish television history in print. Interviewee G’s objective is to contribute to this work with a description of the historical development of Danish news broadcasts, with primary focus on the development in content and form of the news mediation. Interviewee G’s motivation is to broaden the knowledge about the development of the mediation of news in Danish television. The research project is completed, and the results have recently been published.

The primary work task discussed in interview H is concerned with research in the theory and history of Danish television documentaries. The motivation underlying the project is an ambition to contribute to our knowledge about documentaries as a genre, as well as its inherent sub-genres. The project is in its last stage and interviewee H is finalising a book in which the historical development of the genre and the sub-genres is described and analysed.

In the ninth and final interview, interview I, the work task in question is concerned with broadening the scientific understanding of Danish soap operas as a genre. The interviewee analyses five selected soap operas from a genre and reception analytical perspective. The interviewee is an assistant professor, and the research project being discussed is the interviewee’s recently completed PhD project. The motivation underlying the work task is to gather knowledge about the role of the genre in a Danish and Swedish media culture.
In brief, the interviewees give statements about prior, ongoing, and new research projects or work tasks. For one of the assistant professors, the associate professors and the full professors the main work tasks are prior multi-annual research projects. The main work tasks in the interviews with the Master’s student and the two PhD students are ongoing research in relation to the Master’s dissertation and the PhD theses. For one of the PhD students and one of the associate professors, television broadcasts are but one of the media objects being analysed in the main work task in question. While in the remaining seven interviews, television broadcasts are the sole media objects being analysed. The main work tasks in the interviews are recapitulated in Table 7.1.

Table 7.1: Listing of the work tasks of the nine interviews.

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Main work task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewee A</td>
<td>Research project on analysis of television programme format versions.</td>
</tr>
<tr>
<td>Interviewee B</td>
<td>Research project on analysis of children’s reception and use of moving images.</td>
</tr>
<tr>
<td>Interviewee C</td>
<td>Research project on analysis of the management style in two private corporations.</td>
</tr>
<tr>
<td>Interviewee D</td>
<td>Research project on definition and description of traveller guides as a television genre.</td>
</tr>
<tr>
<td>Interviewee E</td>
<td>Research project on description of the historical development of Danish entertainment programmes.</td>
</tr>
<tr>
<td>Interviewee F</td>
<td>Research project on description of the historical development of Danish sports journalism.</td>
</tr>
<tr>
<td>Interviewee G</td>
<td>Research project on description of the historical development of Danish news broadcasts.</td>
</tr>
<tr>
<td>Interviewee H</td>
<td>Research project on description of the historical development of Danish documentaries.</td>
</tr>
<tr>
<td>Interviewee I</td>
<td>Research project on genre and reception analysis of Danish soap operas.</td>
</tr>
</tbody>
</table>

In line with Foster’s recommendations, the interviews also cover other work tasks in order to gather knowledge from a broader range of experiences (Foster, 2005, p. 2). These secondary work tasks are concerned with teaching, including supervision, and the students’ experiences with needs for broadcasts during their studies. Teaching experiences are collected from all interviewees but the Master’s student, and the student experiences are collected from the Master’s student as well as the two PhD students.

Altogether, the main and secondary work tasks covered in the interviews correspond to four of the five types of scholarly moving image work tasks listed by Auffret and Prié (1999, pp. 322-323), and outlined in Section 3.4.3: pedagogical tools, testimony of the past, work of art, and communication acts. The main focus is given to
work task of the second and the latter type. The task concerned with the scholars’ own creation of moving images is irrelevant in the context of a television broadcast retrieval system, and it is consequently not represented in the nine interviews.

7.1.2 Information needs expressed in the nine interviews

Here follows a presentation of the information needs of each of the work tasks (A-I), presented one by one in nine separate sub-sections. Characteristics of the interviewees’ needs are summarised and discussed in a tenth and final sub-section. The aim of this explicit description is to point to the characteristics of the information needs of each interviewee and hereby answer the first research question, as well as to serve as a point of reference for discussions in the remainder of the present work.

7.1.2.1 Interviewee A - Television programme format versions

Interviewee A has a good understanding of the television broadcasts, which are relevant for fulfilling the work task on television programme formats. As stated by the interviewee:

“[…] I know the programmes I need to search for. In my searching the title of the programme is the main search entry”.

Quote 6

The needed broadcasts are at the granularity level of programmes, and the perception of the broadcasts needed is generated through the interviewee’s scientific work in the specific area of interest.

“[…] during my work I have zoomed in on the programmes I wish to analyse, and via my scientific research I probably know all current format versions broadcasted on Danish television”.

Quote 7

A prerequisite for interviewee A’s clear perception of the needed broadcasts is an overview of the primary sources or the format programmes in Danish television. Initially, the interviewee does not consider the generation of this overview part of the need for television broadcasts, but as an isolated part of the research process. The need to generate an overview of all format programmes is not concerned with gaining access to television broadcasts, but instead it is concerned with gathering information about the television broadcasts. That is, the interviewee does not necessarily need access to the
Chapter 7: Research area 1: aspects of information seeking behaviour

broadcasts, if information about format versions is available from other sources. Interviewee A states that:

“[…] it would be my dream: to be able to search for programmes which are Danish versions of foreign programme formats”.

*Quote 8*

However, generating the overview is a prerequisite for identifying the specific programmes the interviewee wishes to analyse, and hence obtain a copy of.

In brief, the interviewee’s information need consists of two parts:

1) A need to generate an overview of format versions in Danish television; and
2) Selection of specific format version programmes for analysis.

In the selection of specific programmes for analysis, interviewee A puts emphasis on the domain knowledge, and especially the specific knowledge gained about the primary sources, e.g., as expressed in Quote 7. More specifically, the interviewee emphasises that the popularity and the recency of the programmes as well as the transmission pattern (first time transmission or repeated transmission), are important selection criteria.

The first part of interviewee A’s information need is of an explorative nature, where the interviewee has a good perception of the features of the broadcasts (format versions) sought, and hence it is comparable to Ingwersen and Järvelin’s third type of information need. The second part is verificative or for known items and it is comparable to Ingwersen and Järvelin’s first type of information need (Ingwersen & Järvelin, 2005, pp. 292-293).

7.1.2.2 Interviewee B - Children’s reception and use of moving images

Interviewee B’s focus is children’s usage and/or reception of television broadcasts:

“I do not start with the programmes. I start with the use, and then I look at the programmes for explanation of the use, with the programmes as the basis”.

*Quote 9*

This starting point is markedly different from a point of departure in the programmes as such, which is the case in interview A. During the observation and analysis of children’s reception of television broadcasts, the interviewee’s needs for broadcasts emerges. In this way, the observations and analyses of the observation data gives the
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

interviewee an unambiguous understanding of the broadcasts that are needed for fulfilling the work task.

The broadcasts needed by the interviewee do not necessarily have to be the specific broadcast that was transmitted during the observation:

“If I have a boy who specifically is interested in machines, e.g., the way they drive, then I could obtain a programme with that specific emphasis. Because, I do not need a direct causality between the specific broadcast the child viewed and the broadcast I analyse”.

Quote 10

In such cases, the interviewee has a clear understanding of the relevant programme universe or serials, e.g., ‘Bob the Builder’ or ‘Postman Pat’, but needs to identify a specific programme in a serial. The identification of a relevant programme, which focuses on a specific event, is done by reading descriptions or summaries of the content of programmes.

Similar to the first interviewee, interviewee B is very knowledgeable about the primary sources in the field. In answering how relevant broadcasts are identified the interviewee states that:

“[…] it is something I know. I do not examine which children’s programmes are broadcasted, because I just know that”.

Quote 11

The interviewee draws on this knowledge in order to analyse the programme universes the children experience. The interviewee monitors the transmission of children’s broadcasts, and is kept up to date with the primary sources in this way. This is comparable to the monitoring feature in Ellis’ information seeking model (Ellis, 1989a; Ellis, Cox & Hall, 1993). However, the interviewee does not conduct this monitoring behaviour explicitly, but through the interviewee’s general interest in children’s broadcasts, e.g., by monitoring and watching broadcasts with the interviewee’s own child.

Interviewee B is not only interested in specific programmes or serials, but also in the broadcasts surrounding the programmes, including the trailers.

“[…] I would like to know how the broadcasts are incorporated in the flow, what surrounds them. I am not just interested in what is between the intro and the outro of a programme. I am also very interested in the surrounding broadcasts, because
often the children hear and recognise the intro, watch for a few minutes, walk to and from during the programme, and returns towards the end for the outro, for the theme tune. Then they watch the trailers, and these trailers are important due to my holistic approach to the children’s behaviour”.

The relevancy of the broadcasts in the interim of programmes is deemed based on what they convey (e.g., what they are trailers of), as well as their placement in the television schedule of the channel in question. We denote this horizontal programming, as opposed to vertical programming, which concerns television scheduling across television channels.

The broadcasts in the interim of programmes are considered to be at the feature level of granularity, and interviewee B consequently needs broadcasts at the serial, programme, and the features level of granularity. The main part of the interviewee’s information need is for known items (Ingwersen and Järvelin’s first type), and only a minor part is known topic or content (e.g., events in programmes), which is the third of Ingwersen and Järvelin’s types of information needs. Most of the interviewee’s observations of the children are done with broadcasts transmitted in reel time. Hence, interviewee B does not need access to these broadcasts, and only occasionally does the interviewee need to obtain a broadcast in order to conduct further analysis of that broadcast.

7.1.2.3 Interviewee C - The management style in two private companies

Interviewee C is concerned with a comparative analysis of the style of management and communication in two specific corporations. The interviewee needs all broadcasts in which either the named managers participate or the named companies are mentioned. In answering how the interviewee would like to search for the broadcasts needed interviewee C states:

“I would like to search for [named person], because he is the person I need information about, and perhaps also [named company], since that is the company’s name”.

The interviewee needs anything from news features to reality shows in which the person is involved or the company is mentioned. The starting point for selection of relevant broadcasts is a retrieved list of all broadcasts in which the person or company appears. From the overview of the broadcasts in which the person or company appears, selection
is mainly based on duration, and genre of the listed broadcasts. The relevant broadcasts are mainly news features, and hence at the granularity level of features, and to a minor extend programmes (e.g., the documentary ‘Himlen over Danmark’).

The interviewee reflects on the conception of the needed broadcasts, and states that:

“If I were to use a broadcast retrieval system, I would never just enter search terms. That is, the search terms do not just appear out of the blue, and I would always know what to retrieve, e.g., by known keywords. I would always have a perception of the items I seek”.

Quote 14

The interviewee did not express whether the named person and/or company instigated the information need, or if it was the need for investigation of management style. Further, we did not prompt persistent enough to derive this information. Interviewee C expresses that information needs of a muddled character are considered to be absent in the interviewee’s retrieval of television broadcasts, and hence would be non-existent in the interviewee’s interaction with a future IR system. This is in line with the interviewee’s type of information need. The interviewee is familiar with known keys or features of the broadcast, which are participation of a named person or mentioning of a named corporation. Hence, interviewee C’s information need is characterised as Ingwersen and Järvelin’s third need, the known topic or contents need.

7.1.2.4 Interviewee D - The genre of traveller guides
In Kuhlthau’s terminology interviewee D is in the pre-focus stage of the research project, the topic is selected and the interviewee is conducting pre-focus exploration, including exploration of the primary data sources (e.g., Kuhlthau, 2004). The interviewee is in the stage of generating an overview of the programmes that are relevant for fulfilling the work task. This is similar to the need for generation of an overview of format versions of interviewee A, as interviewee D aims at getting familiar with all traveller guide programmes broadcasted in Danish television. However, due to the fact that interviewee D’s motivation is identification and definition of the genre in question, the interviewee does not hold a clear perception of the keys or features of the relevant broadcasts. It is very difficult to delineate relevant programmes a priori, since the delineation is one of the aims of the work task. The characteristics of the broadcasts needed by interviewee D is the very phenomenon being investigated in the research project. Therefore, during the generation of the overview the interviewee must come to
some decisions about the programmes considered part of the genre. Interviewee D expresses this difficult situation in the following way:

“It is very difficult to say which programmes are traveller guide programmes and which programmes are not. For instance, is a reality show in which two families exchange holidays a traveller guide? There are many areas that must be demarcated, in order to be able to state whether it is a reality show, a traveller guide programme, or both”.

Quote 15

At this initial stage of the research project interviewee D is very interested in the borderline examples that challenge the definition of the genre. The information need is explorative in nature, and the specific programmes of relevance are conveyed during the interviewee’s research, including the generation of an overview of the primary data sources. The interviewee needs to gather increasing knowledge about the characteristics of the primary data sources in order to fulfil the work task. This increased knowledge is also a prerequisite for identifying the features or characteristics of the broadcasts that are relevant to obtain and analyse. The research process coincides with the information need at this stage of the interviewee’s work task. That is, gaining an overview of the primary sources might very well reveal and instigate new areas or sub-tasks to be investigated.

In brief, the information need can be summarised into three parts:

1. Overview of traveller guide programmes broadcasted in Danish television;
2. Identification of borderline examples of traveller guide programmes; and
3. Selection of exemplars of Danish and foreign traveller guide programmes.

In generating the overview, the interviewee seeks to gather knowledge about the broadcasts, including screening issues (e.g., the number of repeated transmissions, date and time for broadcasting, etc) in a horizontal and vertical programming perspective. A screening analysis is an element in the sub-task of comparative investigation of Danish and foreign produced traveller guides. In this sub-task the interviewee is also interested in information about the producers of the broadcasts, both the producing country, as well as the producing corporation.

The interviewee seeks broadcasts at the granularity level of serials and programmes. Interviewee D’s information need is explorative in nature, but in contrast to interviewee A, the interviewee does not have a clear perception of the keys or features of the programmes sought. The information need is comparable to Ingwersen
and Järvelin’s seventh type of need concerned with seeking based on a muddled perception of topic and/or content. In connection with the generation of the overview, and identification of important broadcast characteristic, the muddled need will most likely develop into a verificative need. However, at the present stage of the research project the interviewee’s information need is muddled, only.

7.1.2.5 Interviewee E - The historical development of Danish entertainment programmes

Interviewee E’s work task is the development of entertainment programmes in Danish television. The interviewee needs to gain an overview of the programmes in question, and in this way, it is comparable to the explorative overview generating needs of interviewee A and interviewee D. In line with interviewee A and D, the generation of the overview is to a large extent concerned with the gathering of information about the television programmes in the specific genre in question. Unlike interviewee D’s work task (traveller guide programmes), the genre of entertainment programmes is well established in the research literature as well as at television stations. Interviewee E uses the television schedules as the main source for generating the overview of the genre.

The information need resample interviewee A’s need.

In addition, interviewee E expresses a need to check up on facts about the broadcasts prior to publication of the results, e.g., the time period of transmission of a specific series of broadcasts. This need is comparable to verifying in Ellis’ terminology, and we hence consider the need to consist of three parts, namely:

1. The generation of an overview of entertainment programmes in Danish television;
2. Selection of specific entertainment programmes for analysis; and
3. Verification of facts to support statements in publications.

The first part of the information need is concerned with gathering knowledge about the programmes, and the interviewee does not need to gain access to broadcasts in this phase, e.g., as expressed by the interviewee:

“[…] for much of what we have discussed so far, I do not need to see the programmes in question. Sometimes I need to see the programmes in order to conduct a detailed analysis or to gain a sense of the content of the programme.”

Quote 16
The selection of the specific programmes for analysis is mainly based on the interviewee’s knowledge about, or ‘feeling’ of, the primary sources. That is, the interviewee’s skills in identifying programmes of unknown content, or to identify specific types of programmes that seem to fill the television schedules at certain points in time. In answering how the programmes are selected interviewee E states:

“[...] I have a feeling that I need access to this type of broadcast. For instance, there are exceptional many soloist programmes in the 1960s, e.g., Raquel Rastenni singing for an hour. I am thinking ‘what are these programmes? I need to see a programme to understand what they are’, and then I select an example of such programmes which has been stored”.

Quote 17

In this quote the interviewee touches upon another important element for selection of broadcasts, namely, whether broadcasts have been stored or they have literally vanished into thin air. This is especially relevant for broadcasts transmitted prior to the establishment of the Danish national collection of television broadcasts in 1987. The interviewee continues to reflect on the selection of programmes and says that:

“The selection of programmes is necessarily pragmatic. If you are to cover approximately 50 years of history, you need to say: ‘it seems there are a lot of similar programmes. I do not know what they are. I need to see an example. This example is available, I will take that’. Initially, I had envisioned that I could do an ideal selection, by selecting programmes from every fifth or tenth year, and thereby retrieve the whole range of programmes, but that is not possible since the programmes do not exist”.

Quote 18

The sources, which the interviewee bases the generation of the overview on, are mainly the television schedules from the television stations. The date or year of transmission is hence important for the interviewee.

The interviewee needs television broadcasts at the granularity level of serials and programmes. Similar to interviewee D, the first part of the information need is concerned with the explorative seeking of programmes from ill-defined keys or features. This is comparable to the muddled topic or content need, which is the seventh need in Ingwersen and Järvelin’s typology. Upon generation of the overview, specific broadcasts are selected. This selection is comparable to the first need in Ingwersen and Järvelin’s typology, namely the need for known items. Lastly, the verificative fact
searching need is comparable to Ingwersen and Järvelin’s fourth information needs on factual data searching.

7.1.2.6 Interviewee F - The historical development of Danish sports journalism

Interviewee F’s information need is somewhat similar to interviewee E’s need, in the way that interviewee F seeks a historical overview of broadcasts within a specific genre. The work task is concerned with generating an overview of sports broadcasts in Danish television from the beginning in 1951 until the 1990s.

Due to the focus on a genre, the interviewee acknowledges the need to identify the borderline examples of the genre:

“[…] it is a genre descriptive project, and therefore I need to find the borders of the genre. In that context, I think it is important to be able to search for content categories. That is categories for the branches of sport”.

Quote 19

In this way the information need is comparable to interviewee D’s need for generating an overview of the genre of traveller guide programmes. The interviewee’s television information need is divided into three:

1) The generation of an overview of sports broadcasts in Danish television;
2) Identification of borderline examples of sports broadcasts; and
3) Selection of specific sports broadcasts for analysis.

The interviewee aims at creating a general impression of the area before picking specific broadcasts. When asked how specific broadcasts are selected, the interviewee says that:

“[…] the television schedules were used to get the general picture, and then I tried to search for titles and persons, and I soon found out that I could not search for branches of sport”.

Quote 20

The selection of specific broadcasts that are relevant for further analysis is based on several aspects e.g., producers, commentators, anchor persons, branches of sport, duration, date and time of broadcasting, summary of content, and the context of the broadcast (horizontal programming). Due to the historical perspective of the information need, the availability of broadcasts is naturally also an important aspect. In
the historical perspective, the context of broadcasts becomes very important due to the fact that often features from programmes are stored, only, and the interviewee needs to be informed about the broadcasts, which surrounded the stored feature. An example of stored features is Harald Nielsen, a Danish footballer in the 1960s, showing training tips for kids (see e.g., Quote 38), where the surrounding studio parts have been lost.

In brief, the granularity level of needed broadcasts is programmes and features, and the information needs are of the types of known item and muddled topic or content needs, or the first and the seventh need in Ingwersen and Järvelin’s typology.

7.1.2.7 Interviewee G - The historical development of Danish news broadcasts

Similarly to earlier interviewees, interviewee G’s information need is divided into two:

1) Generating an overview of news broadcasts in Danish television; and
2) Selecting specific news broadcasts for analysis.

In the first part, the interviewee’s starting point is television schedules and broadcast summaries. Due to the specific type of broadcasts in question (news), the television schedules does not contain any content summaries, instead the interviewee uses programme manuscripts provided by DR. The interviewee emphasises that from a research perspective the need is for access to search for information about all transmitted broadcasts. Interviewee G puts it as follows:

"The main problem in DR’s broadcast retrieval system is that it is constructed as a database of archived broadcasts, whereas our fundamental need is to have an overview of transmitted broadcasts.”

Quote 21

As mentioned previously one of the purposes of generating an overview of the primary data is to conduct a horizontal programming analysis, and derive information about the development of different types of broadcasts or content, including the development in the distribution of a specific type of broadcasts. In this overview, information about all transmitted broadcasts is essential, whether archived or not. Knowledge about the programming is important in the selection of specific broadcasts for further analysis. In the present interview, interviewee G has a clear conception of the television broadcasts that are needed, namely randomly selected broadcasts from specific time-spans, similar to what is considered an idealistic primary source sampling by interviewee E (e.g., see Quote 18). That is, broadcasts are selected from a methodical perspective, based on an overview of the primary sources, including the identification of relevant time-spans. In
addition, the availability of archived material is naturally also important for selecting broadcasts from different time-spans as emphasised by the interviewee. A third element of interest in the selection is the audience ratings, and therefore, to interviewee G, an effective broadcast retrieval system is to:

“[...] construct a combined database where you have what DR has transmitted connected to what exists in the collections, and connected to the TV-Meter rating system. In this way, you combine the transmission dimension, the reception dimension, and the archive dimension in one system”.

Quote 22

The benefits of such a three dimensional system is emphasised several times during the interview. The granularity level of interest for interviewee G is features in programmes, and the interviewee stresses that in order to be able to select relevant broadcasts features, it is necessary that the individual features are indexed.

“You have to have categorisations for the individual programme features. If not, it is very difficult. Such categorisation could be on subject, country, or participants”.

Quote 23

Similar to interviewee C, interviewee G focuses on television news, at the granularity level of features, and both interviewees emphasise the need to search for participants in the broadcasts.

The interviewee’s information needs are twofold, namely

1) Generating an overview of news in Danish television; and
2) Selecting specific news for analysis.

The information needs are characterised as verificative and explorative muddled topical, or again as Ingwersen and Järvelin’s first and seventh type of information need.

7.1.2.8 Interviewee H - The historical development of Danish documentaries

Interviewee H’s work task is comparable to interviewee E, F, and G’s, since interviewee H focuses on the historical development of a genre of television broadcasts. In this case documentaries. At the time of the interview, interviewee H is finishing a book in order to present the results of the research project, and in this work, the interviewee needs to
verify different statements. This activity is comparable to verifying in Ellis’ model of information seeking features. Hence, the need is threefold:

1) Generating an overview of documentaries in Danish television;
2) Selecting specific documentaries for analysis; and
3) Verification of facts to support statements in publications.

In generating the overview, the interviewee uses television schedules. The summaries, duration and programming information in the schedules are to a large extent enough for generating the overview. In generating the overview, the interviewee needs access to broadcasts in order to understand what they convey. When asked where in the research process the interviewee needs access to broadcasts, the respond is:

“That is very early on in the project. Because transmission began in 1951 only few living people remember the first broadcasts [...] The broadcast were just store in the archive, and the content of the early broadcasts is a matter of conjecture, and as it turned out we got surprised [...] It transpires that almost all modern documentary genres already exist in the 1950s, not in the same extent and state of development as is found later on, but they exist. It was a surprise that the universe happened to be much richer than expected”.

Quote 24

Covering broadcasts that are more than 50 years old, gives rise to a need to see broadcasts early on in the research process, since these broadcasts are unfamiliar to the researcher. Indirectly, interviewee H expresses a need to have a ‘feeling’ about the primary sources, which are relevant for the task at hand, which is comparable to the thoughts expressed by interviewee E in Quote 17. In this line of thought, the interviewee reflects on the need for direct access to broadcasts, as interviewee H states that:

INTERVIEWEE H: “[…] the dream is that I as a scholar can access a database with the broadcasts directly from my desktop, and also see and download the broadcasts [...] DR Documentary’s homepage have made all information about programmes from 1994 and forth searchable, and the newest programmes are directly accessible from the homepage, which is a very good feature. In this way, things are getting better, but it is just a specific homepage from a small part of DR, and not the complete interconnection”.

INTERVIEWER: “How about the quality of the broadcasts?”
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

INTERVIEWEE H: “Well, for the first part of my need, where I need to find out: ‘what type of broadcasts is this? I can do with low quality broadcasts’.”

Quote 25

The identification of the broadcast does not require high-quality moving images. The interviewee just needs to recognise the broadcast or briefly see if the broadcast is of an already familiar type, which is similar to the need expressed by interviewee E in Quote 16. The selection of the specific broadcasts that the interviewee wishes to analyse is based on the extent of specific types of broadcasts in the programming. That is:

“[…] the selection of broadcasts is partly based on whether they seem to fill up much space in the programming. For instance, we had an impression of the amount of the interview programmes represented by Karl Bjarnhof, but now we got reel numbers to support our impression. Of course, this is a parameter, and when we saw that some more social reports, which we did not know, took up a lot of space in the programming at an early point in time, we also needed to access those. It has a lot to do with the space filled by different types in the programming. That is, ‘how many are there of these similar broadcasts, and does it indicate a tendency?’ Examples of such broadcasts are of course selected. Usually, the programmes are very structured just conveying different content, and after watching a few you are familiar with the characteristics of that type”.

Quote 26

In this way, the need is comparable to interviewee E’s need. The development of the information need occurs as the research progresses, and the occurrence of the three aspects of the need is not a strictly forward process. Instead, the needs occur concurrently. This said, the selection of broadcasts is always based on the interviewee’s knowledge about the primary data sources, and hence the prior explorative generation of an overview. Further, verification tends to be part of the verifying behaviour in Ellis’ terminology.

The interviewee emphasises that one of the main search entries needed is to be able to search for programmes by genre. However, he also acknowledges that the genre concept is rather difficult, since it is one of the major research foci in Media Studies. The complexity of the concept is exemplified with the development of documentaries as a genre:

“The concept ‘Documentary’ appears along the way, because concepts like report, feature, or montage are used instead. The radio montage, which is the forerunner for TV documentaries, is indeed not denoted radio documentary, but radio
montage, and the department that worked on them is denoted the radio feature department. This is of course a mess, and therefore a scientific view is needed in order to derive a coherent conception”.

_Quote 27_

Interviewee H goes on and stresses that the research already conducted gives a fair basis for a coherent genre concept:

“I would claim that on the basis of the present research in Denmark, Sweden, and Norway we have gathered enough knowledge about the characteristics of the genre and the sub-genres, that it is relatively easy to construct a lasting genre definition, for indexing of broadcasts. It is evident that it is based on fundamental research, and naturally, there will be developments, new ideas, and things that traverse. Nonetheless I claim that it is doable, at present”.

_Quote 28_

At the same time interviewee H acknowledges that there will always be a need to go beyond the developed genres, and explore new genres and trends in television. An example is given by describing the work of a student under his supervision:

“It is helpful if we are able to use the three categories: fact; fiction; and entertainment, and perhaps also news or news and other fact. The ‘full Monty’ is preferable, but even these three are very useful as a point of departure for the researcher, instead of being forced to construct it yourself. I have just had a visit from a student, who works on these new programmes, which she denotes ‘coaching programmes’. Magazine programmes where experts and ordinary people meet in order to correct some everyday problems, e.g., education of children. Those programmes that deluge out of the TV. She had to select the data collected from TV-Meter and then try to understand the data”.

_Quote 29_

The quote illustrates that interviewee H and the student accept that a fully functional genre categorical system of television broadcasts is not available at present. That is, that the actual genre in question is not part of the genre categorisation. This is evident in cases where the researcher aims at developing a genre definition, and in this way the example is similar to interviewee D’s need for broadcasts in order to develop the genre of traveller guide programmes. Nonetheless, a genre categorisation into broader categories is considered an important basis for deeper genre categorisations.
In summary, the interviewee seeks broadcasts at the granularity level of programmes, and the needs are the first, the fourth, and the seventh of Ingwersen and Järvelin’s types of information needs.

7.1.2.9 Interviewee I - Genre and reception analysis of Danish soap operas

Interviewee I’s work task is genre and reception analysis of five selected Danish soap operas. When asked about the selection of the five soap operas, the interviewee states that:

“[…] I could see it from the literature, but it was actually more through other research interests. That is, contacts who suggested that I look at specific examples”.

Quote 30

The clear perception is reached in discussions with colleagues, and by reading the literature of the field. As we have discussed in Section 3.3.2, humanists rely on informal sources in their seeking of previous literature (e.g., Ginman, 1983; Byström, 1999; Green, 2000; 1984). The behaviour expressed in Quote 30 indicates that the reliance on colleagues is important when seeking primary sources for analysis of television broadcasts as well.

The interviewee reflects on the use of a future broadcast retrieval system with reference to the work task in question, and stresses that it is difficult to imagine how the system can help selecting broadcasts in the first place, since this selection is based on so-called research analytical criteria:

“I find it difficult to imagine how a future broadcast retrieval system can help me select. That is, I need to use analytical criteria so to speak in order to clarify a research problem, e.g., in relation to a genre”.

Quote 31

The point is that the research dictates the broadcasts for the present interviewee. The broadcasts are selected prior to any searching, and the selection is for instance based on which broadcasts seem to be trend setting. In order to select broadcasts on this basis the interviewee needs to be familiar with the primary data sources as well as the field of study, prior to conducting a verificative search for the pre-selected programmes.

The interviewee knows the title of the serial that is requested, and the information need is of a verificative nature. However, since short titles are common in a television context, the same title can often be used for different types of broadcasts, over the years.
Therefore, the interviewee emphasises that an indication of genre is an important relevance criteria:

"It would be fine with a broad indication of genre. For instance, ‘Ugeavisen’ [≈The Weekly – a Danish soap opera] could very easily be a local news broadcast or something different, and with the genre indication you would know whether it was fiction or not”.

Quote 32

Another example is given later on in the interview:

"[…] genre should be available, because there is for instance a programme with Anders Lund Madsen [a Danish comedian] with the title ‘Taxa’, which was very confusing in my search [for the soap opera ‘Taxa’]. The title of the broadcasts often reappears”.

Quote 33

For four of the five selected soap operas, interviewee I needs to retrieve all the programmes in each serial. The fifth soap opera comprises 560 programmes or episodes, and it is impossible for the interviewee to analyse all of them. Therefore, the interviewee needs to select relevant programmes from this serial, an important search entry and relevance criterion in this selection is the indication of the broadcast’s placement within a serial, e.g., by indicating the episode number as well as the total number of episodes.

The retrieval of all (or a representative sample of all) broadcasts is connected to the initial stage of the work task. Throughout the progression of the work task (e.g., analysis of specific episodes or writing articles), the interviewee experiences needs for specific broadcasts. An example is given in the following quote:

"[…] then I think, in this analytical example I could really use the episodes where Andreas meets the woman, and I have to retrieve and view 10 episodes, because I know the incident is there, but I just can’t remember where”.

Quote 34

The information needed by interviewee I in order to identify these relevant broadcasts is a description of the course of events or a summary of each programme in the serial.
In brief, the granularity level of the interviewee’s need for broadcasts is programmes and serials, and the information needs are verificative, and for known topic or content. The first and the third of Ingwersen and Järvelin’s need types, respectively.

7.1.2.10 Summary of the nine interviewees’ information needs

This section provides a brief summary of the nine interviewees’ information needs. In Section 7.1.4, the information needs are discussed and related to previous research.

Table 7.2 summarises the nine interviewees’ needs for television broadcasts with respect to the types of broadcasts, and the granularity level of the needed broadcasts. Further, the table depicts how each interviewee’s need for television broadcasts is related to Ingwersen and Järvelin’s eight types of information needs (Ingwersen & Järvelin, 2005, pp. 292-293).

In common to all information needs of the nine interviewees, the four SUL interviewees, and all, but one, questionnaire respondents (see Quote 36 below), is that television broadcasts act as the objects of analysis, or primary sources for analysis. One of the SUL interviewees expressed this focus in the following way:

“There are broadcasts that act as expressions of a culture, and can be researched to derive knowledge about that culture. The researcher’s aim is not to gather knowledge about the broadcast as such. The broadcast is relevant because it acts as a symbol of something else. The perspectives for analysis of the broadcast may then be very different, but I would state that whoever you might ask, contextual knowledge is important, and not as much whether a dog appears in the picture or not”.

Quote 35

This statement is in line with previous research, which identifies human generated artefacts as the primary sources for analysis in the humanities (e.g., Wiberley & Jones, 1994, p. 503). All the interviewees have an active reading of the television broadcasts, in contrast to most viewers’ passive consumption in front of a TV-set (Auffret & Prié, 1999, pp. 319-320). The broadcasts are needed for fulfilling a research project or work task, and the interviewees active viewing includes thorough analysis, contextualisation, and interpretation of broadcast, with the intention of contributing to ongoing research by producing an academic product (e.g., a thesis, an article, an oral presentation, and/or a teaching example). This active consumption is different from the focus on information objects as knowledge generating, learning or entertaining objects, which is the general focus for users’ interaction with IR systems.
Table 7.2: Summary of the nine interviewees’ information needs, the levels of granularity of needed broadcasts, and the type of broadcasts needed. The numbers in the ‘Types of information need’ column refer to Ingwersen and Järvelin’s (2005, pp. 292-293) typology of information needs.

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Types of broadcast</th>
<th>Granularity</th>
<th>Types of information need</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Format versions;</td>
<td>Programmes</td>
<td>Known item (1); Known topic or content (3)</td>
</tr>
<tr>
<td></td>
<td>Reality shows</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Children’s broadcasts; Spots; commercials</td>
<td>Serials; Programmes; Features</td>
<td>Known item (1); Known topic or content (3)</td>
</tr>
<tr>
<td>C</td>
<td>News; reality shows; documentaries</td>
<td>Programmes; Features</td>
<td>Known topic or content (3)</td>
</tr>
<tr>
<td>D</td>
<td>Traveller guides</td>
<td>Serials; Programmes</td>
<td>Muddled topic or content (7)</td>
</tr>
<tr>
<td>E</td>
<td>Entertainment</td>
<td>Serials; Programmes</td>
<td>Known item (1); Factual data searching (4); Muddled topic or content (7)</td>
</tr>
<tr>
<td>F</td>
<td>Sports</td>
<td>Programmes; Features</td>
<td>Known item (1); Muddled topic or content (7)</td>
</tr>
<tr>
<td>G</td>
<td>News</td>
<td>Programmes; Features</td>
<td>Known item (1); Muddled topic or content (7)</td>
</tr>
<tr>
<td>H</td>
<td>Documentaries</td>
<td>Programmes</td>
<td>Known item (1); Factual data searching (4); Muddled topic or content (7)</td>
</tr>
<tr>
<td>I</td>
<td>Soap operas</td>
<td>Serials; Programmes</td>
<td>Known item (1); Known topic or content (3)</td>
</tr>
</tbody>
</table>

One point is important to mention from the questionnaires, namely the aspect conveyed in the following quotation:

“I use them [the broadcasts] often to get an overview of a specific subject at a macro level, prior to going deeply into the subject at a micro level, where I only use textual sources. They [the broadcasts] are good for an overall insight in a problem area. Apart from that, the broadcasts as such can of course be the area of research”.

Quote 36

The work task underlying the need expressed by this questionnaire respondent is concerned with exploring a subject area. The respondent’s motivation for accessing television broadcasts is to gather knowledge about a subject, or ‘extract knowledge’ from the content. This aim is different from the work tasks discussed above, which is to use broadcasts as the source of analysis. The ‘knowledge generating’ type of work task is comparable to a passive consumption of broadcasts. Quote 36 is the sole example of
such a ‘knowledge generating’ type of work task, and we do not consider it further in the present work.

The historical perspective of the Media Studies research is to some extent concerned with re-creating the history of Danish television. That is, the programmes do not exist, and the researchers use whatever information they can gain access to in order to understand and describe what was transmitted, and what the programmes were about. This knowledge seems to be a necessary first step in order to make further analysis of the historical development of Danish television broadcasting. To put it simple, the researchers need to know what was transmitted in order to research how it has developed. The Registrant (Hjarvard & Jespersen, 2006) is an attempt to construct and mediate such knowledge. The historical recreating act is connected to the relative recent access to, and hence scientific focus on television broadcasts in Denmark.

This summary ends our reporting of the nine interviewee’s information needs. Prior to discussing these needs in relation to previous research on information seeking behaviour, and presenting explicit conclusions in relation to research question 1, we report on the information needs expressed in the 108 questionnaire responses.

7.1.3 Information needs expressed in questionnaire responses

This section reports on the knowledge gained from the questionnaires about the respondents’ needs for television broadcasts. The section focuses specifically on the types of broadcasts, as well as the granularity level of the broadcasts needed by the respondents.

In the 3rd section of the questionnaire the respondents are asked to indicate the types of television broadcasts they typically need in order to fulfil the work task at hand (see Section 6.3.1 and Appendix 4). The respondent can tick off as many as he or she sees fit, and the number of types of broadcasts marked by one respondent ranges from none to all 17, with a mean of 4.4 different types of broadcasts marked by one respondent and a standard deviation of 3.0. Approximately 75% of the respondents tick off between 2 and 7 types of broadcasts (e.g., see Appendix 12). The distribution of the broadcast types marked to be most typically needed by the 108 questionnaire respondents is presented in Table 7.3 and Table 7.4. In the first column of Table 7.3 each of the 17 types of broadcasts are listed in accordance with the arrangement in the questionnaire (see Appendix 4). In the second column the distribution of marks from the 108 respondents are shown, both in figures and in percentages. Table 7.3 shows that the distribution does not follow the order in which the types of broadcasts were presented for the respondents in the questionnaire, which attests to the validity of our results.
**Table 7.3:** Distribution of the typically broadcasts needed by the respondents. The table is arranged similar to the arrangement of the types of broadcasts in the questionnaire, which is alphabetical according to the Danish terminology.

<table>
<thead>
<tr>
<th>Type of broadcast</th>
<th>All respondents (N=108)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact-Current Affairs</td>
<td>49</td>
</tr>
<tr>
<td>Fact-Documentaries</td>
<td>78</td>
</tr>
<tr>
<td>Fact-News</td>
<td>52</td>
</tr>
<tr>
<td>Fact-Educational and Informative</td>
<td>17</td>
</tr>
<tr>
<td>Fact-Other</td>
<td>22</td>
</tr>
<tr>
<td>Fiction-Drama</td>
<td>29</td>
</tr>
<tr>
<td>Fiction-Feature Films</td>
<td>29</td>
</tr>
<tr>
<td>Fiction-Serials</td>
<td>54</td>
</tr>
<tr>
<td>Fiction-Other</td>
<td>13</td>
</tr>
<tr>
<td>Children and Youth</td>
<td>27</td>
</tr>
<tr>
<td>Local or Regional</td>
<td>10</td>
</tr>
<tr>
<td>Presentations</td>
<td>18</td>
</tr>
<tr>
<td>Viewer Interaction</td>
<td>8</td>
</tr>
<tr>
<td>Commercials</td>
<td>31</td>
</tr>
<tr>
<td>Entertainment-Music</td>
<td>14</td>
</tr>
<tr>
<td>Entertainment-Sport</td>
<td>4</td>
</tr>
<tr>
<td>Entertainment-Other</td>
<td>19</td>
</tr>
</tbody>
</table>

In contrast, Table 7.4 is arranged according to the type of broadcast marked by most respondents. Further, two columns are added to Table 7.4. The second column shows the distribution of types marked to be most typically needed with reference to the respondents’ academic affiliation. The third column depicts the distribution in relation to the respondents’ academic status. In our discussion below, we firstly pay attention to the overall distribution, and secondly we discuss the distribution in relation to academic affiliation and academic status.

Table 7.4 depicts the overall frequency of the types of broadcasts, and the frequency in relation to the respondents’ affiliation and academic status. Overall, the type of ‘Documentaries’ is indicated by 78 of the respondents (72.2%), while approximately half of the respondents indicate that they usually need access to ‘Fiction serials’, ‘News’, or ‘Current Affairs’ broadcasts. The remaining 12 types of broadcasts are indicated by approximately a quarter of the respondents or less, with
‘Entertainment-Sport’ broadcasts being indicated by fewest respondents, namely 3.7%, only.

**Table 7.4**: Distribution of types of broadcasts typically needed for fulfilling the work task at hand, as indicated by the respondents in the questionnaire. The distribution shows the types marked by affiliation, and academic status, and the pool of all respondents. The types are arranged in accordance to their overall frequency.

<table>
<thead>
<tr>
<th>Type of broadcast</th>
<th>Affiliation (N=52)</th>
<th>Copenhagen (N=56)</th>
<th>Status Scholars (N=22)</th>
<th>Students (N=86)</th>
<th>All respondents (N=108)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact-Documentaries</td>
<td>36 (69.2%)</td>
<td>42 (75.0%)</td>
<td>16 (72.7%)</td>
<td>62 (72.1%)</td>
<td>78 (72.2%)</td>
</tr>
<tr>
<td>Fiction-Serials</td>
<td>23 (44.2%)</td>
<td>31 (55.4%)</td>
<td>7 (31.8%)</td>
<td>47 (54.7%)</td>
<td>54 (50.0%)</td>
</tr>
<tr>
<td>Fact-News</td>
<td>23 (44.2%)</td>
<td>29 (51.8%)</td>
<td>10 (45.5%)</td>
<td>42 (48.8%)</td>
<td>52 (48.1%)</td>
</tr>
<tr>
<td>Fact-Current Affairs</td>
<td>21 (40.4%)</td>
<td>28 (50.0%)</td>
<td>8 (36.4%)</td>
<td>41 (47.7%)</td>
<td>49 (45.4%)</td>
</tr>
<tr>
<td>Commercials</td>
<td>18 (34.6%)</td>
<td>13 (23.2%)</td>
<td>7 (31.8%)</td>
<td>24 (27.9%)</td>
<td>31 (28.7%)</td>
</tr>
<tr>
<td>Fiction-Drama</td>
<td>14 (26.9%)</td>
<td>15 (26.8%)</td>
<td>8 (36.4%)</td>
<td>21 (24.4%)</td>
<td>29 (26.9%)</td>
</tr>
<tr>
<td>Fiction-Feature Films</td>
<td>13 (25.0%)</td>
<td>16 (28.6%)</td>
<td>1 (4.5%)</td>
<td>28 (32.6%)</td>
<td>29 (26.9%)</td>
</tr>
<tr>
<td>Children and Youth</td>
<td>11 (21.2%)</td>
<td>16 (28.6%)</td>
<td>7 (32.7%)</td>
<td>22 (25.6%)</td>
<td>27 (25.0%)</td>
</tr>
<tr>
<td>Fact-Other</td>
<td>9 (17.3%)</td>
<td>13 (23.2%)</td>
<td>8 (36.4%)</td>
<td>14 (16.3%)</td>
<td>22 (20.4%)</td>
</tr>
<tr>
<td>Entertainment-Other</td>
<td>9 (17.3%)</td>
<td>10 (17.9%)</td>
<td>6 (27.3%)</td>
<td>13 (15.1%)</td>
<td>19 (17.6%)</td>
</tr>
<tr>
<td>Presentations</td>
<td>11 (21.2%)</td>
<td>7 (12.5%)</td>
<td>5 (22.7%)</td>
<td>13 (15.1%)</td>
<td>18 (16.7%)</td>
</tr>
<tr>
<td>Fact-Educational and Informative</td>
<td>11 (21.2%)</td>
<td>6 (10.7%)</td>
<td>9 (40.9%)</td>
<td>8 (9.3%)</td>
<td>17 (15.7%)</td>
</tr>
<tr>
<td>Entertainment-Music</td>
<td>7 (13.5%)</td>
<td>7 (12.5%)</td>
<td>1 (4.5%)</td>
<td>13 (15.1%)</td>
<td>14 (13.0%)</td>
</tr>
<tr>
<td>Fiction-Other</td>
<td>5 (9.6%)</td>
<td>8 (14.3%)</td>
<td>2 (9.1%)</td>
<td>11 (12.8%)</td>
<td>13 (12.0%)</td>
</tr>
<tr>
<td>Local or Regional</td>
<td>5 (9.6%)</td>
<td>5 (8.9%)</td>
<td>3 (13.6%)</td>
<td>7 (8.1%)</td>
<td>10 (9.3%)</td>
</tr>
<tr>
<td>Viewer Interaction</td>
<td>5 (9.6%)</td>
<td>3 (5.4%)</td>
<td>3 (13.6%)</td>
<td>5 (5.8%)</td>
<td>8 (7.4%)</td>
</tr>
<tr>
<td>Entertainment-Sport</td>
<td>4 (7.7%)</td>
<td>0 (0.0%)</td>
<td>2 (9.1%)</td>
<td>2 (3.3%)</td>
<td>4 (3.7%)</td>
</tr>
</tbody>
</table>

**Legend**: Association between indication of type of broadcast and ‘Affiliation’ or ‘Status’ is tested with Chi² (light grey cells) or Fisher’s Exact test (dark grey cells). Bold and italic signifies a statistical difference at α=0.05.

It is crucial to emphasise that the respondents are asked to mark the type(s) of broadcasts they typically need, and not to order the types in accordance to their perceived importance. Hence, Table 7.4 shows the distribution of the types of programmes that the respondents have marked as typically needed for completing the work task at hand. Therefore, Table 7.4 should not be used for exclusion of particular types of broadcasts from a future broadcast retrieval system. Instead, the knowledge gained about the types of broadcasts marked by the respondents provides information as to indicate the types of broadcasts that could be given the priority in the short-term. Here it is important to stress that a focus on providing access to the types marked by
many respondents is a self-fulfilling prophesy, in parallel to the Matthew effect in science (Merton, 1968). Generally speaking, the type(s) of broadcasts made accessible are the ones that will be requested and accessed by the users, since it allows for research on that type, which provides fertile soil for further research, and so on. For instance, from our interviews (e.g., interview E and F) we learn that entertainment and sports programmes have not historically been considered important for archival storage, and hence storage of these programmes has been neglected or stored programmes deleted at times where resources were few. Consequently, historical research in entertainment programmes is rendered difficult since a great deal of the programmes has been lost for posterity, and students often chose other types of broadcasts, or neglects a historical perspective in their research. This is expressed in the following quote from interviewee E:

“I remember when the reality wave started at the end of the 90s all students were saying ‘we are witnessing important changes, this is big, all new and completely different’. But in the beginning of the 70s we had a similar trend, where people experiment with social setups. They do it for different purposes than today, but they are reminiscent of each other. For instance, ‘Big Brother’ was made in the 70s by Poul Martinsen. He isolated people on an island and they had to build a bridge to get away from the island [‘Broen’ from 1969]. The students are thinking something like ‘has anybody made television before I was born?’ The lacking of historical knowledge is not good. It is difficult to create interesting research when the perspective is solely forward looking”.

Quote 37

If programmes are neglected in the broadcast storage and retrieval system, our contextual understanding is lost, and consequently new research areas are not discovered and investigated. The students lack of awareness of older broadcasts can to some extent be ascribed the lack of access to broadcasts in the present ‘system’. That is, it is difficult for scholars to include broadcasts in their teaching, and it is difficult for the students to generate an overview of the primary data as well as to obtain access to identified broadcasts. Humanists are dependent on primary sources not previously studied from the same perspective or with the same objective, and hence it would be problematic to disregard infrequently used broadcasts in a future broadcast system context. For instance, disregarding sports programme in the broadcast retrieval system would mean all together disregarding interviewee F’s needs for broadcasts. Therefore, we strongly recommend that broadcast system designers and constructors do not disregard any type of broadcasts. This should be seen in line with the fact that the test
participants’ needs firstly focus on all transmitted broadcasts, and secondly on the broadcasts preserved for posterity.

The distribution for the two departments of affiliation and the two types of academic status is important in order to see whether the association to either of the groups affects the respondents’ response on generally needed types of broadcasts. This is relevant for gathering knowledge about a future broadcast retrieval system. Different research perspectives in the two departments or between scholars and students might affect the test participants’ needs for television broadcasts. Knowledge about difference in aspects of information seeking behaviour, including information needs, between these groups is important in order to understand how an IR system might serve the users. Basically, if essential differences are found between these groups’ needs for television broadcasts, it should be considered whether one common broadcast retrieval system is capable of effectively handling the academics’ needs, or if specific broadcast retrieval systems should be aimed at particular groups of academics. The Chi\(^2\) test is generally applied to test for association, but when the expected count is less than five in more than 20% of the cells, we employ Fisher’s exact test of association (Cramer, 1998, p. 303). The test applied for a particular association is indicated with nuances of grey in Table 7.4. The significance level in these tests is set to 0.05 (see e.g., Section 6.3.5 on our approach for statistical analyses).

We are unable to identify significant differences between the two departments of affiliation. That is, generally no statistical significant differences are found between the respondents from the two departments. Therefore, we conclude that the departmental affiliation is not related to the indication of typical types of broadcasts.

Examination of the data for scholars and students, shows that the chi\(^2\) or Fisher Exact value is high enough to reject the null-hypothesis, indicating that a significant difference between the academic status and the marking of the types of ‘Fiction-Feature Films’, and ‘Fact-Educational and Informative’. Scholars mark significantly fewer feature films and significantly more educational and informative programmes, than students. Here it is important to acknowledge the legislative restraints on the present research access to television broadcasts, namely the fact that users do not have access to independently published broadcasts. Featured films are generally published independently, and can hence not be retrieved from the Danish national collection of television broadcasts, but need to be retrieved from other audiovisual collections at the State and University Library. Further, the fact that scholars, compared to students, generally are more experienced with borrowing broadcasts from the State and University Library (see Appendix 13), makes it reasonable to assume that the scholars are knowledgeable about the legislation. Therefore, it can be assumed that these legal
rules have affected the test participants’ need for this type of broadcast. In this way, we might explain the difference between scholars and students with respect to the type of feature films. For the broadcast type ‘Fact-Educational and Informative’, we do not have a similar explanation for the difference found between scholars and students.

In the free text field of the 3rd section of the questionnaire, few respondents point out that lifestyle programmes and reality programmes are missing from our typology. This is related to a discussion of a genre typology for the broadcasts, which is touched on in several of the interviews, and in our reporting on search entries and relevance criteria, Section 7.2 and Section 7.3, respectively. Though the generation of a genre typology is an important element of a future broadcast retrieval system, the specific construction of such a typology is beyond the objectives of the present work.

In the 4th section of the questionnaire, the respondents are asked to mark the level of granularity, which they find important for searching, and retrieval of broadcasts. This knowledge is important in order to gather knowledge about the preferable indexable units in a future broadcast retrieval system. The respondents mark the level of granularity in similar fashion as the types of broadcasts. Nine respondents did not mark any level of granularity, 16 marked one, 51 marked two, 23 marked three, and nine respondents marked all four levels of granularity. The distribution of the responses is depicted in Table 7.5. The table is generally constructed to resemble Table 7.4, which we presented in Section 7.1.3. The first column of Table 7.5 depicts the four levels of granularity. The second column shows the distribution of the respondents’ marks with reference to their affiliation. The third column shows the distribution with reference to the respondents’ academic status. The fourth column shows the overall distribution of the marks given by the respondents. In addition, Table 7.5 contains an extra column, which depicts the level of granularity expressed in the nine interviews. The low number of interviewees leaves some ambiguity regarding the distribution, and the column is included for comparison purposes, only.

Overall, Table 7.5 shows that 87.0% of the respondents find it important to be able to search and retrieve broadcasts at the ‘Programme’ level of granularity. 14 respondents did not mark this level of granularity. Nine of these respondents did not mark any of the four levels of granularity, and the remaining five respondents marked either ‘Serials’ (2) or ‘Features’ (3). 63.9% of the respondents marks the level of serials, 43.5% the level of features, and 12.0% the level of sequences. All 13 respondents who mark the sequence level have also marked the programme level, 11 of these have in addition marked the serials level, and nine respondents have marked all four levels of granularity. That is, the granularity level of sequences is only marked along with other levels of granularity, and often the respondent have hedge his or her
respond by marking all four level of granularity. In comparison, none of the nine interviewees had needs at the sequence level, four had needs at the features and serials levels, while all nine had needs at the programmes level. Hence, the distribution found in the interviews is roughly comparable to the distribution found for the questionnaire respondents.

Table 7.5: Distribution of the level of granularity. The distribution shows the granularity as preferred by affiliation, and academic status, and the pool of all respondents, as well as the nine interviewees. The table is arranged in accordance to the levels of granularity.

<table>
<thead>
<tr>
<th>Level of granularity</th>
<th>Affiliation</th>
<th>Status</th>
<th>All respondents</th>
<th>Interviewees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aarhus (N=52)</td>
<td>Scholars (N=22)</td>
<td>Students (N=86)</td>
<td>(N=9)</td>
</tr>
<tr>
<td></td>
<td>Copenhagen (N=56)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serials</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>67.3</td>
<td>34</td>
<td>60.7</td>
</tr>
<tr>
<td>Programmes</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>86.5</td>
<td>49</td>
<td>87.5</td>
</tr>
<tr>
<td>Features</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>32.7</td>
<td>30</td>
<td>53.6</td>
</tr>
<tr>
<td>Sequences</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9.6</td>
<td>8</td>
<td>14.3</td>
</tr>
</tbody>
</table>

Legend: Association between indication of level of granularity and ‘Affiliation’ or ‘Status’ is tested with Chi² (light grey cells) or Fisher’s Exact test (dark grey cells). Bold and italic signifies a statistical difference at \( \alpha = 0.05 \).

1) Percentages are ambiguous due to the low frequency of interviewees.

One of the interviewees had ticked off a need for broadcasts at the sequence level, however, during the interview this need turned out to be a need for analysis of specific broadcasts at the programme level. As the interviewee wishes to be able to (automatically) derive figures of distribution of specific sequences in known programmes, e.g., the distribution of sunsets. Therefore, we do not consider this a part of the information need, but rather, we see it as part of the analytical research work to be done after broadcasts are retrieved. This said, it is of course related to the information seeking behaviour, e.g., because digital broadcasts can be progressed automatically, and hence the form or medium of the broadcast (e.g., digitally or analogously stored) may be a relevance criterion. An examination of the responses in the free text field in the questionnaire associated to the sequence level of granularity reveals that several respondents describe a similar need for analytical research in pre-identified broadcasts.

We conclude that access to broadcasts at the granularity level of sequences is not essential for academics in Media Studies. Manual indexing of broadcasts at the level of sequences is very resource demanding, and, at present, automatic approaches are not capable of providing consistent and reliable results for a large heterogeneous collection of broadcasts, similar to the collection at the State and University Library. However, a
lot of focus is devoted to construction of automatic approaches, and the techniques are developing fast. Therefore, it may be of importance that the broadcasts are made available in a digital form and hereby enable the researchers to obtain the selected broadcasts, and use his or her automatic tool to conduct the relevant automatic analyses, e.g., by applying new tools or algorithms developed for the particular automatic analyses. In contrast to our discussion about the inappropriateness of disregarding infrequently used types of broadcasts, disregarding indexing of broadcasts at the sequence level does hamper research to the same extent.

Our comparison of the responses from the two departments shows that a significant relationship exists with regard to the level of features. In other words, the feature level is marked more often by respondents from Copenhagen than by respondents from Aarhus. We believe this is related to the types of broadcasts in focus in the different departments. Though the difference is not statistical significant, more respondents from Copenhagen have marked news broadcasts and current affairs programmes as depicted in Table 7.4. Broadcasts are characterised by comprising many features. The point we wish to make here is that the granularity needed is determined by the characteristics of the types of broadcasts needed. This point is further emphasised by the association found between the types of news and current affairs programmes, and the granularity level of features, which is outlined below in relation to Table 7.6.

In Table 7.6 we show the result of the comparison between how the respondents marked the types of broadcasts they typically need, and the level of granularity they find important for searching and retrieval. The Chi$^2$ and Fisher’s exact tests of association have been conducted for test of association between the 17 types and the 4 levels of granularity. In 12 cases, we identified a significant association between the responses. In other words, Table 7.6 shows whether the respondents’ indication of a typical need for a particular type of broadcast is related to an indication of a need for broadcasts at a particular level of granularity. Knowledge about the association between the types of broadcasts and the needed access level of granularity is important in order to understand the granularity level of indexing needed for the different types of broadcasts. An X in the table indicates that an association is found (see statistical details in Appendix 14).
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

Table 7.6: Relation between respondents marking of type of broadcast and level of granularity. Types of broadcasts are arranged according to overall frequency. The types of broadcasts are arranged according to the overall frequency in marking by the respondents.

<table>
<thead>
<tr>
<th>Type of Broadcast</th>
<th>Serials</th>
<th>Programmes</th>
<th>Features</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact-Documentaries</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fiction-Serials</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fact-News</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fact-Current Affairs</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Commercials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fiction-Drama</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fiction-Feature Films</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children and Youth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fact-Other</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entertainment-Other</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fact-Educational and Informative</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entertainment-Music</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fiction-Other</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local or Regional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viewer Interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entertainment-Sport</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend: Association is tested with Chi² (light grey cells) or Fisher’s Exact test (dark grey cells).

X indicates a significant statistical difference at α = 0.05.

The test of association does not tell us anything about the type of association found, but this is identified by examining each 2x2 matrix with statistical significant associations. The association between broadcast type ‘Fact-News’ and granularity level ‘Serials’ is identified to be a so-called negative association. The respondent who marks one (e.g., ‘Serials’) tends not to mark the other (e.g., ‘Fact-News’), and the other way around. In other words, respondents who typically need access to news broadcasts do not mark the serials level of granularity. This is in line with the fact that news broadcasts consists of many features, and that the association between ‘Fact-News’ and ‘Features’ is a positive association. The respondent who marks one is likely to mark the other as well. A respondent who needs news broadcasts is more likely to consider it significant to gain access at the features level of granularity. The remaining 10 identified associations are also positive. In the ‘Serials’ column, the ‘Fiction-Serials’ is self-evident, along with ‘Fiction-Drama’ and ‘Fiction-Other’ since these types of broadcast often are serials. ‘Entertainment-Other’ and ‘Entertainment-Music’ are often a series of broadcasts e.g.,
the music serial ‘Boogie’ on DR1. The type of ‘Fact-Other’ is not similarly evident to explain. Documentaries are most often stand-alone broadcasts, and the association with the ‘Programmes’ level of granularity is expectable. The association between programmes and fiction serials is explainable with reference to Interviewee I’s need to gain access to particular programmes in the serial. ‘Fact-Current Affairs’ are generally magazine programmes containing several feature broadcasts. Hence, the association with the features level of granularity is just as evident as the association between news broadcasts and the features level. Commercials are usually on the forefront with employing new mechanisms in order to get the messages through to the audience, e.g., by using fast cutting. In this way the association between ‘Commercials’ and ‘Sequences’ is related to our discussion about automatic tools for analysis of pre-selected broadcasts.

In brief, we conclude that access is needed at the programmes level of granularity, and to a lesser extent, the test participants express a need for access at the serials and features level. The relations found between the types of broadcasts and the levels of granularity needed for access does not reveal big surprises. The characteristics of the types of broadcasts determine to a large extent the level of granularity. For instance, programmes which are characterised by comprising many ‘independent’ features e.g., news or magazine programmes, entails access at the features level of granularity. In an IR system context this is related to the level of the analysable or indexable units, and it entails that a future broadcast retrieval system should index broadcasts at the serials, programmes, and features level of granularity, with programmes as the main indexable unit. In comparison to indexing in traditional library settings, the programmes level is comparable to the physical tangible units of documents (e.g., books), the serials level is derived from a traditional library setting (see Section 6.3.1) and is hence directly transferable to textual serials, and the features level is comparable to independent intellectual parts of a physical unit (e.g., the articles in a scientific journal or the intellectual parts of an anthology). In essence, serials, programmes, and features are all considered significant indexable units. That is, surrogate records should be constructed for broadcasts at all three levels. The programmes surrogate record is the main access level for the users, and hence programmes are the main indexable unit. The serial surrogate record collocate all programmes belonging to a particular serial. In comparison to textual information objects, it is similar to the information provided in the serial fields of a MAchine-Readable Cataloguing record (MARC record), or in the sixth area of an International Standard Bibliographic Description record (ISBD record). The features surrogate record provides access to the individual features of a programme. The features are component parts of a programme or a host item in MARC terminology.
(Library of Congress, 2007), and a relational links should naturally be established from the component or feature to the host item or the programme. Serials and features surrogate records should only be constructed for the relevant types of broadcasts but the present work is not designed to derive exact causal relations between each type of broadcast and the granularity level needed of the indexable units. A well defined, general agreeable, and easily understandable broadcast genre typology is a pre-requisite for such investigations which needs to be conducted with more analytical emphasis on the characteristics of the different types or genres of television broadcasts.

This ends our reporting in relation to the first research question. In the following, we summarise our results, and provide an explicit answer to research question 1.

7.1.4 Summary and concluding statements

The present section summarises the results that compose the answers to the first research question, concerning: what characterises a given group of users’ information needs in the context of television broadcasts. The investigated group of users investigated is scholars and students affiliated to two Danish University departments of Media Studies, as outlined in Section 5.3. Knowledge about these test participants’ information needs is mainly derived through in-depth interviews and web questionnaires, and we have identified five characteristics of the test participants’ television needs. These characteristics are interconnected and multifaceted, and consequently, our answer to the first research question comes with some length.

First, the test participants’ television needs are characterised by a need for the broadcasts to serve as data objects in scientific analyses, which is similar to our results from analysis of the genuine requests in Section 5.3.1. Though to some extent trivial this knowledge is important because it affects the way access is needed to the broadcasts. If the needed knowledge about the broadcasts is available, e.g., in a broadcast retrieval system, the test participants do not necessarily need to obtain access to the broadcasts. In this way, the effectiveness of a future broadcast retrieval system is not to be solely determined by the number of broadcasts accessed, since an effective IR system provides the academic with information about the broadcasts, and hereby actually reduces the need to gain access to specific broadcasts.

Second, the test participants’ needs are characterised by addressing three different dimensions of the broadcasts. These dimensions are:

1) The transmission dimension;
2) The archive dimension; and
3) The reception dimension
Chapter 7: Research area 1: aspects of information seeking behaviour

Media Studies is generally concerned with the broadcasts that are transmitted to the public, and hence the focus is on the transmission of broadcasts as expressed in Quote 21. The need for the transmission dimension is related to the test participants need to see each broadcast in its originally transmitted context or the programming as we denote it. The test participants first and foremost need to be able to search for all transmitted broadcasts. This is connected to the ‘criticism’ of production archives, for only conveying what has been stored, instead of conveying what has been transmitted, e.g., as expressed by Hjarvard (2006a, p. 22). In an IR system context this means that high recall is expected from the users and that representation of all transmitted broadcasts are searchable whether the broadcasts are stored or not. This entails that user equally should be able to search for existing as well as no-longer-existing television broadcasts, and hence that bibliographic records should be constructed for existing as well as non-existing broadcasts. This is fundamentally different from a traditional bibliographic context, where bibliographic records are defined as aggregates of data associated to real life entities, whether store in digital or analogue form (e.g., IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998, p. 7), and indexing is based on examination of the primary sources or documents (e.g., International Organization for Standardization, 1985; Lancaster, 2003). If surrogate records are to be constructed for no-longer-existing broadcasts it naturally follows that external sources are the basis for the data in such records, and hence the bibliographic records can only be as good as the information available in the external sources. Further, the test participants’ needs are concerned with information about whether the broadcasts have been stored. The so-called archival dimension. This dimension is important in order to know which broadcasts are available for the test participant, and hence can be selected for further analysis. However, prior to the selection the test participants need to be informed about the non-archived as well as the archived broadcasts, since the contextual information is important in order to identify relevant broadcasts. This need is related to a need to gather information about broadcasts in order to generate an overview of the primary sources. Furthermore, the test participants wish to be informed about the audiences’ reception of the broadcasts. The reception dimension provides further contextual information, which is important for the test participants information needs. The reception dimension is largely associated with TV-Meter’s different measures of audience rating, e.g., rating (average number of viewers), share of viewing (share of actual viewers), or reach (number of individual viewers) (Hjarvard, 2001, p. 17).
Essentially, this means that a future IR system ought to combine the transmission dimension of the Registrant, with the archive dimension of the stations databases (e.g., DR), and the reception dimension of TV-Meter into one coherent retrieval system.

Third, the test participants’ needs are characterised as concerning known items, known topic or content, muddled topic or content, and factual data. The two types of topic or content needs are related to the generation of explorative overviews of the primary sources, while the latter is related to obtaining access to specific broadcasts. The need to generate an overview is characterised to be explorative in nature, and it is connected to the development of the research and the work task (see e.g., Quote 20). In relation to Kuhlthau’s model of the information search process (Kuhlthau, 1991), the explorative need for an overview of the primary sources is comparable to the pre-focus vague conception of the work task at hand. To generate an overview of the primary sources is a step towards gaining a clear and focused perception of the work task at hand. Therefore, the two types of information needs go hand in hand, and they might appear several times during the fulfilment of a particular academic work task. For instance, interviewee H’s experience with gaining an overview requires the access to particular broadcasts early on in the research project (see Quote 24), and it reoccurs several times throughout the fulfilment of the work task. The need for television broadcasts is highly related to the different phases of the work task. This is in line with Vakkari’s work on the relation between the information need and the progression of the work task (Vakkari, 2001a). The information needs of academics in Media Studies are characterised by a progression in line with the progression of the research task. The needs for television broadcasts occur as the research progresses. In other words, a researcher might need television broadcasts in the beginning of the research project in order to generate an overview of the field, and then as the research on the initially selected television broadcasts progresses, new aspects or foci for analysis might appear for the academic and he or she might again need an overview, gain access to the specific broadcasts again, or to gain access to new broadcasts. For instance, interviewee H’s needs are explorative in the beginning but later on (e.g., as the interviewee is finishing a book on the subject) the needs are verificative and factual in nature, and are concerned with gathering specific facts about broadcasts and access to particular broadcasts. That is, factual data or known item searching in Ingwersen and Järvelin’s terminology. The distinction between the four part of the information needs are comparable to Meadows aforementioned division between know item retrieval, fact retrieval, subject retrieval, and explorative retrieval.

Fourth, the broadcast information needs are characterised to consist of four parts, which are depicted in Figure 7.1. Not all four parts are identified in each of the nine
interviews, and we do not claim that all television seekers’ needs consist of all four parts. Neither do we claim an exact order of appearance of the parts, though the order above seems to be intuitive. For instance, the generation of an overview or the selection and analysis of specific broadcasts may prompt a need or sub-task for generating a new overview with another perspective. The parts of the information needs may occur in an intermingled way, which is signified with dotted lines in the figure. Figure 7.1 bears a slight resemblance with Wilson’s (1999b, p. 255) depiction of Ellis’ information seeking features model. The overview part is comparable to the starting, browsing, and monitoring features in Ellis’ model, identification of borderline exemplars is comparable to the feature of differentiating, selection is comparable to extracting, and verification is comparable to verifying and ending. The present work did not aim at an investigation or verification of Ellis’ feature model in a television broadcasts context, and our study is hence not designed to give answers about the applicability of Ellis’ model in a television context. However, the identification of several of Ellis’ features is an indication of the general applicability of Ellis’ model, and future research could focus more specifically on its use in a television broadcast context.

As a prerequisite for identification of specific broadcasts for analysis the test participants’ need to generate an overview of the primary sources in their area of research. One test participant (interviewee B) states that this overview is generated through the everyday experience with the primary sources, whereas for the remaining interviewees it is an overview, which needs to be generated as part of the work task. Here it is important that the starting point for interviewee B is the reception of broadcasts, and hence that relevant programmes are selected by the receivers (the children in this case), see e.g., Quote 9. In contrast, the broadcasts are the point of departure in the work tasks discussed in the remaining eight interviews. In short, the broadcasts are selected as the objects of analysis, and subsequently the analysis may or may not include a given group of users and their use of the broadcasts.
The overview generating part of the test participants’ needs is of an exploratory nature where the test participants have a clear or a muddled perception of the topic or content of the broadcasts in question. It is comparable to Ingwersen and Järvelin’s third and seventh type of information need. The test participants with a muddled information need may have a clear perception of their work task (e.g., generation of a genre), but the complexity of the task may be considered to be very high, and especially in relation to identification and selection of relevant primary sources (broadcasts) for analysis. This is related to Byström and Järvelin’s studies (Byström & Järvelin, 1995), and we claim that the muddled needs are likely to be related to complex work tasks or sub-tasks. With reference to Kuhlthau’s stages in the information search process the explorative overview generating information needs are at the pre-focus stage, and the test participants are aiming at generating a clear focus of the primary sources in question.

In the selection part of the information need, the test participant has a very clear perception, or focus in Kuhlthau’s terminology, of the broadcasts that are needed. The overview instigates a verificative or known topical need for broadcasts, which is comparable to Ingwersen and Järvelin’s first and third type of information need.

It is important to differentiate between the need to generate an overview, and thereby be informed about all transmitted television broadcasts, and the need to gain access to television broadcasts in order to study the broadcast as such. Hereby the nine interviewees’ information needs can be summaries to consist of two types of needs:

1) A need to gain knowledge about broadcasts; and
2) A need to gain access to specific broadcasts.

The first type is for instance concerned with generating an explorative overview of some specific broadcasts e.g., a genre of broadcasts, including a brief examination and analysis of specific broadcasts, as expressed by interviewee E in Quote 16 or by interviewee H in Quote. Several of the interviewees express that they have gained extensive knowledge about the primary sources of their interest. For instance, by gathering information about the broadcasts from external sources. Knowledge about the broadcasts is a prerequisite for the second type of need, namely the identification or selection of the specific relevant broadcasts, which are needed for analysis. However, the overview of the primary sources may be generated prior to interaction with an IR system, e.g., as expressed by interviewee B in Quote 11.

Fifth, a characteristic of the test participants’ television needs is that they are concerned with all types of broadcasts, that access is needed at the serials, programmes and features levels of granularity, and that the characteristics of the different types or
genres of broadcasts determines the level of granularity needed. In a future broadcast retrieval context this mean that surrogate records should be constructed for serials of programmes, programmes, and features in programmes. However, the present work is not designed to investigate the characteristics of the different types of broadcasts, or the specific cataloguing practices to apply, and therefore it is recommended that future work focus on such issues.

This concludes our reporting regarding research question 1. In the following section, we focus on another aspect of users’ information seeking behaviour, namely the search entries preferred by test participants when searching for television broadcasts.

7.2 Research question 2: which search entries are preferred for searching of television broadcasts, and why and when are they preferred?

This section provides answers to the second research question regarding which search entries are preferred by the test participants, including when and why the entries are preferred. The ‘which’ of the question is primarily based on the questionnaire responses, and further expanded with knowledge about when and why from the nine interviews. In addition, the interviewees bring attention to search entries not included in the questionnaire. In the following Chapter on the appropriate access points for television broadcasts, we give attention to a comparison between access points in a traditional textual context and the television broadcasts context, but the present section only briefly touches on such comparisons where it is considered relevant for a coherent presentation.

In the fourth section of the questionnaire, the respondents are asked to mark the search entries they prefer for searching television broadcasts. The individual respondent can mark as many of the 20 search entries listed in the questionnaire as he or she feels like. The main part (93.5%) of the 108 respondents have marked between 5 and 16 search entries, with a mean of 10.3, and a standard deviation of 3.8 (see Appendix 15 for details).

Table 7.7 depicts the results of all 108 respondents’ marking of preferred search entries. In the first column, the search entries are listed in accordance to the way they are listed in the web questionnaire. The table also includes the illustrative search terms provided in the questionnaire (e.g., see Appendix 4). The second column contains a categorisation of the search entries. The categorisation is inspired by Hertzum’s (2003) six categories of so-called need attributes in film requests and the our categorisation comprises five categories: 1) bibliographic search entries; 2) screening search entries; 3)
content search entries; 4) archival search entries, and; 5) relational search entries. We have combined Hertzum’s content and subject categories, and added an additional ‘Archival’ category. Further, we do not apply Hertzum’s ‘Other’ category. Hertzum’s categories are described in Section 3.4.1. This categorisation is chosen in order to point to the similarities with surrogate records for textual information objects, as well as to point to the peculiarities of television broadcasts in a bibliographic retrieval context. The third column depicts the number and percentages of respondents who have marked the search entry in question.

**Table 7.7:** Search entries preferred by the respondents. The table is arranged similar to the arrangement of search entries in the questionnaire, which is alphabetical according to the Danish terminology.

<table>
<thead>
<tr>
<th>Search entry</th>
<th>Search entry category</th>
<th>All Respondents (N=108)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of production</td>
<td>Bibliographic</td>
<td>50  46.3</td>
</tr>
<tr>
<td>Date of transmission</td>
<td>Screening</td>
<td>86  79.6</td>
</tr>
<tr>
<td>Keywords</td>
<td>Content</td>
<td>90  83.3</td>
</tr>
<tr>
<td>Format of transmission (e.g., 16:9, 4:3)</td>
<td>Bibliographic</td>
<td>8   7.4</td>
</tr>
<tr>
<td>Genre</td>
<td>Content</td>
<td>87  80.6</td>
</tr>
<tr>
<td>Geographical location for content/storyline</td>
<td>Content</td>
<td>18  16.7</td>
</tr>
<tr>
<td>Geographical location for recording</td>
<td>Bibliographic</td>
<td>10  9.3</td>
</tr>
<tr>
<td>Time of transmission</td>
<td>Screening</td>
<td>56  51.9</td>
</tr>
<tr>
<td>Borrow options (e.g., access to broadcasts)</td>
<td>Archival</td>
<td>67  62.0</td>
</tr>
<tr>
<td>Medium (e.g., download, DVD, VHS)</td>
<td>Archival</td>
<td>76  70.4</td>
</tr>
<tr>
<td>Participant (e.g., interviewee, actor)</td>
<td>Bibliographic</td>
<td>79  73.1</td>
</tr>
<tr>
<td>Author (e.g., programme planner, scriptwriter)</td>
<td>Bibliographic</td>
<td>67  62.0</td>
</tr>
<tr>
<td>Audience ratings</td>
<td>Relational</td>
<td>53  49.1</td>
</tr>
<tr>
<td>Black/white or colours</td>
<td>Bibliographic</td>
<td>13  12.0</td>
</tr>
<tr>
<td>Spoken language</td>
<td>Bibliographic</td>
<td>21  19.4</td>
</tr>
<tr>
<td>With or without subtitles</td>
<td>Bibliographic</td>
<td>22  20.4</td>
</tr>
<tr>
<td>Title</td>
<td>Bibliographic</td>
<td>91  84.3</td>
</tr>
<tr>
<td>Transmitting channel (e.g., DR1, TV2, TV2 Zulu)</td>
<td>Screening</td>
<td>91  84.3</td>
</tr>
<tr>
<td>Duration</td>
<td>Bibliographic</td>
<td>56  51.9</td>
</tr>
<tr>
<td>Waiting on loan</td>
<td>Archival</td>
<td>73  67.6</td>
</tr>
</tbody>
</table>

Overall, Table 7.7 shows that 14 of the 20 search entries are marked by 46.3% of the respondents or more. These are five of the ten bibliographic search entries (‘Date of production’, ‘Participant’, ‘Author’, ‘Title’, and ‘Duration’), the three screening search
Chapter 7: Research area 1: aspects of information seeking behaviour

entries (‘Date of transmission’, ‘Time of transmission’ and ‘Transmitting channel’), two out of the three content search entries (‘Keywords’ and ‘Genre’), the three archival search entries (‘Borrow options’, ‘Medium’, and ‘Waiting on loan’), and the relational search entry (‘Audience ratings’). The remaining six search entries (‘Format of transmission’, ‘Geographical location for recording’, ‘Black/white or colours’, ‘Spoken language’, ‘With or without subtitles’, ‘Geographical location for content/storyline’) are marked by less than 25% of the respondents.

At first hand, the categorisation of medium as an archival search entry may seem strange. The reason for the categorisation is due to the non-physical nature of television broadcasts, and the fact that the archive decides the medium, or AV document in Auffret and Bachimont’s (1999, p. 60) terminology, for storage of the television broadcasts, as well as the medium for the users access to the broadcasts. These media include VHS tapes, DVD, download, streaming etc. The last search entry (‘Audience ratings’) is of a relational character.

Table 7.7 shows that the respondents’ do not merely mark the first search entries on the list. The two most frequently marked search entries are ‘Title’ and ‘Transmitting channel’, and they are listed as the 17th and 18th search entry in the questionnaire. Instead, the pattern of the respondents’ marking reflects a purposeful marking by the respondent, which increases the validity of our results regarding the respondents’ preferred search entries.

We report more specifically on each of the five categories of search entries in separate sections, and the 20 search entries in Table 7.7 are expanded with the supplementary search entries discovered in the nine interviews. In these discussions, we give a primary attention to the 14 search entries most often preferred for searching television broadcasts by the 108 respondents. In addition, the interviewees provided information about the importance of specific search entries, detailed aspects of other entries, and essentially, when and why the search entries are preferred. In this discussion we include the general functional requirements for bibliographic records (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998). We do not consider the relational links between the four entities in the framework, namely work, expression, manifestation and item, but such relational links are naturally important in order to bind the collection of television broadcasts to the remainder of the bibliographic universe, e.g., the work ‘Hamlet’ in its many different expressions and manifestations, including the original book by Shakespeare, a featured film (e.g., transmitted via television), and the television transmission of Hamlet as a musical by Århus Studenterrevy [Aarhus Student Revue].

7 Transmitted on Monday 28th of March 1960 from 8 pm till 10.15 pm (Hjarvard & Jespersen, 2006).
the difference between responses from test participants with different affiliation or different academic status, before summarising, discussing and providing answers to the second research question in Section 7.2.7.

7.2.1 Bibliographic search entries

Three of the 14 frequently marked search entries are related to the title or statement of responsibility in a bibliographic sense. These are ‘Title’, ‘Participant’, and ‘Author’. In the nine interviews, the ‘Title’ stands out as one of the main search entries. Titles of television broadcasts are often short as expressed by interviewee I (see Quote 32 and Quote 33), and the titles are not very informative, e.g., in comparison to titles in scientific LIS articles. Consequently, ‘Title’ is generally used for retrieval of known items (see e.g., Quote 6 by interviewee A). In relation to the general functional requirements for bibliographic records (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998), the ‘Title’ is used to search for and find the manifestation or the broadcast. Further, ‘Title’ is the main search entry for Interviewee E and H when conducting factual data searching.

The statement of responsibility is divided between participants and authors. ‘Participant’ denotes the responsible body of persons in front of the camera, e.g., an interviewee, a host in a game show, or an actor, while ‘Author’ denotes the responsible individuals behind the camera, e.g., a producer, a subtitler, or an animation designer. The statement of responsibility search entries are used for finding all broadcasts by key responsible persons (see e.g., interviewee D), or broadening the overview of the primary sources. For instance, interviewee F searched the archives of DR by known participants in order to find programmes, which were not categorised to be sports programmes. In other words, searching for known responsible persons is important in order to find borderline examples in genre studies.

In addition to the listed search entries four respondents used the free text field in the questionnaire, to emphasise that a specification of the roles of the ‘Author’ is needed, e.g., by indication of ‘director’, ‘cutter’, ‘cameraman’, or ‘image medium translator’. Similarly, the two search entries of ‘Author’ and ‘Participant’ are further divided by the interviewees, whom find it important to be able to search for the particular role of the responsible persons. The exact specification or categorisation of the roles of the responsible people involved is beyond the scope of the present work, but it is an important aspect for further studies.

‘Date of production’ is marked by 46.3% of the respondents, but in the interviews, we are not informed about the use of ‘Date of production’ for searching broadcasts. The data of production coincides with the date of transmission for live television broadcasts,
e.g., news broadcasts, while for ‘delayed’ transmissions the two dates are naturally dissimilar. If the programme is a re-run of a live transmission, the date of production coincides with the information about previous transmissions or the so-called transmission pattern. In several of the nine interviews, we are informed about the importance of other production related search entries, e.g., the producing/responsible department, the producing country, and copyright holders in relation to screening. Producing/responsible department and the copyright information regarding screening of the broadcast is the main information needed in interviewee A’s need for knowledge about format versions. Compared to a motion picture context, the information needed is similar to the ‘based on the book’ relation in bibliographic records (see Appendix 16). For other interviewees, e.g., interviewee F and interviewee H, information about producing or responsible department and country is used in order to gather information about the perspective of broadcasts. An example of this is given in the following quote:

“‘Harald Nielsen teaches kids in technical skills’. There is a series of such broadcasts, and they are not produced by the Sports Department, but instead by the Children and Youth Department and that is an essential point. Similarly, critical documentaries about the Olympic Games or the International Olympic Committee are related to the Culture Department, and are typically imported. Knowledge about the origin of broadcasts is essential to me”.

Quote 38

When a person is knowledgeable about the domain and the primary sources of that domain, the departments or countries that are responsible for a broadcast gives indication as to the content of, or perspectives conveyed in the broadcast. In this way, the production related search entries are used as a mean to access the broadcasts in relation to the content they convey. These production related search entries are comparable to the publisher information in a textual context, and they are used for differentiating or identifying the broadcasts from other broadcasts at search time as well as when assessing the relevancy of the broadcasts. The latter aspect is further expanded in Section 7.3. One of the six infrequently marked search entries in Table 7.8 is concerned with production related issues, namely ‘Geographical location of recording’. This search entry is not discussed in any of the nine interviews, and we are consequently unable to state information about when and why the search entry is applied.

Five of the 20 search entries listed in Table 7.8 are concerned with the physical description in a bibliographic sense. These are: ‘Duration’, ‘Subtitles’, ‘Spoken language’, ‘Black/white or colours’, and ‘Format of transmission’. ‘Duration’ is
marked by 51.9%, while the remaining four are among the six infrequently marked search entries. ‘Duration’ is comparable to indication of the extent of the information objects in the form of pages in a textual context. ‘Duration’ is mainly used for differentiating between broadcasts, and it is applied in combination with other search entries, e.g., keywords. The interviewees also focus on duration as a relevance criteria and this use is discussed in Section 7.3. The interviewees briefly touch on the four infrequently physical description search entries, ‘Subtitles’, ‘Spoken language’, ‘Black/white or colours’, and ‘Format of transmission’. These search entries are used for narrowing the search to programmes with such specific characteristics. Here it is important to emphasise that the test participants’ research foci makes it important to investigate previously unexplored broadcasts, or investigate broadcasts from different perspectives. Therefore, all imaginable broadcast characteristics may turn out to be important in specific future research perspective. For instance, interviewee A states that it is important to be informed about all facts about or characteristics of a broadcast. Similarly to ‘Duration’, these search entries are used as secondary search entries, e.g., by specifying ‘spoken language’ in a topical search.

7.2.2 Screening search entries

‘Channel’, ‘Date of transmission’, and ‘Time of transmission’ are concerned with transmission or screening of the broadcast. ‘Channel’ is marked by 84.3% of the respondents and it is hence a major search entry. Research in Media Studies is often concerned with the development of broadcasts from a single station (e.g., DR1’s coverage of the Iraq war) or a comparative investigation of two or more channels (e.g., comparison between DR1’s and TV2’s coverage of the Iraq war). In any case, the channel of origin is often a known and important parameter for Media Studies research, hence it is used for narrowing the search, and it is naturally used in combination with other search entries. In all work tasks with a historical perspective the channel is not an issue since DR is the sole channel prior to 1988. However, the interviewees mention it as an important entry in relation to needs for more contemporary broadcasts.

Interviewee E and G describe a so-called ideal selection of broadcasts, in which broadcasts with some specific characteristic (e.g., the genre of news broadcasts) are selected based on date and time of transmission. The selection is methodically based, and the interviewees find the search entries essential for ensuring sound methodical research.

Further, the interviewees express a need to be able to search for more transmission information, e.g., to be able to search for first time transmissions, only. The heavy use of repeated programming in the television context makes information about
transmission pattern very important in retrieval of broadcasts. Whether the user is retrieving a first time transmission or a re-run is for instance significant in relation to the audience rating. That is, the audience rating may very well be significantly different for re-runs than for first time transmissions, and since ‘Audience rating’ is an important relevance criterion (see the following Section 7.3) the screening entries including transmission pattern are important search entries. The transmission pattern is explicitly mentioned by interviewee I in relation to retrieval of serials, and the serial title, as well as the placement within a serial are mentioned to be important search entries.

The screening search entries are particular to television broadcasts in comparison to textual information objects, and they function as secondary search entries.

7.2.3 Content search entries

Two of the highly marked search entries in Table 7.8 are concerned with the topic or content of the broadcasts, namely ‘Genre’ and ‘Keywords’. Genre gives an indication of the content and the form of the broadcasts in line with Bazerman (1994) and Swales’ (1990) conceptions of the concept. Similar to ‘Channel’, ‘Genre’ is mentioned as a search entry for narrowing a search to particularly interesting broadcasts. Such narrowing is an important point of departure for the test participants’ information needs, since they are often concerned with specific genres of broadcasts, and are hence able to specify the genres that are either relevant or irrelevant for their study. Overall, genre is the search entry that is emphasised to be of most importance for retrieval of broadcasts in the nine interviews. Nevertheless, the interviewees also point to the difficulties in constructing and applying a genre typology. The genre concept is an ever evolving aspect and a genre categorisation is considered an essential basis for conducting further scientific development of the genre concept, e.g., as expressed by interviewee H in Quote 28 and Quote 29. ‘Genre’ is used as a primary topical search entry, and as a secondary search entry in combination with other search entries e.g., the screening access points for differentiating between broadcasts.

In addition to genre, several interviewees point out that a controlled list or typology of keywords is needed in order to describe the content of the broadcasts. This is for instance pointed out in relation to needs at the features level of granularity, e.g., news broadcasts by interviewee G in Quote 23. Keywords are also essential for interviewee C’s need for broadcasts about a specific named corporation, but it is also considered as important by interviewee D and F for exploring the primary sources in the investigation of a particular genre, and especially in relation to searching for borderline examples of a genre.
The sixth less frequently marked search entry ‘Geographical location of content/storyline’ is a topical or content related search entry. The interviews do not provide information about the use of the search entry, and together with the other five infrequently marked search entries we consider it of minor importance for test participants’ searching of television broadcasts. However, the search entry might be essential for other user groups with different purposes and work tasks, e.g., journalists.

In Quote 39, interviewee G also expresses a need for free text searching in summaries. Free text searching in summaries is not listed in the questionnaire, but the feature is called for by several of the interviewees.

Interviewee C and interviewee G express a need for searching in the spoken words of a broadcast as expressed in the following dialogue with the interviewer:

INTERVIEWEE G: “[…] it should be possible to conduct free text searches”
INTERVIEWER: “In the spoken word or in summaries?”
INTERVIEWEE G: “In a summary. Well, that depends. It is obvious that if we imagine that the broadcasts are digitalised, and it is possible to search in the broadcasts then”
INTERVIEWER: “When you say ‘search in’ do you then mean the spoken words?”
INTERVIEWEE G: “Yes, if you could imagine that it is possible to search in the broadcast itself. Is that what you are thinking?”
INTERVIEWER: “Yes”
INTERVIEWEE G: “Then it would be a whole different ball game. That would be amazing, but I do not imagine that is a realistic possibility, though it might be. I do not know enough about IR systems to make my mind up about that”.

Being able to conduct searches in the spoken words of the broadcasts is considered a utopian possibility, and the interviewee does not consider the possibility until prompted about it. However, when prompted, the interviewees are very fond of the possibility of having that search entry as an option. Interviewee C and interviewee G both have needs to access specific features of news broadcasts, and their appraisal of ‘Spoken words’ as a search entry is very likely connected to the type of broadcasts in question.

These content and topical search entries are naturally used as primary search entries for topical searching, whether the users’ perception of the topic is muddled or clear, but at different levels. In this way, they give rise to a discussion of the indexing policy to apply in a television context, including the depth or exhaustivity of indexing,
Chapter 7: Research area 1: aspects of information seeking behaviour

and the type of aboutness. These indexing policy issues are briefly touched on in Chapter 0, and further discussed in relation to indexing of the broadcasts in Chapter 9.

7.2.4 Archival search entries

The search entries ‘Medium’, ‘Waiting on loan’ and ‘Borrow options’ are important for the interviewees in order to be informed whether the broadcast is available, and the form it is available in. The search entries are used for narrowing a search and the entries are for instance mentioned in relation to teaching work tasks, where the broadcasts are needed within a tight time schedule and examples need to be obtained prior to teaching. The three search entries are related to issues of an archival nature. The archival search entries are strongly related to copyright issues, and the legislation on access to broadcasts, but again it is important to emphasise that the test participants are interested in being informed about all transmitted broadcasts, including broadcasts, which they cannot gain access to for one reason or another. Hence, the archival search entries are used as secondary search entries for narrowing the search or differentiating between broadcasts, and it is closely related to the use of the information for relevance assessments.

7.2.5 Relational search entries

‘Audience ratings’ is the only relational search entry listed in the questionnaire. ‘Audience ratings’ comprises different measures of the impact of the broadcast in question. Highly rated broadcasts have many viewers and are consequently considered to be more interesting in the eyes of the broader public. With reference to Ellis and colleagues’ eight information seeking features (Ellis, Cox & Hall, 1993), audience ratings are generally used for starting or differentiating seeking activities, and in this way it is also related to assessing the relevancy of broadcasts for the work task at hand. Further, audience ratings are comparable to the bibliometrician’s use of citations in the identification and selection of primary data for analysis. The test participants are often interested in retrieving broadcasts, which have a high impact in the public, and audience ratings are the most easily available measure to give such an indication of societal impact. ‘Audience ratings’ are used as a secondary search entry in topical information needs.

The preferred search entries are naturally related to the level of granularity of the needed broadcasts. Interviewee F and G both emphasise the need for a table of content of the features in a broadcast, similar to the table of content found in a textual context, e.g., a book. The need is expressed in the following quote:
“It is essentially a table of content with indication of duration of each feature [...] 
It would be optimal if that information was available”.

In contrast, interviewee I needs to be able to search for series of soap operas, including the placement of a particular programme within the serial. As mentioned earlier (see Section 7.1.3), the granularity level of the needed broadcasts is related to the type of broadcast, which naturally affects the search entries preferred. In other words, the need to search for the sequential order of broadcasts in a serial is relevant in a soap opera context, while irrelevant in a news context, where the table of content is relevant instead.

These two internal relational search entries are related to our discussion of the granularity of the indexable units in a future broadcast retrieval system. Based on the knowledge gained in Section 7.1.3 about the indexable units, and the associated surrogate records, relational links are to be established between ‘Programmes’ and ‘Serials’ and ‘Features’, respectively. For instance, with insight from a textual context it is evident that the title of serials, programmes, as well as features should be searchable for the future users of the broadcast retrieval system. We did not focus explicitly on the association between search entries and the level of granularity in our data collection, but it is obvious that some search entries are related to particular types of broadcasts, and hence particular levels of indexable units. An example is interviewee C and G’s mentioning of spoken words as a search entry, and its relation to news broadcasts or interviewee I’s mentioning of transmission pattern in relation to serials. Here the focus has been on broadcasts at the programmes’ level of granularity. When gathering knowledge about the characteristics of the different types of broadcasts, further research should focus on the search entries associated to the serials as well as features level of indexable units. For instance, serial titles often coincide with programme titles (e.g., see the series in the weekly schedule in Appendix 5). In contrast, feature titles are generally not explicitly stated, consider e.g., the titles of the features in a news broadcast. Consequently, features titles are to be constructed by the indexer or derived elsewhere, e.g., from programme manuscripts or other external sources. The construction of television broadcast access points is discussed in Chapter 9.
7.2.6 **Comparison between search entries and affiliation or academic status**

In Table 7.8 the search entries mark by respondents are listed according to the overall frequency. The overall frequency is depicted in column four, in figures as well as in percentages. In the first column, each of the 20 search entries are listed according to the overall frequency depicted in column four. The letter in brackets indicates whether the search entry is characterised as bibliographic (B), screening (S), content (C), archival (A), or relational (R). Similar to Table 7.4, the second and third columns of Table 7.8 show the distribution of the search entries preferred by the respondents, according to the respondent’s affiliation and academic status. We test for association between marking a search entry to be preferred for searching, and the respondents’ affiliation and academic status, respectively. The colour of the cells indicate whether the chi² test of association (light grey) or Fisher’s exact test of association (dark grey) is applied (see Section 6.3.5 for our reasoning for applying the two tests of association). Further, in the two cases where a statistical significant association is found, the numbers are shown in a bold typeface.

We are not able to identify a difference between the responses from respondents affiliated to the two departments, except for the search entry ‘Waiting on loan’; at which respondents from Aarhus have a significantly higher positive response. Apparently, respondents from Aarhus find it more important to be informed about the time they have to wait before they receive the broadcast. This might be related to the geographical distance of the two departments and the State and University Library. The department at Copenhagen University is located more than 150 km from the State and University Library, while respondents at the department at Aarhus University, can walk the distance in 10 minutes. Therefore, the respondents in Copenhagen might to a larger extent have found other ways of accessing broadcasts (colleagues, small private archives, etc.). By constructing such alternative ways of access, the respondents affiliated to Copenhagen University are less dependent on the collection at the State and University Library. Another explanation is that due to the geographical distance respondents from Copenhagen know that they have to wait in order to receive a broadcast by postal mail, so the waiting is unavoidable, and consequently the waiting time does not matter.
Table 7.8: Comparison of search entries preferred by respondents according to affiliation, academic status, and the pool of all respondents. The search entries are listed in accordance to the overall marking by the 108 respondents.

<table>
<thead>
<tr>
<th>Preferred search entries</th>
<th>Affiliation</th>
<th></th>
<th>Status</th>
<th></th>
<th>All Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aarhus (N=52)</td>
<td>Copenhagen (N=56)</td>
<td>Scholars (N=22)</td>
<td>Students (N=86)</td>
<td>Respondents (N=108)</td>
</tr>
<tr>
<td>Transmitter channel (S)</td>
<td>47</td>
<td>44</td>
<td>18</td>
<td>81.8</td>
<td>91</td>
</tr>
<tr>
<td>Title (B)</td>
<td>43</td>
<td>48</td>
<td>18</td>
<td>81.8</td>
<td>91</td>
</tr>
<tr>
<td>Keywords (C)</td>
<td>42</td>
<td>48</td>
<td>18</td>
<td>81.8</td>
<td>90</td>
</tr>
<tr>
<td>Genre (C)</td>
<td>41</td>
<td>46</td>
<td>17</td>
<td>77.3</td>
<td>87</td>
</tr>
<tr>
<td>Date of transmission (S)</td>
<td>45</td>
<td>41</td>
<td>20</td>
<td>90.9</td>
<td>86</td>
</tr>
<tr>
<td>Participant (B)</td>
<td>34</td>
<td>45</td>
<td>15</td>
<td>68.2</td>
<td>79</td>
</tr>
<tr>
<td>Medium (A)</td>
<td>36</td>
<td>40</td>
<td>18</td>
<td>81.8</td>
<td>76</td>
</tr>
<tr>
<td>Waiting on loan (A)</td>
<td>40</td>
<td>76.9</td>
<td>33</td>
<td>58.9</td>
<td>73</td>
</tr>
<tr>
<td>Author (B)</td>
<td>29</td>
<td>55.8</td>
<td>13</td>
<td>59.1</td>
<td>67</td>
</tr>
<tr>
<td>Borrow options (A)</td>
<td>32</td>
<td>61.5</td>
<td>14</td>
<td>63.6</td>
<td>67</td>
</tr>
<tr>
<td>Duration (B)</td>
<td>28</td>
<td>53.8</td>
<td>13</td>
<td>59.1</td>
<td>56</td>
</tr>
<tr>
<td>Time of transmission (S)</td>
<td>31</td>
<td>59.6</td>
<td>17</td>
<td>77.3</td>
<td>56</td>
</tr>
<tr>
<td>Audience ratings (R)</td>
<td>26</td>
<td>50.0</td>
<td>13</td>
<td>59.1</td>
<td>53</td>
</tr>
<tr>
<td>Date of production (B)</td>
<td>24</td>
<td>46.2</td>
<td>10</td>
<td>45.5</td>
<td>50</td>
</tr>
<tr>
<td>With or without subtitles (B)</td>
<td>8</td>
<td>15.4</td>
<td>4</td>
<td>18.2</td>
<td>22</td>
</tr>
<tr>
<td>Spoken language (B)</td>
<td>7</td>
<td>13.5</td>
<td>4</td>
<td>18.2</td>
<td>21</td>
</tr>
<tr>
<td>Geographical location for content/storyline (C)</td>
<td>7</td>
<td>13.5</td>
<td>11</td>
<td>19.6</td>
<td>18</td>
</tr>
<tr>
<td>Black/white or colours (B)</td>
<td>5</td>
<td>9.6</td>
<td>4</td>
<td>18.2</td>
<td>13</td>
</tr>
<tr>
<td>Geographical location for recording (B)</td>
<td>3</td>
<td>5.8</td>
<td>7</td>
<td>12.5</td>
<td>3</td>
</tr>
<tr>
<td>Format of transmission (B)</td>
<td>2</td>
<td>3.8</td>
<td>3</td>
<td>13.6</td>
<td>8</td>
</tr>
</tbody>
</table>

Legend: Association between indication of the search entries marked and ‘Affiliation’ or ‘Status’ is tested with Chi² (light grey cells) or Fisher’s Exact test (dark grey cells). Bold and italic signifies a statistical difference at α=0.05.

Generally, the scholars and the students have marked the same search entries, and only for the search entry ‘Time of transmission’ the chi² value is high enough to reject the null-hypothesis, indicating that a significant difference between scholars and students is found for this variable. It does not seem clear to us why this difference exist. One explanation might be the scholars’ larger experience (see Appendix 13), and possibilities for borrowing from the State and University Library, and therefore they are very aware of the importance of being able to state the time of transmission in order to receive a copy from the collection. Three of the scholars in our interview express that during the retrieval of broadcasts for their work tasks, the identification of time of
transmission was a problematic issue, and a data element which they spend considerable
time at identifying.

7.2.7 Discussion and summarising statements

In the present section, we provide discussion and summary of the search entries
preferred by the test participants. Further, the section provides answers to the second
research question stating: which search entries are preferred for searching of television
broadcasts, and why and when are they preferred?

Table 7.9 encapsulates the search entries preferred by the test participants for
searching television broadcasts, and in this way Table 7.9 answers, which search
entries, are preferred by the test participants. The first column contains the five
categories described above, and the second column contains the sub-categories derived
through our description of the search entries. Hereby, the bibliographic search entry
category is divided into three sub-categories (Title and Statement of responsibility,
Production, and Physical description), and the relational category is divided into three
sub-categories: 1) serials; 2) features; and 3) external). In the third column, we list the
search entries that belong to each category and subcategory. In addition to the search
entry ‘names’, the overall percentages from the questionnaire are provided for each
search entry. For the search entries derived through the nine interviews this information
is naturally not available, hence the (n/a).

The 20 search entries from the questionnaire responses and the additional eight
search entries identified through the nine interviews are placed in one of the five
categories. For each of the search entries applied in the questionnaire survey, we have
added the percentage of respondents who have marked the search entry. Similar figures
are obviously not available in the additional eight search entries derived from the
interviews.

With reference to the distribution of marked search entries in Table 7.8, the two
sub-categories of ‘Physical description’ and ‘Production’ comprise the search entries,
which are marked by the fewest respondents. Apart from ‘Duration’ (51.9%) and ‘Date
of production’ (46.3%), the remaining six search entries are marked by 1/5 of the
respondents or less. The search entries in the screening category are highly marked
compared to the results by Hertzum (e.g., ‘Channel’ and ‘Date of transmission’ are
marked by approximately 80%). This difference is probably due to the different
contextual and situational settings of the two studies. Hertzum’s work is concerned
with information needs at a film archive, while we are concerned with information
needs in the context of transmitted television broadcasts. In the television context,
screening is often the point of departure for the work task, or a known feature, which can be used to narrow down the retrieved records.

Table 7.9: Categorisation of search entries for television broadcasts.

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-category</th>
<th>Search entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographic</td>
<td>Title and Statement of responsibility</td>
<td>Title (84.3%); Participant (73.1%); Author (62.0%)</td>
</tr>
<tr>
<td></td>
<td>Production</td>
<td>Date of production (46.3%); Geographical location of recording (9.3%); Production/responsible department (n/a); Production country (n/a); Copyright information (screening) (n/a)</td>
</tr>
<tr>
<td></td>
<td>Physical description</td>
<td>Duration (51.9%); Subtitles (20.4%); Spoken language (19.4%); Black/white or colours (12.0%); Format of transmission (7.4%)</td>
</tr>
<tr>
<td>Screening</td>
<td></td>
<td>Channel (84.3%); Date of transmission (79.6%); Time of transmission (51.9%); Transmission pattern (n/a)</td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td>Keywords (83.3%); Genre (80.6%); Geographical location of content/storyline (16.7%); Spoken words (n/a); Summary (n/a)</td>
</tr>
<tr>
<td>Archival</td>
<td></td>
<td>Medium (70.4%); Waiting on loan (67.6%); Borrow options (62.0%)</td>
</tr>
<tr>
<td>Relational</td>
<td>Serials</td>
<td>Placement within a serial (n/a)</td>
</tr>
<tr>
<td></td>
<td>Features</td>
<td>Table of content (n/a)</td>
</tr>
<tr>
<td></td>
<td>External</td>
<td>Audience ratings (49.1%)</td>
</tr>
</tbody>
</table>

The answers to the why and when parts of the research question are more complex. In the following, we mainly summarise our answers in relation to the types of information need, as identified previously. The ‘Title’ is the primary search entry applied when the test participants search for verificative or items already familiar for the test participants’, and it is also the main search entry for factual data searching. Secondary search entries for known item or factual data searching are the four screening related entries.

For the explorative topical or content searching, several search entries are applied. The primary search entries for these information needs are ‘Genre’, ‘Keywords’, ‘Participant’, and ‘Author’. Five search entries are considered secondary for topical or
Chapter 7: Research area 1: aspects of information seeking behaviour

content related information needs. These are being able to search for production or responsible department, production country, free text searching in summaries, audience ratings, and searching the spoken words in relation to broadcasts at the features level of granularity. Other granularity level dependent search entries are the placement within a series of broadcasts, and the table of content for features in broadcasts. Both of these are along with ‘Duration’ used as secondary search entries for topical or content searching, though the main use is for relevance assessments, which is the focus of the following Section 7.3. In addition, the screening search entries are applied for narrowing the topical search. The remaining seven bibliographic or content search entries (‘Geographical location of recording’, ‘Copyright information (screening)’, ‘Subtitles’, ‘Spoken language’, ‘Black/white or colours’, ‘Format of transmission’ and ‘Geographical location of content/storyline’), are considered to be of minor importance in relation to searching for television broadcasts. In such, the seven entries are concerned with characteristics of broadcasts, which are useful in relation to specific research projects, only, and not considered of primary relevancy in the broader retrieval context.

The search entries in the screening category of Table 7.9 are the primary search entries in the ideal or methodical selection of data for analysis, and ‘Genre’ functions as a secondary search entry for narrowing the search.

The three search entries in the archival category are useful in relation to obtaining access to particular broadcasts. The search entries are concerned with the time it takes from ordering to reception as well as the playback equipment needed. This category of search entries is related to the statement in Quote 3. This means that the test participants are not willing to wait to long to be able to access the broadcasts they need. Often they simply do not have the time to wait, and the test participants wish to be able to filter out broadcasts, which are not available. In the interviews, the archival search entries are often discussed in relation to teaching tasks, since the scholars often have a tight schedule for preparing their teaching. However, the archival entries are also mentioned for narrowing searches in relation to known item, factual data, and the content or topical information needs.

‘Date of production’ is highly marked by the respondents, but unfortunately, we did not prompt probably for this search entry in the interviews. Hence, we are unable to give statements about the test participants’ use of this search entry. However, in consideration of the knowledge gained and with reference to a traditional textual context ‘Date of production’ is most likely used as a secondary search entry for known item, and factual data searching.
Table 7.10 summarises our answers concerning why and when the search entries are preferred. That is, the table relates the search entries to the four types of information needs identified in Section 7.1.4, and indicates whether the entries are primary, secondary or of minor or no value for the information need. The first column depicts the search entry category, and the second column list each of the 28 search entries identified in a broadcast context. The remaining part of the table depicts each of the four types of information needs identified in Section 7.1, and information about the assessed value of each search entry. The assessment is based on the knowledge learned during our investigations, and our experience within LIS, including knowledge about LIS literature (e.g., IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998). The criteria applied for indication of the relative value is comparable to the criteria used in the functional requirements for bibliographic records (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998, pp.83-84). We distinguish three levels of value assessments. Primary signifies that a search entry is highly valuable for supporting the fulfilment of a particular type of information need. Primary value is assigned when the search entry is considered the typical search entry applied. Secondary signifies that the search entry is assessed to be of moderate value for fulfilling a type of information need. Secondary value is assigned when the search entry is used for qualifying a search under a primary search entry, or the search entry is used for sub-dividing a larger search set. The absence of a marking signifies that the search entry is of no or low importance for fulfilling the type of information need, while ‘–’ signifies that the search entry is considered of no or low value for all four types of information needs. A search entry is considered of low value if it is used for very specific sub-divisions under particular circumstances.

The summarising knowledge provided in Table 7.10 is essential for construction of a future broadcast retrieval system. By assessing the value of each search entry the designer and/or constructor of the future retrieval system is provided with an assessment of the benefits achieved by facilitating each of the 28 search entries. The value is made more explicit in our discussion of the appropriate access points in Chapter 0. Prior to that, we identify and assess the value of relevance criteria applied for broadcasts in the following section.
Table 7.10: Summarisation of the importance of search entries in relation to the four types of information needs.

<table>
<thead>
<tr>
<th>Category</th>
<th>Search entry</th>
<th>Verificative</th>
<th>Factual data</th>
<th>Clear topical</th>
<th>Muddled topical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographic</td>
<td>Title</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Author</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Date of production</td>
<td>S</td>
<td>S</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Geographical location for content/storyline</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Production/responsible department</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Production country</td>
<td>S</td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Copyright information</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>-</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Subtitles</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Spoken language</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Black/white or colours</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Format of transmission</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Screening</td>
<td>Channel</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Date of transmission</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Time of transmission</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Transmission pattern</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Content</td>
<td>Keywords</td>
<td>S</td>
<td></td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Genre</td>
<td>S</td>
<td>S</td>
<td>P, S</td>
<td>P, S</td>
</tr>
<tr>
<td></td>
<td>Geographical location for recording</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Spoken words</td>
<td>-</td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>-</td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Archival</td>
<td>Medium</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Waiting on loan</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Borrow options</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Relational</td>
<td>Placement within a serial</td>
<td>-</td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Table of content</td>
<td>-</td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Audience ratings</td>
<td>-</td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

Legend: P signifies that the search entry is of primary value; S signifies secondary value; and - signifies no or low value for all four types of information needs.
7.3 Research question 3: which relevance criteria are applied when evaluating television broadcasts, and why and when are they applied?

In this section, we report on our results concerning the criteria applied for assessing the relevancy of television broadcasts, as well as when and why the criteria are applied. Similarly, to the previous section, the results for the ‘which’ of the research question is mainly based on the questionnaire responses, whereas the results regarding why and when generally are based on the interview data.

In the fifth part of the questionnaire, the respondents are asked to mark the relevance they wish to apply in a television broadcast context. Similar to the responses of types of broadcasts, level of granularity, and search entries, the questionnaire respondents are free to mark as many relevance criteria as they see fit. The main part (87%) of the respondents has marked between 4 and 18 relevance criteria with a mean of 11.3, and a standard deviation of 5.1 (see Appendix 17 for details). The distribution of the marked relevance criteria is shown in Table 7.11. Similar to Table 7.7, Table 7.11 contains three columns. The first column lists the relevance criteria according the listing in the web questionnaire, including the indicative information in brackets. In the second column, the relevance criterion category is depicted, and in the third column, the overall distribution of the respondents’ answers is provided.

Overall, Table 7.11 shows that more than half of the respondents mark 17 of the 24 relevance criteria. These are five of the ten bibliographic relevance criteria (‘Date of production’, ‘Participant’, ‘Author’, ‘Title’, and ‘Duration’), the three screening criteria (‘Date of transmission’, ‘Time of transmission’, and ‘Transmitting channel’), five of the seven content criteria (‘Keywords’, ‘Genre’, ‘Clip’, ‘Summary’, and ‘Images’), the three archival criteria (‘Borrow options’, ‘Medium’, and ‘Waiting on loan’), and the relational criteria (‘Audience ratings’). The remaining seven criteria (‘Description of image sequences’, ‘Format for transmission’, ‘Geographical location for content/storyline’, ‘Geographical location for recording’, ‘Black/white or colours’, ‘Spoken language’, and ‘With or without subtitles’) are marked to be applied by less than 25% of the respondents.

The 20 search entries in the fourth section of the questionnaire are also listed as relevance criteria (e.g., see the questionnaire in Appendix 4), and a statistical significant association is found between the respondents’ marking of these corresponding data elements (see Appendix 18). In other words, if a data element (e.g., ‘Title’) is marked in the search entry section, the same data element tends to be marked in the relevance criterion section. This could indicate that the respondents do not clearly differentiate
between the two types of data elements. This problem is to some degree resolved with
the methodical use of in-depth interviews, since the interviewer uses the ability to
provide the interviewees with additional information and to prompt for further details.
However, with the benefit of hindsight additional methodical approaches could have
been applied to get around this issue, e.g., by designing and applying screens shots from
a possible future broadcast retrieval system. These issues are further addressed in
Chapter 10.

Table 7.11: Relevance criteria applied by the respondents. The table is arranged similar to the
arrangement of the types of broadcasts in the questionnaire, which is alphabetical according to the Danish
terminology.

<table>
<thead>
<tr>
<th>Relevance criterion</th>
<th>Relevance criterion category</th>
<th>All Respondents (N=108)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of image sequences</td>
<td>Content</td>
<td>16</td>
</tr>
<tr>
<td>Date of production</td>
<td>Bibliographic</td>
<td>49</td>
</tr>
<tr>
<td>Date of transmission</td>
<td>Screening</td>
<td>79</td>
</tr>
<tr>
<td>Keywords</td>
<td>Content</td>
<td>67</td>
</tr>
<tr>
<td>Format for transmission (e.g., 16:9, 4:3)</td>
<td>Bibliographic</td>
<td>11</td>
</tr>
<tr>
<td>Genre</td>
<td>Content</td>
<td>76</td>
</tr>
<tr>
<td>Geographical location for content/storyline</td>
<td>Content</td>
<td>17</td>
</tr>
<tr>
<td>Geographical location for recording</td>
<td>Bibliographic</td>
<td>12</td>
</tr>
<tr>
<td>Time of transmission</td>
<td>Screening</td>
<td>49</td>
</tr>
<tr>
<td>Clip</td>
<td>Content</td>
<td>36</td>
</tr>
<tr>
<td>Summary</td>
<td>Content</td>
<td>87</td>
</tr>
<tr>
<td>Borrow options (e.g., access to broadcasts)</td>
<td>Archival</td>
<td>80</td>
</tr>
<tr>
<td>Medium (e.g., download, DVD, VHS)</td>
<td>Archival</td>
<td>68</td>
</tr>
<tr>
<td>Participant (e.g., interviewee, actor)</td>
<td>Bibliographic</td>
<td>69</td>
</tr>
<tr>
<td>Author (e.g., programme planner, script writer)</td>
<td>Bibliographic</td>
<td>58</td>
</tr>
<tr>
<td>Audience ratings</td>
<td>Relational</td>
<td>43</td>
</tr>
<tr>
<td>Black/white or colours</td>
<td>Bibliographic</td>
<td>14</td>
</tr>
<tr>
<td>Spoken language</td>
<td>Bibliographic</td>
<td>20</td>
</tr>
<tr>
<td>With or without subtitles</td>
<td>Bibliographic</td>
<td>25</td>
</tr>
<tr>
<td>Images (thumbnails)</td>
<td>Content</td>
<td>32</td>
</tr>
<tr>
<td>Title</td>
<td>Bibliographic</td>
<td>87</td>
</tr>
<tr>
<td>Transmitting channel (e.g., DR1, TV2, TV2 Zulu)</td>
<td>Screening</td>
<td>84</td>
</tr>
<tr>
<td>Duration</td>
<td>Bibliographic</td>
<td>72</td>
</tr>
<tr>
<td>Waiting on loan</td>
<td>Archival</td>
<td>69</td>
</tr>
</tbody>
</table>

209
Similar to our reporting on the search entries applied by the test participants in the preceding section, we report on the use of relevance criteria in separate sub-sections concerned with the five categories introduced in Section 7.2. A sixth sub-section focuses briefly on the difference between responses from the different departments and different academic status. Lastly, Section 7.3.7 summarises our answers to the third research question.

7.3.1 Bibliographic relevance criteria

The respondents frequently mark all the three bibliographic relevance criteria concerned with title or statement of responsibility. That is, ‘Title’ is marked by 80.6%, ‘Participant’ by 63.9%, and ‘Author’ by 53.7%. In verificative and factual data searching the title of the broadcasts is the main search entry as well as the main relevance criteria, indicated by the fact that 91 and 87 respondents mark it, respectively. However, since titles often are short and indistinct, the test participants’ use several other criteria for judging the relevancy in verificative or factual data retrieval. Further, the indistinctiveness of the titles provides little information for topical or content-related information needs. Nonetheless, due to the familiarity with the primary sources, the interviewees emphasise that the title is important because it often makes them recall the broadcast in question, which prompts the recovery of knowledge about the broadcast from the back of the test participants mind.

‘Participant’ and ‘Author’ are used for topical information needs. For instance, in interviewee C’s work task the participant is the main search entry and the main relevance criteria, and in interviewee F’s work task known authors are used for broadening the topical searches, and the authors and participants are used for assessing the relevancy of the retrieved broadcasts. The two relevance criteria are marked by approximately 10 percent point less than the corresponding search entries, which indicates that though significant for relevance assessments, the data is considered more important in searching. Similar to the responses in the free text field in the search entry section of the questionnaire, the respondents break down ‘Author’ into detailed roles (e.g., director, cutter, or cameraman). This need for specific indication of the role of authors as well as participants is also expressed in the interviews, but as with the search entries, it is outside the scope of the present work to comprise the categorisation or list of the different roles of authors and participants.

The two production related relevance criteria, ‘Date of production’ and ‘Geographical location of recording’, are roughly marked by the same percentages as the corresponding search entries. In the interviews, we did not prompt properly for information about the ‘Date of production’, and we are hence not able to provide
answers as to why and when the relevance criterion is applied. However, intuition and knowledge about its textual counterpart, namely publication year, gives indication about the use of the relevance criterion in judging the recency of the broadcasts. In a television context, other data elements are also used for judging the recency of broadcasts. These are discussed in relation to the screening relevance criteria in the following Section 7.2.2. Similar to our reporting on bibliographic search entries, the interviews bring about production related relevance criteria not listed in the questionnaire. These are ‘Production/responsible department’, ‘Production country’ and ‘Copyright information (screening)’. The two former are mainly used for judging the perspectives in the content of the broadcasts in question, and the relevance criteria are used in order to identify and differentiate the broadcasts. This is due to the topical information these data elements hold for the user familiar with the domain. An example would be three programmes entitled “Princess Diana”, and produced by the news, documentary, and drama departments, respectively. This is interconnected to the use of the content related relevance criteria, which is discussed in Section 7.3.3. Another use of these production related criteria is for judging the nationality of the broadcasts, which is relevant for many investigations in Media Studies, e.g., a comparative study of Danish and Australian format versions, as expressed by interviewee A.

The marking of the relevance criteria related to the physical description of the television broadcasts, are somewhat scattered, but generally comparable to the distribution found for the corresponding search entries. ‘Duration’ is the only physical description criteria marked by more than 25% of the respondents, and it is one of the only data elements marked more often as a relevance criterion than as a search entry, namely 14.8 percent points more. This shows that the duration or length of the broadcasts is extremely important when judging whether a broadcast should be obtained or not. ‘Subtitles’ and ‘Spoken language’ are used in combination with other relevance criteria for judging the nationality or country of origin of broadcasts, as mentioned above. However, the two criteria are only considered of minor importance for this task. The remaining two physical description relevance criteria (‘Black/white or colours’ and ‘Format of transmission’) are not mentioned by any of the nine interviewees and we are hence unable to provide reasonable explanation of their use.

7.3.2 Screening relevance criteria

Three relevance criteria in Table 7.12 are concerned with screening related aspects of television broadcasts. These are ‘Channel’, ‘Date of transmission’, and ‘Time of transmission’. ‘Channel’ is marked by 77.8%, and it is hence considered important by a majority of the respondent. Information about the transmitting channel is important in
order to identify the right broadcasts in a verificative or factual data need, and ‘Channel’ serves as a secondary relevance criterion for these two types of information needs, as well as for comparative topical investigations, e.g., comparison between the coverage of the Iraq war on DR1 and TV2. Date and time of broadcasting are used in a similar fashion. In addition, they function as the main criteria in work task where the user aims at an ‘ideal’ data collection. Further, date of transmission is used in assessing the recency of broadcasts. At first glance the date of transmission is somewhat comparable to the data of printing in a textual context, since both dates are not concerned with creation of the intellectual content, but merely the ‘release’ of the content. However, the date of transmission is very important for the test participants since this is the specific date the broadcast is issued in society.

Similar to our reporting on the preferred search entries in the previous section, the transmission pattern is important in order to judge the relevancy of broadcasts. The user needs to know whether the broadcast is a first time transmission or a re-run. This information is important in relation to verificative, factual data, or topical related information needs. In the same way, the serial title, placement within serial, as well as table of feature content in a programme, are mentioned as relevance criteria by the interviewees. These are discussed in Section 7.3.5 about relational relevance criteria.

In all nine interviews, the programming is considered an extremely important relevance criterion. That is, the interviewees mention horizontal as well as vertical programming. The purpose of horizontal and vertical programming in the exploratory generation of overviews of the primary sources is discussed in our reporting of the interviewees’ work tasks in Section 7.1.1. In the overview generating tasks, the information about the surrounding broadcasts is important in order to select relevant television broadcasts. Horizontal and vertical programming is similar to the presentation of television broadcasts in the television schedules in newspapers. An electronic example of horizontal and vertical programming is found at TRILT’s homepage (see e.g., an example in Appendix 19). The placement in the program schedules is very important when assessing the relevancy of retrieved broadcasts for the nine interviewees.

In brief, the relevance criteria in the category of screening are considered important by the test participants, including whether we are dealing with a first time transmission or a re-run (‘Transmission pattern’).

7.3.3 Content relevance criteria

Seven relevance criteria are categorised as content related. These are ‘Summary’, ‘Genre’, ‘Keywords’, ‘Clips’, ‘Images’, ‘Geographical location of content/storyline’,
and ‘Description of images sequences’. Summaries or description of the content of the broadcasts in question are used for gaining knowledge about the content of broadcasts. These summaries are especially important in the overview generating work tasks. They serve as relevance criteria in relation to explorative topical or content related information needs, and in some cases, they even provide the user with the needed information, and the broadcast itself is consequently not relevant to access. In contrast to the summaries at the programme level, only 14.8% of the respondents consider descriptions at the level of sequences important for relevance judgments. The relatively few respondents who have marked ‘Description of image sequences’ is in line with the few respondents (12.0%) who have marked broadcasts at the granularity level of sequences to be of importance. However, despite the intuitive connection between the two responses, we are not able to identify a statistical significant relation between the respondent marking the granularity level of sequences and marking the relevance criteria of ‘Description of images sequences’ (see Appendix 20).

Indication of the genre is used in order to gain insight as to the main perspectives of the broadcasts, and genre is considered important for searching as well as for judging the relevancy of broadcasts, which is expressed in the following quote from interviewee C:

“After retrieving broadcasts by their title it is important for me to be informed whether the broadcast is a news broadcast or a documentary, or something else. Further, I would like to be able to search on it as well. It will make the searching easier to be able to search all news broadcasts for that company or that person, and then move on to other categories afterwards”.

Quote 41

An example would be “Princess Diana” – news, documentary or motion picture. In this way, genre functions as a relevance criterion for all four types of information needs.

‘Keywords’ is considered less important as a relevance criterion than as a search entry. To be exact, ‘Keywords’ is marked by 20.1 percent points less as a relevance criterion than as a search entry. This indicates that the main purpose of keywords is for searching and not for relevance judgments. Keywords are considered to be of secondary importance for relevance assessments. This is in line with the fact that ‘Summary’ is the main topical relevance criterion.

‘Clips’ and ‘Images’ are both marked by approximately 1/3 of the respondents, only, which indicates that such visual content related criteria are of minor importance for judging the relevancy of broadcasts. In Quote 25, interviewee H expresses that such visual data is important for gathering information about the content of the broadcasts,
and in this way, these visual criteria are used for assessing broadcasts in explorative topical information needs.

7.3.4 Archival relevance criteria

The three archival relevance criteria, ‘Borrow option’, ‘Waiting on loan’, and ‘Medium’ are frequently marked by the respondents, and the interviewees emphasise them to be important in order to see which broadcasts are available, and how and when the broadcasts may be accessible. These relevance criteria are considered significant independently of the type of information need, and are hence applied for verificative, factual data, clear as well as muddled topical information needs. The need for this kind of information is logical in consideration of the need to be informed about all transmitted broadcasts as well as the restrictions on access, including the many legislative changes taking place over the years. These many changes have entailed that the test participants have difficulties in being acquainted with the precise rules for accessing the broadcasts. This again, makes it difficult for the scholars to inform and help the students in obtaining access to broadcasts, and it is necessary for the test participants to get this archival related information at retrieval time. Providing such information in the IR system, requires extensive access right management, as well as unequivocal identification of each user at the time of retrieval.

7.3.5 Relational relevance criteria

‘Audience ratings’ is considered important in order to identify the impact a particular broadcast have had in society. Since a major aspect in Media Studies is the relation between the media and society, audience ratings play an important role in the selection of relevant television broadcasts. This corresponds to our discussion of the use of audience ratings as a search entry. The relevance criterion is used in relation to clear or muddled topical information needs, only.

Further relational relevance criteria, are ‘Placement within a serial’ and ‘Table of content’. Both criteria provide the user with contextual information in relation to the broadcast in question, though they are concerned with information at different levels of granularity. The present work have had a primary focus on relevance criteria at the programmes level of granularity, and future work will need to give more attention to the relevance criteria applied at the serial as well as the features level of granularity. We are naturally aware that several relevance criteria are applied at different levels of granularity, including the title of serials and features, and summaries at the features
level, but since it is outside our focus, we are not able to provide proper statements in this respect.

7.3.6 Comparison between relevance criteria and affiliation or academic status

Table 7.12 depicts the respondents’ expression concerning application of relevance criteria, according to the two types of affiliations (Aarhus and Copenhagen), and the two types of academic status (scholar and student). As follows, the table provides the figure for comparison between application of relevance criteria and affiliation or academic status. The table is constructed like Table 7.8 for preferred search entries. The relevance criteria are listed in the first column according to the overall frequency, which is depicted in the fourth column. The second and third column shows the distribution according to affiliation and academic status, respectively. In the second and third column we have tested for an association between each relevance criterion and either affiliation or academic status. In the light grey cells, we applied the chi² test of association, and in the dark grey cells, we applied Fisher’s exact test of association. For the four statistical significant associations the figures are shown in bold typeface.

The comparison of relevance criteria marked by respondents from the two departments shows, that generally, the academic affiliation is not related to the relevance criteria marked. However, for ‘Time of transmission’ and ‘Audience ratings’ we are able to identify a significant difference between the two groups of respondents. In both cases, the respondents affiliated to the Copenhagen department are more reluctant to consider the relevance criteria to be of importance. Both criteria are also listed as search entries (see Table 7.8), and here they get approximate the same attention from the Aarhus department, while the marks from the Copenhagen department drops with six and ten respondents, respectively. In comparison to the respondents from Aarhus, the respondents from Copenhagen consider ‘Time of transmission’ and ‘Audience ratings’ to be search entries rather than relevance criteria.

Scholars and students predominantly agree about the relevance criteria they consider important. For two criteria, ‘Waiting on loan’ and ‘Time of transmission’ we are able to identify a statistical significant difference. ‘Waiting on loan’ is considered considerably more important by students than by scholars. In contrast, scholars mark ‘Time of transmission’ significantly more often than do students.

In the following section, we summarise and discuss the results obtained regarding the third research question, concerned with the test participants’ application of relevance criteria.
Table 7.12: Comparison of relevance criteria applied by respondents according to affiliation, academic status, and the pool of all respondents. The relevance criteria are listed in accordance to the overall marking by all 108 respondents.

<table>
<thead>
<tr>
<th>Relevance criterion</th>
<th>Affiliation</th>
<th>Status</th>
<th>All Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aarhus (N=52)</td>
<td>Copenhagen (N=56)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n     %</td>
<td>n     %</td>
<td>n     %</td>
</tr>
<tr>
<td>Summary</td>
<td>44</td>
<td>84.6</td>
<td>43</td>
</tr>
<tr>
<td>Title</td>
<td>43</td>
<td>82.7</td>
<td>44</td>
</tr>
<tr>
<td>Transmitting channel (e.g., DR1, TV2, TV2 Zulu)</td>
<td>44</td>
<td>84.6</td>
<td>40</td>
</tr>
<tr>
<td>Borrow options (e.g., access to broadcasts)</td>
<td>39</td>
<td>75.0</td>
<td>41</td>
</tr>
<tr>
<td>Date of transmission</td>
<td>40</td>
<td>76.9</td>
<td>39</td>
</tr>
<tr>
<td>Genre</td>
<td>39</td>
<td>75.0</td>
<td>37</td>
</tr>
<tr>
<td>Duration</td>
<td>39</td>
<td>75.0</td>
<td>33</td>
</tr>
<tr>
<td>Participant (e.g., interviewee, actor)</td>
<td>30</td>
<td>57.7</td>
<td>39</td>
</tr>
<tr>
<td>Waiting on loan</td>
<td>37</td>
<td>71.2</td>
<td>32</td>
</tr>
<tr>
<td>Medium (e.g., download, DVD, VHS)</td>
<td>34</td>
<td>65.4</td>
<td>34</td>
</tr>
<tr>
<td>Keywords</td>
<td>30</td>
<td>57.7</td>
<td>37</td>
</tr>
<tr>
<td>Author (e.g., programme planner, script writer)</td>
<td>25</td>
<td>48.1</td>
<td>33</td>
</tr>
<tr>
<td>Date of production</td>
<td>23</td>
<td>44.2</td>
<td>26</td>
</tr>
<tr>
<td>Time of transmission</td>
<td>30</td>
<td>57.7</td>
<td>19</td>
</tr>
<tr>
<td>Audience ratings</td>
<td>26</td>
<td>50.0</td>
<td>17</td>
</tr>
<tr>
<td>Clip</td>
<td>18</td>
<td>34.6</td>
<td>18</td>
</tr>
<tr>
<td>Images (thumbnails)</td>
<td>15</td>
<td>28.8</td>
<td>17</td>
</tr>
<tr>
<td>With or without subtitles</td>
<td>11</td>
<td>21.2</td>
<td>14</td>
</tr>
<tr>
<td>Spoken language</td>
<td>11</td>
<td>21.2</td>
<td>9</td>
</tr>
<tr>
<td>Geographical location for content/storyline</td>
<td>8</td>
<td>15.4</td>
<td>9</td>
</tr>
<tr>
<td>Description of image sequences</td>
<td>9</td>
<td>17.3</td>
<td>7</td>
</tr>
<tr>
<td>Black/white or colours</td>
<td>5</td>
<td>9.6</td>
<td>9</td>
</tr>
<tr>
<td>Geographical location for recording</td>
<td>5</td>
<td>9.6</td>
<td>7</td>
</tr>
<tr>
<td>Format for transmission (e.g., 16:9, 4:3)</td>
<td>6</td>
<td>11.5</td>
<td>5</td>
</tr>
</tbody>
</table>

Legend: Association is tested with Chi² (light grey cells) and with Fisher’s Exact test (dark grey cells).

**Bold and italic:** significantly different at \( \alpha = 0.05 \).

### 7.3.7 Discussion and summarising statements

Table 7.13 outlines the relevance criteria applied in relation to television broadcasts. In this way, Table 7.13 provides an answer to the first part of the third research question, concerned with which relevance criteria are applied for evaluating the relevancy of television broadcasts. Similar to our recapitulation of the search entries in Section
As to the when and why part of research question 3 it is concluded that for verificative and factual data information needs, ‘Title’ is the general primary relevance criterion. In information needs of a topical or content related nature, ‘Summary’ is generally used as
the primary relevance criterion. In addition, several relevance criteria are used to supplement these primary criteria. ‘Duration’, the five screening criteria, ‘Genre’, ‘Clips’, ‘Images’, and the three archival criteria are used as secondary relevance criteria in relation to all four types of information needs (known item, factual data, known topic or content, and muddled topic or content). ‘Participant’, ‘Author’, ‘Production/responsible department’, ‘Production country’, ‘Keywords’, and the three relational criteria are all used as secondary relevance criteria in relation to clear as well as muddled topical information needs. The respondents frequently mark ‘Date of production’, but as with its corresponding search entry, we did not prompt properly for this relevance criterion in the interviews. Hence, we are unable to give statements about the test participants’ use of this search entry. However, similar to our conclusion regarding ‘Date of production’ as a search entry, the relevance criterion ‘Date of production’ is most likely used as a secondary search entry for known item, and factual data searching. ‘Spoken words’ is particularly mentioned in relation to topical information needs for broadcasts at the granularity level of features. That is, the transcriptions of the spoken words are used for estimating the relevancy of features, similarly to the use of summaries at the programmes level of granularity. The remaining relevance criterions: ‘Geographical location of recording’, ‘Copyright information’, ‘Subtitles’, ‘Spoken language’, ‘Black/white or colours’, ‘Format of transmission’, ‘Geographical location of content/storyline’, and ‘Description of images sequences’ are considered to be of minor importance for assessing the relevancy of broadcasts.

Table 7.14 summarises our answer concerning why and when the relevance criteria are applied. That is, the table relates the criteria to the four types of information needs identified in Section 7.1.4, and indicates whether the criteria is of primary or secondary importance for fulfilling the information need. The assessment of the value of relevance criteria in relation to types of information needs is conducted similar to the assessment of the value of search entries, and the criteria for the assessment are described in Section 7.1.4. Table 7.14 is constructed like Table 7.10 in Section 7.1.4.

An important finding in the present work is the use of programming as a relevance criterion. Though we characterise it as a secondary criterion according to the criteria for assessment described in Section 7.1.4, the contextual information provided in the ‘Programming’ is important for the test participants, and it is essential that the feature of programming is made available in a future broadcast retrieval system.
Table 7.14: Summarisation of the importance of relevance criteria in relation to the four types of information needs.

<table>
<thead>
<tr>
<th>Relevance criterion</th>
<th>Verificative</th>
<th>Factual data</th>
<th>Clear topical</th>
<th>Muddled topical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date of production</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographical location for content/storyline</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production/responsible department</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production country</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copyright information</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Subtitles</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Spoken language</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Black/white or colours</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Format of transmission</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Channel</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Date of transmission</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Time of transmission</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Transmission pattern</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Programming</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Summary</td>
<td></td>
<td></td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Genre</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Keywords</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Clips</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Images</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Geographical location for recording</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Description of image sequences</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Spoken words</td>
<td></td>
<td></td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Medium</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Waiting on loan</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Borrow options</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Placement within a serial</td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Table of content</td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Audience ratings</td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

Legend: P signifies that the search entry is of primary value; S signifies secondary value; and - signifies no or low value for all four types of information needs.
With reference to Sandom and Enser’s (2002) results on the relevance criteria applied at 11 operating moving image archives (see Section 3.5.2.1) our distribution of relevance criteria depicted in Table 7.12 is generally comparable. However, a few differences are identified; differences, which we believe, are related to the different work tasks of the users in question. The users in Sandom and Enser’s study are professionals who need broadcasts in order to produce new broadcasts (e.g., news broadcasts or documentaries). In Sandom and Enser’s study, the users typical need access to broadcasts at the granularity level of features, and they have a strong focus on relevance criteria, which we have categorised within the ‘Physical description’ sub-category of ‘Bibliographic’ relevance criteria. In brief, the contextual and situational differences between the users in Sandom and Enser’s study and the users in the present study is most certainly a determinant factor in the different emphasis on particular relevance criteria. This is further verified by Yang (2005), who finds that academic work tasks and information needs differs from moving images editors and producers work task and information needs. For instance, moving images editors and producers have a primary need for broadcasts at the features and sequence level of granularity, whereas academics generally need broadcasts at the programmes level of granularity. ‘Clips’ and ‘Images’ are particular for audiovisual information objects, and our study verifies the results regarding what Yang denotes criteria that relates to visual elements. That is, similar to Sandom and Enser, and Yang, our results show that the visual criteria ‘Clips’ and ‘Images’ are considered to be secondary for relevance assessments.

Here it is important to emphasise that often the test participants need information about broadcasts, and that it often is important for the test participants to identify broadcasts which stand out in some way or another, and hence it is sometimes difficult for the interviewees to identify the needed relevance criteria upfront, e.g., broadcasts which are representative for a type of broadcast in a particular period (historical work tasks) or borderline exemplars in relation to a genre. Further, since many information needs are concerned with gathering information about broadcasts, information needs can be solved without obtaining access to any broadcast at all. Hence, the information available in the broadcast retrieval system, not the broadcasts as such, is assessed for relevancy in fulfilling the need.

Previous studies on relevance criteria in different contexts and with different types of information objects (e.g., Schamber, 1991b; Park, 1992; Choi & Rasmussen, 2002; Yang, 2005), emphasise the importance of topical related relevance criteria, which is verified in the present work. In an empirical study Conniss, Ashford and Graham (2000) found that ‘recency’ is an important relevance criterion for information professionals retrieving still images. To be exact they found that the most recent
information objects were requested (e.g., the latest mug shot of a criminal). The users in the study by Conniss and colleagues had somewhat different work tasks than the users in our study (e.g., police officers seeking information in order to solve crimes). We do not believe the results from our study should be interpreted similarly to the results obtained by Conniss, Ashford and Graham. The fact that the users to a large degree have to rely on their memory when requesting programmes due to the lack of electronic search entries of the present collection probably does affect the television programmes users request, e.g., as emphasised by interviewee E in Quote 37. However, in our investigation we find that the interviewees have an emphasis on recent as well as historical broadcasts, which to some extent is in opposition to the genuine request received at the Danish national collection of television broadcasts as reported on in Kirkegaard and Borlund (2006).

This concludes our reporting on the relevance criteria applied for evaluating television broadcasts. In the following section, we discuss and summarise the results obtained in relation to the first research area, and hereby close the present chapter concerned with three aspects of the test participants’ behaviour when seeking television broadcasts, namely information needs, search entry preferences and application of relevance criteria.

7.4 Summary and discussion of aspects of seeking behaviour: research area 1

Research area 1 is concerned with aspects of users’ seeking behaviour in the context of television broadcasts, which we have investigated with a focus on scholars and students in the field of Media Studies.

In our reporting, we pay attention to investigation of differences between the responses from respondents with different departmental affiliation or academic status. Generally, the groups do not demonstrate different information seeking behaviour with respect to characteristics of their information needs, search entry preferences, and application of relevance criteria. The few differences identified are considered to be of minor importance in respect to design and construction of a future broadcast retrieval system. In other words, the behaviour exercised by the test participants in relation to information needs, search entry preference and relevance criteria application, does not indicate that major information behavioural differences exist, which would affect the establishment of a broadcast retrieval system. Essentially, we are unable to identify any difference that argues in favour for specific broadcast retrieval system for any of the groups of test participants, and we conclude that one general system should be able to
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

meet the needs of all the test participants. With reference to the results of previous investigations of students and scholars’ information seeking behaviour which we reviewed in Section 3.3.2, we conclude that the difference in seeking behaviour is not related to a difference in the search entries preferred and relevance criteria applied.

The focus on a historical development, which is found in four of the nine interviews, is related to the fact that the primary sources have only recently become available for the researchers, and hence the field does not have a general conception of the historical development of the primary sources. This is for instance reflected in the statement about the state in other Scandinavian countries in Quote 28. In this statement interviewee H argues that research on the genre concept in a Danish television context, have reached a point where the genre is applicable for indexing the broadcasts. With reference to a future broadcast retrieval system, it could be claimed that the generation of an overview of the primary sources is a one-time event for the field. That is, once the historical overview of transmitted broadcasts is (re)created subsequent scholars and students can learn about this overview by reading the literature e.g., Hjarvard (2006b) or searching a database e.g., the Registrant (Hjarvard & Jespersen, 2006). However, the media is constantly being developed and the primary sources, historical as well as contemporary, can continually be approached from new perspectives. An example is the need to define and describe a new genre as expressed in Quote 15. In addition, generation of the historical overview allows researchers to dig deeper into the characteristics of the primary sources, e.g., as expressed in Quote 29.

Our results verifies the five tasks for users of broadcast retrieval systems which are identified by Auffret and Bachimont (1999, p. 62). These five tasks are to be able to: 1) search for broadcast, 2) browse broadcasts, 3) navigate broadcasts, 4) interpret broadcasts in context, and 5) annotate broadcasts. The task of searching for broadcasts is evident and underlying all work tasks, we have paid attention to. The task of browsing is for instance expressed by interviewee H in Quote 25. Identification of the specific functions to apply are outside the present work, but it is closely related to the task of navigation at different levels of granularity, e.g., by providing at table of content for features in a programme, or establishing links between the programmes of a serial. Again, the test participants’ departure in the transmission dimension of the broadcasts signifies that the respondents need to be able to browse and navigate, not necessarily the broadcasts themselves, but also the index of surrogate records for all transmitted broadcasts. The task of interpretation of broadcasts in context is clearly identified in our results, and e.g., exemplified with the need to see the horizontal as well as the vertical programming, as well as audience ratings. The context plays an essential role in the test participants’ television needs. The fifth task of annotation of broadcasts is not
the main focus of our work, but it is apparent that support should be provided for the users. The respondents have a need to work on/manipulate/process the broadcasts, as well to further process the metadata or surrogate records. A simple way is to allow the users to download digital copies of the broadcasts and thereby enabling them to manipulate the primary source, but features that are more sophisticated should be considered. For instance features which allows annotation in connection to the broadcast retrieval system. The relevant tasks are naturally related to the type of information need and e.g., explorative overview generating needs would most certainly be closely related to needs for navigation and interpretation in context.

Another aspect of the test participants’ information needs in a television broadcasts context is the distinguishing between two kinds of needs:

1) The need to gain knowledge about broadcasts; and
2) The need to gain access to specific broadcasts

Most of the needs discussed with the test participants are concerned with gathering knowledge about broadcasts, e.g., in relation to generating an overview of the primary sources. The test participants are concerned with the broadcasts as primary sources for their scientific analysis, and if the information they need about the broadcasts is available in external sources they do not need to conduct the cumbersome task of extracting the information from the primary sources themselves. An example is interviewee A’s task of identifying all programme formats in Danish television. If information about copyright is available in external sources, the interviewee does not need to view any of the insurmountable amounts of broadcasts. The need to gain access to specific broadcasts appears after generating the overview, and hence the need has often developed into a verificative need.

In wrapping up the first research area regarding the characteristics of a given group of users’ information needs for television broadcasts, the following quote adds a further aspect.

“This is the way it has been with TV-Meter. In the beginning, we just searched for the number of viewers. Then you learn that the system opens up for a lot of possibilities, and that adds to the research in ways, which are impossible to identify beforehand. Suddenly, you discover that you can do certain things and it adds to your analysis in a way that you hadn’t considered possible”.

Quote 42
That is, the aspect that new retrieval possibilities add new perspectives to the research and hence the work tasks to be conducted. Such work tasks are difficult if not impossible to identify, up front, mainly because scholars and students largely choose their own focus for the work tasks.

This concludes our reporting on characteristics of the test participants’ information needs, search entry preferences, and application of relevance criteria in the context of television broadcasts. In the following chapter, we pay attention to the access points that are appropriate for handling the reported behaviour, and hence we answer research question 4.
Chapter 8: Research area 2: access point for television broadcasts

In this chapter, we answer research question 4, which corresponds to research area 2 that concerns the identification of appropriate access points for indexing and retrieval of television broadcasts. To be exact, research question 4 states:

Which access points in a surrogate record are appropriate for television broadcasts, and what are their functions?

The answering of the research question is based on the previously verified search entries and relevance criteria. The structure of the chapter follows the structure established in Chapter 7. That is, Section 8.1 comprises our answers to the fourth research question, including a summary and discussion of the results. The purpose of the present chapter is to relate the results obtained in Chapter 7 to an IR system’s design context. We do this by translating the verified search entries and relevance criteria into appropriate access points in surrogate records. In this translation, we include the knowledge learned about the characteristics of the test participants’ information needs. Hereby, the chapter relates the results obtained in answering the research questions 1, 2, and 3, to system aspects, and the chapter comprises knowledge about the access points that needs to be constructed in order to facilitate the users’ seeking behaviour in a television broadcast context. Further, we consider the type of data conveyed in the access points, and we concentrate on the function of each access point in relation to retrieval of broadcasts. The function of the identified access points is described by relating to the four types of information needs identified in Chapter 7, and the four generic user tasks defined in the study of functional requirements for bibliographic records (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998, pp. 8-9). These four generic user tasks are:

- **Find**: using the access point to find broadcasts that corresponds to the user’s search;
- **Identify**: using the access point to identify a television broadcast, e.g., to confirm that the broadcast corresponds to the broadcast sought;
- **Select**: using the access point to select a broadcast that is appropriate to the user’s need at hand; and
- **Obtain**: using the access point to acquire or obtain access to broadcasts.
In Chapter 9, we take the knowledge learned in the present chapter a step further by focusing on construction of the access points, including the indexing that needs to be conducted in order to facilitate the identified behaviour. However, when relevant for our line of reasoning, the present chapter briefly touches on indexing issues at a general level.

8.1 Research question 4: which access points in a surrogate record are appropriate for television broadcasts, and what are their functions?

The aim of this section is to provide answers to the fourth research question. The search entries and the relevance criteria listed in Table 7.9 and Table 7.13, form the basis for the section. The access points are related to access points in a traditional textual context, and we discuss and conclude on the function and significance of the access points in a television broadcast context. The ascertaining of the significance of the access points is based on the results on aspects of the test participants’ information seeking behaviour in a television broadcasts context as reported on in Chapter 7. The marking of most used search entries and/or relevance criteria by more than 25% of the respondents provide a crude thumb rule for assessing it as a significant access point. 25% is naturally an arbitrary threshold, and therefore we also give emphasis to the viewpoints of the interviewees, as well as the nature and characteristics of the test participants’ work tasks and information needs in assessing whether a specific access point is considered significant or not. In addition, we include the values assigned to the identified search entries in Table 7.10 and to the identified relevance criteria in Table 7.14.

In the following five sub-sections, we discuss the appropriate access points in relation to each of the five categories listed in Table 7.9 and Table 7.13: bibliographic access points, screening access points, content access points, archival access points, and relational access points. The structure follows the structure introduced in Chapter 7. Since our categorisation to some extent is inspired by categorisation of metadata in a traditional textual context (e.g., the areas in an ISBD description or fields in a MARC record), this structure is beneficial for comparing the television broadcast access points to access points in a textual context. Further, it is useful for future work regarding construction of a broadcast retrieval system, including construction of relational links between information objects in the bibliographic universe. Finally, in the 6th sub-section, we summarise and discuss our results concerning appropriate access points for television broadcasts, and we explicitly provide answers to research question 4.
8.1.1 Bibliographic access points

This section comprises 18 access points categorised to be bibliographic in nature. The bibliographic access points are divided into three sub-categories in line with the categorisation of search entries in Table 7.9, and relevance criteria in Table 7.13. Hence, bibliographic access points are considered according to the categories: 1) Title and Statement of responsibility (three access points); 2) Production (five access points); and 3) Physical description (five access points), and treated separately in the following three sub-sections.

8.1.1.1 Title and Statement of responsibility

The ‘Title and Statement of responsibility’ sub-category of bibliographic access points comprises three access points: 1) ‘Title’; 2) ‘Participant’; and 3) ‘Author’. The ‘Title’ of a broadcast is considered important for searching as well as for relevance assessments. It is marked by 84.3% and 80.6% of the respondents, and it is mentioned by the interviewees to be very important. Therefore, we consider the ‘Title’ a significant access point in relation to television broadcasts. The ‘Title’ is comparable to the title in textual contexts, however, it is not an unambiguous access point in a moving image context as pointed out by Hertzum (2003, p. 180). Here the point of departure is broadcasts at the granularity level of programmes, and the ‘Title’ consequently refers to broadcasts at this level of granularity, only. However, the title is important in relation to serials and features as well, which is discussed in Section 8.1.5 on relational access points. The ‘Title’ access point contains so-called descriptive data. The data is extracted from the information object itself, and reproduced in the form of appearance in the broadcast. The title of a broadcast is mainly used for searching and relevance assessments in verificative or factual information needs, and information about the title is important in order to identify and represent the broadcast. This is similar to the function of titles in a textual context. As expressed by interviewee I in Quote 32 and Quote 33, titles are often short and identified titles might be used for several different programmes. Therefore, the user needs further information in order to differentiate between several programmes with the same title, e.g., an indication of genre as expressed by the interviewee. This is basically not different from a textual context where two different books might have the same title, however the indistinctiveness of titles of television programmes adds to this need for additional information. Further, the same programme might be broadcasted several times and it can very well be a specific transmission the user needs e.g., the first transmission. This distinction is for instance important when the so-called reception dimension is connected to the transmission dimension. If a user is interested in knowing about the audience ratings, it is important
to be able to specify the specific transmission(s) of a programme. In this way, the ‘Title’ is related to the two access points of ‘Transmission pattern’ and ‘Audience ratings’, and the issue is further discussed in Section 8.1.2 on screening access points and Section 8.1.5 on relational access points. In brief, the ‘Title’ is an essential access point for television broadcasts, and its main function is to find and identify broadcasts in verificative and factual information needs, though it does not identify broadcasts unambiguously.

‘Participant’ is considered an important search entry by 73.1% of the respondents, and a key relevance criterion by 63.9% of the respondents. In addition, the interviewees mention participants as search entries as well as relevance criteria. The interviewees point out that the roles of the participants are important, e.g., by specifying that the participant is an interviewee or a debater. Hence, the access point should be further specified into the roles of the participants. The exact specification of the roles of the ‘Participant’ access point is beyond the scope of the present work, but a controlled list of roles should be considered. In this way, the users will be able to conduct detailed and structured searches.

The ‘Author’ is considered important by approximately 10 percentage points less than ‘Participant’, both as a search entry and as a relevance criterion. In the free text fields of the search entry section as well as the relevance criteria section, questionnaire respondents express the need for a further specification of the roles of the author. This specification is further emphasised in the nine interviews. Examples of roles in such a specification are director, actor, cameraman, commentator, programme planner and anchorperson. As with the ‘Participant’ access point, the exact specification of the ‘Author’ is beyond the scope of the present work, but from the present work it seems that users consider significantly more roles for ‘Author’ than for ‘Participant’. Therefore, it is even more important to emphasise the value of constructing and applying a controlled list of roles of authors for indexing and retrieval.

‘Participant’ and ‘Author’ are mainly used as the search entry and relevance criteria in relation to topical information needs, and their main functions are to disambiguate broadcasts e.g., broadcasts with similar titles, and to collocate broadcasts with or by the same person. The participants and the authors may appear with different names in the broadcasts, and normative data is therefore needed in order to relate and collate the broadcasts by or with the same person. In addition, it cannot be expected that the users know the different names or spellings of the participant or author, and therefore a descriptive form of the names should also be available for the retrieval of broadcasts to be most efficient. Therefore, authority control in the form of a controlled
list of roles as well as an authority record for the personal names is vital for effective retrieval.

The sub-category of ‘Title and Statement of responsibility’ is comparable to categorisation of access points in a traditional textual context. It resembles area 1 in ISBD records and field 245 in a MARC record. This said, the different roles of the responsible persons (authors as well as participants) differ from the roles in a traditional context (e.g., illustrator). Nevertheless, the differences are few, and future IR system designers and constructors should be able to incorporate these differences in the indexing and cataloguing practices.

8.1.1.2 Production

In this section we discuss the five access points in the bibliographic sub-category of production, namely: ‘Date of production’, ‘Geographical location of recording’, ‘Production/responsible department’, ‘Production country’, and ‘Copyright information (screening)’.

As mentioned in Section 7.2.1, ‘Data of production’ is comparable to publication year in a textual context. However, the time of production or publication is not as important in a television context as it generally is in a textual context. For television broadcasts the date and time of transmission is considered to be of more importance than the actual production date. This is naturally related to the fact that the general basis for academics is the transmission dimension of the broadcasts, including the focus on all transmitted broadcasts. With an indication from approximately 50% of the respondents, we consider ‘Date of production’ a significant access point in a future broadcast retrieval system. The function of the access point is to assist in identifying and selecting broadcasts.

The ‘Production/responsible department’ access point is not listed in the questionnaire, neither as a search entry nor as a relevance criterion. However, several interviewees mention that information about department is important for searching as well as assessment of relevancy. The ‘Production/responsible department’ access point contains information about the ‘local’ department, which is responsible for the broadcasts, whether it concerns production or purchase. That is, the purpose of the information is for the interviewees to be able to see the department, which is responsible for either producing or buying the programme. As mentioned in Section 7.2.2 the information about producing or responsible department is used to specify the perspectives in the programmes, and hence for gathering information about the content in order to select relevant broadcasts. In this way, the access point is comparable to the publisher access point in a textual context.
Similar to ‘Production/responsible department’, ‘Production country’ is not listed in the questionnaire, but several of the interviewees mention it. This access point lets the user retrieve programmes originating in particular countries. The access point is comparable to the registered address of the publisher in a traditional textual context, and in the present television context it is considered relevant for scientific analysis, similar to the bibliometric analyses conducted with ISI’s citation indexes (The Thomson Corporation, 2007). In the present context, such an analysis could be concerned with audience ratings related to the country of origin of programmes. Further, ‘Production country’ is used for selecting broadcasts relevant for the task at hand.

‘Geographical location of recording’ is marked as useful by approximately 10% of the questionnaire respondents, only, and it is not considered by any of the interviewees. The counterpart in a textual context would be something like geographical location of writing, and such information is only considered important in relation to very specific information needs. Consequently, we do not consider ‘Geographical location of recording’ a significant access point for television broadcasts, which is related to the test participants’ work tasks, and research foci. For instance, if an academic focuses on city development, than ‘Geographical location of recording’ would be important for the academic.

One interviewee (Interviewee A) mentions a need to be informed about the ‘Copyright information (screening)’ of the broadcasts, since this particular aspect of television broadcasts is important for the interviewee’s research. From that perspective, we do not consider it a significant access point in a future television broadcast retrieval system.

Two of the three significant access points from the production category: ‘Production/responsible department’, and ‘Production country’ should be descriptive as well as normative. Descriptive to facilitate the users in retrieval of broadcasts, and normative to relate and collate broadcasts that are produced by the same department. In order to facilitate effective retrieval the ‘Date of production’ should be specified in a standardised normative form, like any other indication of date or time. The ‘Production’ sub-category of access points is comparable to area 4 in an ISBD record or field 260 in a MARC record. In this way, the access points are comparable to access points in a traditional textual context.

8.1.1.3 Physical description
The five access points in the physical description sub-category are ‘Duration’, ‘Subtitles’, ‘Spoken language’, ‘Black/white or colours’, and ‘Format of transmission’. 
More than 50% of the questionnaire respondents consider ‘Duration’ to be important for searching and relevance judgments. Further, duration is considered in several interviews, and it is a significant access point for television broadcasts. If the starting and ending time is known (‘Time of transmission’), the duration is easily obtained and expressed in a normative form. ‘Duration’ is used for identification as well as selection of broadcasts.

The remaining physical description access points of ‘Subtitles’, ‘Spoken language’, ‘Black/white or colours’, and ‘Format of transmission’ are not mentioned by any of the interviewees, but considered important for searching and judgments of relevance by approximately 20% of the respondents or less. Indication of these physical access points is utilised for narrowing the search by particular characteristics of the broadcasts, as well as to assist in the choice of a broadcast. Further, the access points make it possible to collocate broadcasts with that particular characteristic. However, the low frequency of marking of these types of information both as search entries and relevance criteria, and the fact that they are not considered relevant by the interviewees makes us conclude that these four access points are not essential in the present context.

The categorisation of access points in the physical description sub-category is equivalent to the physical description categories in description of textual information objects, e.g., area 5 in ISBD, and field 300 in MARC.

The three sub-categories of bibliographic access points are comparable to categorisation of access points in a traditional textual context. By applying this categorisation with resemblance to the traditional textual context, future work can somewhat easily compare the access points to the traditional context, and hence index the television broadcasts in a way that enables the establishment of relational links. For instance, for different types of information objects with the same responsible person, e.g., a book and a television broadcast by the same person. However, though the sub-categorisation resembles the categorisation of traditional information objects, some of the specific access points in the categories differ from access points in a textual context, e.g., the extent exemplified by duration for broadcasts instead of pages for textual objects. In contrast, the access points related to the screening of the broadcasts are particular for television broadcasts, and hence not comparable to a textual context. These access points are discussed in the following section.

8.1.2 Screening access points

The screening category comprises five access points: 1) ‘Channel’; 2) ‘Date of transmission’; 3) ‘Time of transmission’; 4) ‘Transmission pattern’; and 5) ‘Programming’. ‘Channel’ and ‘Date of transmission’ is considered important for
searching and relevance assessment by more than 75% of the questionnaire respondents. ‘Channel’ is not relevant in a historical context, due to the monopoly of DR prior to 1988. Nonetheless, it is considered very important by questionnaire respondents as well as by interviewees, and it is an essential access point in a future broadcast retrieval system. The access point is an important characteristic for identification of specific broadcasts, as well as for collocation of broadcasts, since broadcasting channel is often a parameter in the scientific investigations. ‘Channel’ is incomparable to access points in a textual context.

As mentioned in Section 8.1.1.2 ‘Date of transmission’ coincides with the ‘Date of production’ for live broadcasts. Along with the ‘Time of transmission’ it gives information, which is important in order to establish when the programme was broadcasted, and it is naturally essential for finding, identifying and selecting television broadcast.

The ‘Transmission pattern’ indicates whether the present programme is a first-time transmission or a re-run. Further, it connects the same programme transmitted at different time. However, it is important that the two or more versions of the programme is kept in the archive, since the academics are interested in all transmitted broadcasts, and it may very well be the small peculiarities of a re-run which is the object of analysis. That is, in a re-run of a live broadcast, the subtitles could be an important part of a scientific analysis, or the user could be interested in knowing how many times a particular programme has been repeated, e.g., how often has the motion picture ‘Pretty Woman’ been transmitted? Or are American programmes more often repeated than Danish or European? The ‘Transmission pattern’ is not listed in the questionnaire, but discovered in our interviews, only. The intellectual content of a first time transmission and a repeated transmission is the same, and consequently, it is to some extent comparable to information about the printing in a textual context. The characteristics of the transmission pattern are important for finding, identifying, and selecting broadcasts.

‘Programming’ is not listed in the questionnaire, but it is extensively discussed in the interviews, since the surrounding broadcasts are an important relevance criterion for the test participants. They need to be informed about the programmes prior and subsequent to the programme in question, as well as programmes from other channels, which the programme in question is up against, so to speak. The ‘Programming’ is a visual presentation of a condensation of information from several access points. ‘Programming’ utilises information about ‘Channel’, ‘Date of transmission’, and ‘Time of transmission’ in order to present the programmes in their ‘natural’ order, and the ‘Title’ of the broadcast is important in order to distinguish the programmes from one another. ‘Programming’ is comparable to an interactive television schedule (see e.g.,
Chapter 8: Research area 2: access point for television broadcasts

Appendix 19), and in this way it is related to the design and construction of the interface. In addition to the transmission access points, the visual presentation of the ‘Programming’ can include several of the bibliographic and/or content access points discussed above. The specific access points to include in the design of the interface is beyond the scope of the present work, but ‘Author’ ‘Participant’, ‘Keywords’, ‘Genre’, and/or ‘Summary’ should be considered. The ‘Programming’ is particular for television broadcasts, and hence not comparable to a textual context, and it is very significant for the academic users in order to select relevant broadcasts. In relation to interviewee B’s information need (see Section 7.1.2.2), ‘Programming’ is also concerned with broadcasts in the interim of programmes, e.g., spots or trailers. In addition, ‘Programming’ is important for the interviewees in relation to the scientific analysis they conduct.

With reference to description of textual information objects e.g., in ISBD and MARC records, the screening category of access points is unique for television broadcasts. This said, on one hand, the four access points could be categorised as bibliographic, since they are concerned with publication or distribution of the broadcasts (slightly comparable to area 4 in ISBD and field 260 in MARC). On the other hand, the uniqueness of the screening access points makes it relevant to consider them in a separate category, which is the approach chosen in the present work.

Nevertheless, the five screening access points are considered very important for the test participants’ information searching behaviour in the context of television broadcasts. ‘Table of content’ and ‘Placement within a serial’ are also related to the screening category, since they are concerned with the order of transmission of broadcasts. However, in the present work we put more emphasis on their relational characteristics, and therefore, they are categorised in the ‘Relational’ category. As mentioned earlier the time and date indications should be in a standardised normative form. In describing the channel, a standardised expression should be agreed upon and applied in order to facilitate consistency in indexing as well as retrieval. Indication of the transmission pattern should also use normative expressions, e.g., re-run. In addition, transmission pattern should include the date and time of previous transmissions as well as the channel of the previous transmission and hereby establish links between relevant programmes.

8.1.3 Content access points

Eight access points are categorised within the category of content, and each access point is dealt with in the following. The eight access points are: 1) ‘Summary’; 2) ‘Keywords’; 3) ‘Genre’; 4) ‘Clips’; 5) ‘Images’; 6) ‘Geographical location of
content/storyline’; 7) ‘Description of image sequences’; and 8) ‘Spoken words’. The access points in the content category are divided into two sub-categories, textual content access points and visual content access points. This corresponds to the distinction made by Yang (2005) in relation to relevance criteria in a moving image context.

8.1.3.1 Textual content access points

‘Summary’ is listed in the relevance criteria section of the questionnaire, only. However, several interviewees express a need for the feature in relation to free text searching as well as relevance assessments. 80.6% of the respondents mark ‘Summary’ as important for relevance assessments, and together with the attention from the interviewees, we conclude that ‘Summary’ is a very significant access point for broadcasts in relation to topical information needs. ‘Keywords’ and ‘Genre’ receives similarly a great deal of attention from the questionnaire respondents, with reference to search entries (83.3% and 80.6%, respectively) and relevance criteria (62.0% and 70.4%, respectively). In addition, the two access points are extensively mentioned in the nine interviews, in relation to searching as well as relevance judgments. For ‘Keywords’ the interviewees express a need for a controlled categorical list, e.g., as expressed by interviewee G in Quote 23. In a few words, the same goes for ‘Genre’, but for this access point, the test participants emphasise the difficulties of the categorisation. That is, the fact that the research have reached a point where genre categorisation is applicable, though at the same time it is considered an ever evolving categorisation. The evolving nature of the categorisation is caused by the constantly developing research on the topic, as well as the development of the media itself, including the establishment of new genres. Here it is important to emphasise that this evolving nature of the categorisation is not particular to a genre categorisation, but a characteristic of all the categorisations discussed in the present work. The access points of ‘Summary’, ‘Keywords’, and ‘Genre’ are comparable to similar access points in a textual context. In addition, all three access points contain textual condensation of the content of the programmes in question. ‘Summary’ is an uncontrolled subject access point, which contains a textual condensation of the content of the broadcast. ‘Keywords’ can be uncontrolled, but the interviewees express a need for a controlled vocabulary. In this way, it is comparable to the formalisation needed for genre categorisation. ‘Summary’, ‘Genre’, and ‘Keywords’ are used for searching for broadcasts. In addition, they are used for identification and selection of particular relevant broadcasts.

‘Geographical location of content/storyline’ and ‘Description of image sequences’ are only considered being important by a minority of the respondents, and the access points does not seem significant in a broadcast retrieval context.
Chapter 8: Research area 2: access point for television broadcasts

Searching in the spoken words is related to broadcasts at the features level of granularity. It is comparable to free text or natural language searching in the full text of textual information objects and its function is to allow for detailed searching, and hence an increase in the retrievability of the broadcast (Lancaster, 2003, p. 253). However, such searching requires transcriptions of the oral speech to text e.g., by applying automatic speech recognition techniques (e.g., Gauvin, Lamel & Adda, 2001; Barras et al., 2002; Evans, 2003; Bailer et al., 2005). The application of automatic speech recognition techniques requires that the broadcasts are available in digital form. Transcription of the spoken words enables automatic textual indexing techniques to be applied. This is particularly applicable for broadcasts, which have a precise vocabulary, and have a factual content. From the interviews with interviewee C and G we learn that the possibility to search in textual representations of ‘Spoken words’ is considered very important. The two respondents are both focusing on news broadcasts, and the factual nature of this type of broadcasts is related to their focus on this access point. ‘Spoken words’ is an uncontrolled subject access point. The function of the access point is to find, identify, and select broadcasts.

The function of these textual content access points (‘Summary’, ‘Keywords’, ‘Genre’, and ‘Spoken words’) is related to a discussion of the indexing policy and/or accuracy to apply in the future broadcast retrieval system, including a discussion of the depth or exhaustivity, the breadth or specificity, assignment vs. derivation of index terms, recall and precision, and controlled vs. uncontrolled vocabulary. ‘Keywords’ and ‘Genre’ collocate broadcasts with similar content, which requires the data to be in a normative or controlled form, similar to the authority control needed for the statements of responsibility as discussed in Section 8.1.1.1. This said the keywords might be expressed by applying an uncontrolled or natural language vocabulary. In contrast, the function of the summary is to allow for selection of broadcasts by conveying the topical content of the broadcasts in a condensed form. In addition, free text searching in the summaries as well as in the spoken words allows for detailed topical access and a general increase in recall. However, this increase in recall will generally mean an increase in the possibility for spurious relationships or false relations to occur, and hereby lower the precision in retrieval (Lancaster, 2003, p. 253). This is in line with the test participants’ needs for high recall retrieval, as reported on in Chapter 7. In contrast, specifying a searching on controlled keywords or genre will probably decrease recall, and increase precision. In consideration of the interviewees proficient procedural search task skills e.g., as expressed in relation to searching and processing data from the TV-Meter database, it is expected that the academics are able to benefit from the use of an ‘expert’ based retrieval system with a diversity of functionalities. In our discussion of
‘Summary’, ‘Keywords’, and ‘Genre’, the access points are considered to concern indexer aboutness as defined in Section 4.1.1, while ‘Spoken words’ is concerned with author aboutness.

8.1.3.2 Visual content access points

In contrast to the textual representations, ‘Clips’ and ‘Images’ contain visual representations of the programmes, and these visual access points are naturally particular for visual information objects. ‘Clips’ and ‘Images’ are only listed in the relevance section of the questionnaire, and they are marked by approximately 1/3 of the respondents. We are aware that images can be applied as a request in a search for (moving) images, but we did not consider the feature relevant in relation to the present investigation. ‘Clips’ are mentioned in two interviews, but otherwise the interviewees do not consider this feature important in relation to retrieval of broadcasts. In the present work, these visual access points do not seem overly significant in the minds of the users. However, we do believe that the access points will be increasingly significant for relevance assessments in the future. The content access points of ‘Images’ and ‘Clips’ are particularly important for selecting broadcasts of relevancy for the work task at hand. They provide a visual supplement to the obvious limitations of the textual content access points. In the present work, we have not focused on the use of visual search techniques, but visual techniques can be applied for searching in combination with traditional textual in a hybrid system as proposed by Enser (2000). This is further discussed in the following Chapter 9.

The textual content access points (‘Summary’, ‘Keywords’, ‘Genre’, and ‘Spoken words’) are comparable to subject access points in a textual context, e.g., the subject access fields (fields 600-699) in a MARC record. The two visual content access points (‘Images’ and ‘Clips’) are particular for visual information objects, and are obviously not comparable to access points for textual information objects.

8.1.4 Archival access points

The archival access points are more related to the archive or library function than to the television broadcasts as such. In this way, the archival access points are all concerned with the way users can obtain the broadcasts rather than the way the broadcasts can be described. Further, the archival access points of ‘Medium’, ‘Waiting on loan’, and ‘Borrower options’ are comparable to access points in a traditional library setting, since the archival management data to a large extent is independent of the types of information objects in question.
The three archival access points are listed both as search entries and relevance criteria, and they are marked by between 60% and 75% of the respondents. Further, the interviewees mention the importance of being informed about the programmes they can access, as well as having the opportunity to narrow their search to programmes, which are available within a short period of time. Hence, the archival access points are essential in retrieval of television broadcasts, and they are especially important for relevance assessments and for obtaining access to the broadcasts. The data in the archival access points are comparable to so-called local data in a textual library context, since they are concerned with data, which are particular to the archive function and not the broadcasts per se.

8.1.5 Relational access points

The relational access points are concerned with the relation between a broadcast and some other entity (e.g., another broadcast or a user). The relations are divided into media internal relations and media external relations. The media internal relations comprise relations at the granularity level of serials and features, respectively. The point of departure for the access points is the granularity level of programmes, and consequently, the internal relational access points are important in order to facilitate retrieval at other level of granularity than the programmes level. The three sub-categories of relational access points are considered in the following three sub-sections.

8.1.5.1 Serial

The ‘Placement within a serial’ is an indication of the total number of programmes or episodes in the serial and the position of the programme in question, in relation to the whole serial (e.g., 2:3). In this way, it is not as linear in time as the ‘Programming’, since succeeding episodes seldom are transmitted in direct continuation of each other. This relationship is comparable to the so-called sequential relationship for bibliographic records (Tillet, 2001, p. 20). The access point operates at the serial level of granularity, and other access points should naturally also be available at this level of granularity, most notably the title of the serial, as well as authors responsible for the serial. The serial information is comparable to the serial information in a textual context, e.g., area 6 in ISBD and field 440 in MARC records, and it is important for retrieval since many information needs are at the serial level of granularity. The function of the serial access points is primarily to relate and to collocate broadcasts, which are part of the same serial, and it is necessary in order to facilitate retrieval at the serial level of granularity. In this way, the serial access points serve as to provide relational links between programmes and the series of broadcasts, which the programme in question is part of.
8.1.5.2 Features

With focus on the features level of granularity, the ‘Table of content’ is an important access point mentioned by the interviewees. It is a listing of the features of a programme, and it is consequently associated to types of programmes with many features, e.g., magazine programmes or news broadcasts. The access point is similar to the table of content in textual information objects, e.g., journals or books. In relation to the description of textual information objects, this type of information is generally provided in note fields e.g., area 7 in ISBD or field 505 in MARC records. The ‘Table of content’ requires that the features of the programmes are identified, and that specific access points are derived for each feature, e.g., access points in relation to title and statement of responsibility. The specific access points that should be provided at this level of granularity is beyond the scope of the present work, but e.g., the possibility to search in transcriptions of the spoken words is related to programmes at the features level of granularity. Feature access points are comparable to the serial access points, but only at a lower level of granularity. That is, the feature access points’ function is to provide access to broadcasts at the features level of granularity, and to collocate features, which are parts of the same programme.

8.1.5.3 External

‘Audience ratings’ is the sole access point which addresses a relation to external entities, in this case the viewers of the broadcasts. The ‘Audience ratings’ is comparable to citations in a textual context, and it is considered to be a measure of the impact of particular broadcasts. The questionnaire respondents consider it important in relation to searching (49.1%), as well as for relevance assessments (39.8%), and all the interviewees use information about audience ratings in their retrieval and assessments. The interviewees express that they do not consider the facility as a part of the IR system, since it is already available in relation to the TV-Meter system. However, they do express a need for the information in relation to their retrieval, and we find it an important access point for retrieval. Further external relationships are naturally also necessary if the surrogate records of the television broadcasts are to be related to the other surrogate records in the bibliographic universe. For example, a search for ‘Rosens navn’ [The Name of the Rose’] in the State and University Library’s OPAC, should retrieve surrogate records for the book by Umberto Eco (in its many manifestations), the CD soundtrack composed by James Horner, the DVD of the motion picture directed by Jean-Jacques Annaud, and the television transmissions of the motion picture (though copies of these may not be accessible for the user).
This ends our reporting on and discussion of each of the search entries and relevance criteria mentioned in the previous chapter. In the following section, we summarise our results regarding appropriate access points in a television broadcasts context, including a listing of the access points which the present work have identified to be significant in a future broadcast retrieval system, and summary of the functions of the appropriate access points.

8.1.6 Discussion and summarising statements

Table 8.1 recapitulates the significant access points for retrieval of television broadcasts. Hereby, the table answers which access points, are appropriate for indexing of television broadcasts, which is the first part of the fourth research question. The appropriate access points are: ‘Title’, ‘Participant’, ‘Author’, ‘Date of production’, ‘Production/responsible department’, ‘Production country’, ‘Duration’, ‘Channel’, ‘Date of transmission’, ‘Time of transmission’, ‘Transmission pattern’, ‘Programming’, ‘Summary’, ‘Keywords’, ‘Genre’, ‘Spoken words’, ‘Clips’, ‘Images’, ‘Medium’, ‘Waiting on loan’, ‘Borrow options’, ‘Placement within a serial’, ‘Table of content’, and ‘Audience ratings’. The table is constructed in a similar way to Table 7.9 and Table 7.13. That is, the first column contains category, the second sub-category, and the third column lists the access points. Further, the fourth column in Table 8.1 relates the 24 access points to the three dimensions of a retrieval system, as discussed in Section 7.1.4. Hence, the table provides a focus on the significant access points and leading to a clearer understanding of the use of the access points in relation to the archive, transmission, and relational dimension of the future broadcast retrieval system.

Table 8.1 is not meant as a definitive listing of access points in a television broadcast context, but merely as a proposal of the most significant access points that ought to be considered when indexing broadcasts for academic users. In addition to the access points listed, it might be considered to include other access points. First of all, access points connected to the Serial and Features levels of granularity are needed in the indexing. Secondly, designers and constructors of a future broadcast retrieval system should considered including other access points, e.g., some of the physical descriptive access points not listed in Table 8.1 though they did not attract much attention from the test participants. The physical description access points are used for narrowing the search to broadcasts with particular characteristics, which is relevant for specific research foci, and in a cost benefit analysis it could be considered whether to include them and thereby help the researchers in their work. This is expanded in the following Chapter 9. Thirdly, additional relational access points should be considered in order to bind the broadcasts to the adjoining bibliographic universe.
Table 8.1: Summary of significant access points for retrieval of television broadcasts. Listed according to category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-category</th>
<th>Access point</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographic</td>
<td>Title and Statement of responsibility</td>
<td>Title; Participant; Author</td>
<td>Archive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Date of production; Production/responsible department; Production country</td>
<td>Archive</td>
</tr>
<tr>
<td></td>
<td>Production</td>
<td>Duration</td>
<td>Archive</td>
</tr>
<tr>
<td></td>
<td>Physical description</td>
<td>Channel; Date of transmission; Time of transmission; Transmission pattern; Programming</td>
<td>Transmission</td>
</tr>
<tr>
<td>Screening</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>Textual</td>
<td>Summary; Keywords; Genre; Spoken words</td>
<td>Archive</td>
</tr>
<tr>
<td></td>
<td>Visual</td>
<td>Clips; Images</td>
<td>Archive</td>
</tr>
<tr>
<td>Archival</td>
<td></td>
<td>Medium; Waiting on loan; Borrow options</td>
<td>Archive</td>
</tr>
<tr>
<td>Relational</td>
<td>Serial</td>
<td>Placement within a serial</td>
<td>Archive</td>
</tr>
<tr>
<td></td>
<td>Features</td>
<td>Table of content</td>
<td>Archive</td>
</tr>
<tr>
<td></td>
<td>External</td>
<td>Audience ratings</td>
<td>Reception</td>
</tr>
</tbody>
</table>

The second part of the fourth research question concerns the function of the identified access points. Here we relate each of the 24 access points to the four types of information needs identified in Chapter 7, and the four generic user tasks defined in the study of functional requirements for bibliographic records (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998, pp. 8-9). The answers to the second part of the fourth research question are summarised in Table 8.2. The table is constructed similarly to Table 7.10 and Table 7.14 in that the first column contains the category, the second column contains the access points, and the relative value in relation to each of the four types of information needs are provided in the last part of the table. In the middle of Table 8.2 the relative value in relation to the four generic user tasks or functions is depicted. ‘P’ signifies that the access point is of primary value for the function or the need. ‘S’ signifies secondary value, and no marking signifies low or no value.
The aspect of time plays an important role for television broadcasts, in comparison to the description of traditional textual information objects. Apart from the publication year, or ‘Date of production’ in our context, the indication of the date and time of transmission is a key element in the retrieval of broadcasts. This is shown in our placement of the different time indications. ‘Date of production’ is placed in the bibliographic category, which contains access points that are akin to the access points for textual information objects. In contrast, the indication of the date and time of

Table 8.2: The function of the appropriate access points for television broadcasts.

<table>
<thead>
<tr>
<th>Access points</th>
<th>Find</th>
<th>Identify</th>
<th>Select</th>
<th>Obtain</th>
<th>Verificative</th>
<th>Factual Data</th>
<th>Clear Topical</th>
<th>Muddled Topical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>P</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date of production</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production/responsible</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channel</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Date of transmission</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Time of transmission</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Transmission pattern</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Programming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>P</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Genre</td>
<td>P</td>
<td>S</td>
<td>S</td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Keywords</td>
<td>P</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Spoken words</td>
<td>P</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Clips</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Images</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Archival</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borrow options</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waiting on loan</td>
<td>S</td>
<td>P</td>
<td>S</td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Medium</td>
<td>S</td>
<td>P</td>
<td>S</td>
<td></td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Relational</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placement within a serial</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table of content</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audience ratings</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend: P signifies that the access point is primary in relation to the specified function, or type of information need, while S signifies a secondary relation.
transmission is placed in the screening category, which contains access points that are particular for the television broadcast context.

In Section 4.1 we defined subject indexing to be concerned with creation of representations of content or intellectual characteristics of information objects, while descriptive indexing is concerned with the formal data or the isness of the information object. With these definitions in mind the six access points in the content category are concerned with subject indexing while the 15 access points in the categories of ‘Bibliographic’, ‘Screening’, and ‘Relational’ are concerned with descriptive indexing. The remaining three archival access points are considered archive management data.

Though the present work does not focus on cataloguing, our results show that the objectives of a catalogue for television broadcasts are not fundamentally different from the objectives of a catalogue for textual information objects, which is expressed in our use of the four generic user tasks defined in the study of functional requirements for bibliographic records (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998). The similarity is also expressed by the interviewees, e.g., interviewee G in the following quote:

“I do not perceive it [retrieval of broadcasts] to be basically different from the Royal Libraries approaches for systematic description of literature, both scientific literature and fiction. The retrieval tool is in a sense excellent, but of course the nature of the audiovisual information objects require a need for content descriptions, which are not available for books in the Royal Library”.

Quote 43

The appropriate access points are somewhat different from access points in a traditional textual context, and the knowledge gained about these access points is important in order to design a future broadcast retrieval system. First, the access points in the screening category are particular to transmitted information objects; including the access point ‘Programming’ which depict the broadcasts in the transmitted context. Secondly, the visual content access points (‘Clips’ and ‘Images’) are particular for visual information objects. The knowledge learned about these access points should be included in the construction of a future broadcast retrieval system in order to facilitate effective retrieval. Nevertheless, as indicated with our five categories applied for categorisation of search entries in Section 7.2, relevance criteria in Section 7.3, and access points in the present chapter, some access points are comparable to access points in a textual context (e.g., ‘Title’, ‘Summary’, and ‘Borrow options’). Consequently, we believe it should be possible to construct relational links to other types of information objects in the bibliographic universe.
In the succeeding Chapter 9, we focus on the construction of the access points identified in the present chapter. That is, first and foremost we focus on the applicability of information available in textual television schedules. Secondly, we give attention to other approaches for constructing the access points in question, approaches that are outlined and reviewed in Chapter 4.
9 Research area 3: construction of surrogate records

This chapter focuses on the third research area on construction of surrogate records for television broadcasts, and more specifically the aim of the chapter is to provide answers to research question 5, which states:

To what extent does external textual and other sources contain or provide information valuable for constructing the identified access points?

In this way, the present chapter is a natural continuation of the previous two empirically based chapters concerning aspects of the users’ information seeking behaviour in relation to television broadcasts, and the thereof derived appropriate access points for indexing and retrieval of broadcast. The present chapter provides answers as to the construction of surrogate records for television broadcasts, based on the knowledge gained about characteristics of the test participants’ information needs, search entry preferences, and relevance criteria applied. In other words, the chapter is concerned with the third research area stated in introductory chapter. As indicated in research question 5, the main focus of the present chapter is the applicability of textual external and other sources, an in particular television schedules. The primary reason for this focus is the television schedules’ readily available textual content (e.g., see our reasoning in Section 1.2). The television schedules are explored in relation to each of the 24 appropriate access points identified in the previous chapter. Thus, we relate aspects of the scholars’ and students’ seeking behaviour identified in Chapter 7, and the implication for design of surrogate records, discussed in Chapter 8, to considerations about indexing and construction of television broadcast surrogate records.

Secondly, we cursorily pay attention to other approaches for indexing the identified access points, including the manual and automatic approaches outlined in Chapter 4. These approaches (e.g., Hauptmann & Witbrock, 1998; Bimbo, 1999; Satoh, Nakamura & Kanada, 1999; O'Connor et al., 2001; Lienhart, 2003; Hanjalic, 2004; Bailer et al., 2005) are discussed in connection to exploration and discussion of the usefulness of television schedules for describing each of the 24 access points. Automatic indexing of a continuous stream of heterogeneous television broadcast is a very complex research area, which comprises comprehensive algorithmic insight, and the area have given rise to several PhD theses on its own (e.g., Lee, 2001; Petkovic, 2003). Therefore, the present work does not aim at a comprehensive discussion of the usefulness of automatic indexing of television broadcasts. Instead, we focus on those
automatic approaches we find most applicable in relation to the access points identified in the present investigation.

Thirdly, the chapter focus on analysis of existing surrogate records for television broadcasts, in order to learn the value of exchanging records with operating archives.

9.1 Research question 5: to what extent does external textual and other sources contain or provide information valuable for constructing the identified access points?

In order to investigate the usefulness of television schedules for indexing of television broadcasts, we analyse a selected television schedule from DR1. In addition, we analyse data obtained in the nine interviews about the appropriateness of television schedules for retrieval of television broadcasts. The television schedule comprises programme descriptions from one week, namely week 16 in 2007 (April 16th – April 22nd). See the full television schedule in Appendix 5. This week is selected at random as a representation of a ‘normal’ week of broadcasting, and schedule captures the different types of descriptions in the television schedule. In the selection we purposefully chose not to focus on a week with special events, e.g., the soccer world cup. In this way, the objective of the present work is to provide answers as to the general applicability of television schedules. Future work can focus specifically on the appropriateness in relation to the different types of broadcasts, and the usefulness of schedules from different television channels.

Since no search facilities are available for television broadcasts, the test participants often use television schedules as data sources and retrieval tools for their research. Therefore, the nine interviewees are familiar with the schedules and they are able to give information about the usefulness of television schedules for retrieval of television broadcasts. For instance, Interviewee F clearly states that television schedules are political statements.

*INTERVIEWEE F:* “They [the television schedules] are relative informative in the early period, but I would have liked to be informed for instance about the name of the commentator of every sports transmission. The television stations have been inclined to include the information when it is of some political importance, and thereby mark that it was Gunnar Hansen [a famous Danish sports commentator] who commented”

[...]

246
INTerviewee F: “and that is why I mean that it is obvious that they [the television schedules] are political statements. The television schedules are strategically statements from the organisations”

[...]

INTerviewee F: “Now with Karin Palshøj. That was important because she were the first female host. At a particular point in time, the media people gain a star function, and that development is traceable in the television schedules. However, it would be interesting to have some of the waterboys as well. Hereby I would have been able to prove that Gunnar Hansen always took the best pieces of the cake”.

Quote 44

Interviewee F further states that the comprehensiveness of the programme descriptions changes in line with the television station’s need for advertising for a specific programme, e.g., extensive descriptions in the release of a new serial. The television schedules are the television station’s way of ‘selling’ a programme to the audience. In light of the increasing competition on the media market e.g., as described in Section 5.1, the television stations are becoming more aware of the television schedules ‘marketing value’. In the third part of the interview guide, we collect the interviewees’ knowledge and perception of the information provided in television schedules (e.g., see the interview guide in Appendix 8 and our description of the guide in Section 6.4.1).

The present section reports on each of the 24 access points in the five main categories of Table 8.1. Prior to the discussion of the usefulness of the information available in the television schedule, Table 9.1 depicts examples of ‘entries’ in DR1’s television schedule for week 16 in 2007. In our reporting in the succeeding subsections, we primarily focus on these examples, and if not otherwise stated, programme titles refer to programmes listed in Table 9.1. The complete television schedule for week 16 is reproduced in Appendix 5, and the examples in Table 9.1 are selected in order to show the diversity in the descriptions in the weekly television schedule.

The television schedule is constructed at the granularity level of programmes, which coincides with our definition of the granularity level in Section 4.2.2. However, the schedule also contains what we denote framework programmes that consist of two or more entries for individually scheduled programmes e.g., ‘Barracuda’, ‘Fjernsyn for dig’ or ‘TV AVISEN med Horisont og SportNyt’ in Table 9.1. These framework programmes challenges our definition of programmes, because an abrupt cut or abrupt transition does not necessarily exist between each programme in a framework programme. In total, the schedule for week 16 contains descriptions of 269 programmes and 20 framework programmes.
The present section comprises eight sub-sections. In Section 9.1.1, we pay attention to the segmentation of broadcasts into different structural units, since segmentation is the basis for indexing of broadcasts. Subsequently, we focus on each of the 24 access points in five sub-sections (Sections 9.1.2 - 9.1.6). Section 9.1.7 relates and discusses the access points identified in the present work in relation to access points in existing broadcast retrieval system. Finally, Section 9.1.8 summarises and discusses our results and provides answers to research question 5.
Chapter 9: Research area 3: construction of surrogate records

9.1.1 Segmentation of broadcasts

Television broadcasts consist of a continuous stream of image and sound, normally starting in the morning and ending at some point during the night. In some cases (e.g., during the Olympics or for the channel: dk4) the stream is not ended during the night, and it is consequently an endless audiovisual stream. In order to provide access to the broadcasts, the streams must be broken down into structural units, as mentioned in connection to Figure 4.1. With reference to the present work, the main structural unit is programme, and secondary structural units are serials and features. Segmentation of the broadcasts does not necessarily mean that the broadcasts are archived in a segmented form, since date and time codes can be used for identifying segments (e.g., programmes) in a continuous audiovisual stream. However, it is necessary to operate with surrogate records for segmented units of broadcasts in the indexing and retrieval of broadcasts. Surrogate records are needed for broadcasts at the programmes, serial, and features levels of granularity, as we discussed in Section 7.1.3, but the surrogate records may contain time codes to point to different parts of the same computer file or be provided with a link that gives direct access to the searched unit of broadcasts. This is the approach taken for the digital part of the collection of broadcasts at the State and University Library, where one file comprises the transmission from one channel for one day, and time codes are used to identify specific structural units or segments within that day’s transmission. These time codes provide the direct connection between the surrogate record and the specific segment of the audiovisual file in question, and in this way the surrogate records deals with segmented broadcasts, whether the broadcasts are store in a segmented form or not. Further, such an approach can be used for broadcasts stored in an analogue form as well, though direct linkage to the segment is not possible for the analogously stored broadcasts. Furthermore, different archives or indexes might link their surrogate records to different segments of the same audiovisual computer file. As such, though different institutions need to have access to television broadcasts, e.g., DR and the State and University Library, the broadcasts can be stored in one central digital archive, and hence save storage resources, since it is unnecessary to store the same broadcast in two places.

An essential element in the television schedule is the indication of start time of programmes, and consequently it is possible to extract time codes from the television schedule, and hereby provide information that is useful for segmentation in the construction of the surrogate records. However, the television schedule only provides crude approximated time stamps for the beginning and ending time of programmes. For instance, if a programme is delayed 5 minutes, this information is not available in the
television schedule, as the schedule is constructed prior to transmission of the broadcasts. Therefore, relying on such crude time stamps entails inaccurate retrieval, where the first minutes of an obtained broadcast may not be the programme in question. The problem is even more profound in cases where the schedule is not followed, e.g., because of extra news programmes as on September 11, 2001, or when a soccer transmission is prolonged due to extended play.

The accuracy of the segmentation of the continuous stream of broadcasts into programmes could be supported by applying automatic segmentation techniques. The time codes provided in the television schedule could be used to indicate probable places for segmentation, which could be a parameter or weight for an automatic segmentation algorithm. Further, additional data may be applied to ensure the accuracy of the time stamps, e.g., the data provided in the Programme Delivery Control (PDC) system. In the PDC system, time stamps are ‘hidden’ in the teletext and transmitted along with the programmes in order to indicate to a video recorder when a programme starts and ends. Furthermore, manual checking of the identified segmentations can be applied to ensure accuracy in segmentation. Manual checking is very resource demanding and consequently random samples could be considered.

Subsequent to segmentation into programmes, relevant programmes (e.g., news and magazine programmes) are to be further segmented into features. The television schedule does not hold any information for segmentation at the features level. For this segmentation, it is necessary to design algorithms that are type specific, e.g., one algorithm for news broadcasts, and another for magazine programmes and so forth. To ensure accurate segmentation it might even be necessary to design algorithms for the specific serials of programmes (e.g., the series of ‘Kontant’ programmes on DR1). As reported in Section 4.2.2, specific algorithms for structured types of programmes provide the best results in segmentation (Bimbo, 1999, p. 225), and promising results have been obtained with segmentation of news broadcasts (e.g., O'Connor et al., 2001; Stokes, Cathy & Smeaton, 2002). With these results in mind, and the fact that news broadcasts are important for academics (see Section 7.1.3), it is recommendable to consider automatic segmentation of news broadcasts into features. Furthermore, textual external sources, which are available in connection to the Internet transmission of news broadcasts, might be applied to enhance the accuracy of the segmentation (e.g., see http://www.dr.dk/DR1/TVAVISEN/tvaindslag.htm for transmission of news from DR).

The segmentation of broadcasts into the granularity level of serials is not discussed in this section, because it is concerned more with creation of relational links between programmes, since serials are not a continuous stream of broadcasts.
In brief, the television schedule provides data, which can be applied for crude or slightly inaccurate segmentation of the broadcasts into programmes, only. Future work should focus on ensuring the accuracy of the segmentation by applying a hybrid of approaches, including the television schedule, data from the PDC system, and automatic segmentation techniques. The television schedule does not provide any information for segmentation into features, and feature segmentation must hence be conducted by designing algorithms for the specific type(s) of programmes in question.

In the following five sections, we discuss how surrogate records for the segmented broadcasts might be constructed, and in particular, we discuss the value of information in television schedules for such construction.

9.1.2 Bibliographic access points

The bibliographic access points of 1) ‘Title and Statement of responsibility’, 2) ‘Production’, and 3) ‘Physical description’ are discussed in the following three sub-sections, which is in line with the structure applied in the previous chapter.

9.1.2.1 Title and statement of responsibility

The title and statement of responsibility sub-category comprises three access points, namely: 1) ‘Title’; 2) ‘Participant’; and 3) ‘Author’. The title of a programme is the main entry in the television schedule, and title information is hence readily available. However, the title entry in the schedule is sometimes the serial title and not the programme title. An example from Table 9.1 is ‘Gurli Gris’, where the subtitle (‘Museet’) is the title of the programme in question, while the main title (‘Gurli Gris’) is the title of the serial. For the programme ‘Den europæiske mand’ – [‘The European man’] we are only provided with the serial title, and the fact that it is the first of four programmes in the serial (1:4). Hence, the programme title must include information about serial number in order for it to be distinctive. Further, for several programmes the title solely refers to the serial. Examples are ‘Aftenshowet’, TV avisen’, ‘Horisont’, and ‘Nyheder på tegnsprog’. For these programmes, it is necessary to provide additional information about the date of transmission in order to differentiate the programmes unambiguously. This is comparable to the title of textual journal issues, which are not distinct unless additional data is provided, e.g., volume and number. For foreign productions, we are provided with the original title of the programme. An example is given in ‘Inspector Rebus’, with the original title in parenthesis (Rebus). Extraction of this additional information about original title of foreign productions should be considered, similar to the way it is considered for textual information objects. Despite the fact that the title information is available for the programmes, the ‘level’ of the
stated title needs to be established before the information can be extracted from the schedule.

In contrast, information about the participants is not readily available in the schedule. The information ranges from no information at all (e.g., ‘Den europæiske mand’) to a detailed listing of the participants (e.g., ‘Hammerslag’), with the weight given to the former. In between are programmes where participants are briefly mentioned in the summary description (see Appendix 5, the programme ‘Kongehuset’, Monday 08.30, where Prince Henrik is mentioned in the summary). In brief, information about participants is seldom available in the television schedule, and when available it is often unstructured in the summary description. The availability of participant information is very different from programme to programme, which is in line with the fact that television schedules are political statements as expressed in Quote 44. In other words, when the participant is known in advance or have gained a star function, the information is likely to be included in the schedule in order to sell the programme to the audience (see e.g., Quote 44).

Participant information is an openness attribute (see Table 4.2), and therefore the semantic gap between computational processing and human understanding must be bridged in order to apply automatic recognition and extraction of people in pictures, e.g., participants. An algorithmic representational model is a prerequisite for the present automatic tools for recognition of people (e.g., Li et al., 2001; Mohan, Papageorgiou & Poggio, 2001). For instance done in analysis of news programmes by Satoh, Nakamura, and Kanada (1999) in the experimental Name-it retrieval system. The reliance on the representational means that only persons known in advance can be recognised automatically. Participants are often listed in the roller captions at the end of the television programmes, and automatic video optical character recognition (OCR) tools (e.g., Lienhart, 2003) can be applied to identify names from these roller captions, and hereby help bridging the semantic gap. However, the listing of a name does not necessarily mean that the person is a participant in the programme. Further, all participants are not necessarily mentioned in the roller captions, and would hence be missed by solely relying on video OCR tools. Traditional indexing requires the viewing of each programme and manual recognising and describing each participant. From the three visits at television stations we know that this is the present approach taken for recognition and description of participants in broadcasts (e.g., see medvirkende [participants] in Appendix 21). We agree with previous researchers (e.g., Smeaton, 2004), that automatic approaches are to be considered a complement to traditional human approaches for recognition of participants, but the information provided by automatic video OCR can provide a list of names, which subsequently needs to be
checked manually, also denoted semi-automatic indexing. As follows, the application of automatic video OCR can reduce the resources needed for indexing of participants.

Authors are often explicitly mentioned in the television schedule. Examples are ‘Troldspejlet’ and ‘Inspector Rebus’ where the roles of the authors are indicated as well (e.g., Vært [Host], Manuskript [Manuscript], and Instruktion [Direction]). The explicit mentioning of authors is nonetheless not present for all programmes. For instance, in ‘Aftenshowet’ and ‘Mission integration’ the hosts (Sisse Fisker and Mads Steffensen, and Özlem Cekic, respectively) are mentioned in the summary, and it is not explicitly stated that they are the hosts. For the remaining programmes in Table 9.1 we do not have any information about authors. Authors are defined as the responsible people behind the camera. Therefore automatic content-based analysis of broadcast content is futile for recognition of authors. Instead, automatic video OCR techniques as described in relation to recognition of participants are applicable for semi-automatic indexing of ‘Author’.

Approaches for indexing of authors are similar to the approaches available for indexing of participants, and hence, automatic video OCR tools could be used as a complement to the traditional human intellectual approaches.

For the three ‘Title and Statement of responsibility’ access points, the indexing needed is of a descriptive nature and the information available in television schedule is only explicitly available for indexing the ‘Title’. For both the ‘Participant’ and ‘Author’ access point, the information available in the schedule is sporadic, and extraction of the information will not result in consistent indexing. Automatic video OCR techniques can be applied to extract information about participants as well as authors, and future designers and constructors should consider using automatic video OCR techniques, to supplement the information extractable from the schedule.

9.1.2.2 Production

The production sub-category of bibliographic access points comprises the access points ‘Date of production’, ‘Production/responsible department’, and ‘Production country’.

The date of production is mentioned for the fiction programme ‘Inspector Rebus’, alone, or rather the year of production is mentioned. For the live transmissions, the date of production is similar to the date of transmission, but for the remaining programmes, the schedule does not hold any information about date of production, and the information is hence not available. The information is generally available in the roller credits of each programme, and hence automatic video OCR techniques might be applied to extract such information.
Production or responsible department is mentioned in two cases of the whole weekly schedule, only. That is, the mentioning of DR TV-Drama in the programmes ‘Forestillinger (5:6)’ on Saturday and ‘Forestillinger (6:6)’ on Sunday in Appendix 5. For all remaining programmes, the weekly schedule does not hold any information about the production or responsible department. In both cases, it is the producing department, and information about the department responsible for purchasing a programme is not mentioned in the schedule. Further, responsible department is information, which is external to the broadcast itself and as such not likely to be found in the audiovisual data. Automatic content-based approaches are not applicable for deriving information about departments responsible for purchasing and transmitting the programmes. The co-producing television stations are mentioned in some cases, e.g., DR, YLE (Finland), TV P (Poland), and ERT (Greece) in ‘Den europeiske mand’, or ‘Nordvision fra Sverige’ in ‘Italienske fristelser’ (see Monday in Appendix 5). The need for information about production or responsible department is to support the user in getting the picture of the content of the programme in question, and the latter mentioning of production corporations between television stations does not provide such information. The mentioning of DR TV-Drama in the two former cases provides such information, only. The latter cases provide information about the producing country, which is also a significant access point for retrieval of television broadcasts. For all foreign fiction programmes, information about producing country is provided just below the title, e.g., for ‘Inspector Rebus’ it is stated that it is from Scotland. However, information about producing country is generally not available for the non-fiction programmes.

In summary, information valuable for the production category of access points is provided in few cases in the schedule, only. Similarly to our discussion of the ‘Title and statement of responsibility’ access points, automatic video OCR scanning of roller credits should be considered as a complement to the diminutive information available in the schedule, and the traditional manual approaches for indexing.

9.1.2.3 Physical description
In the previous Chapter 0 we identified ‘Duration’ to be the sole important physical description access points. Duration of programmes is easily available, since start time is one of the key elements provided in the schedule. The duration is simply the time passed from one start time until the next start time. As follows, the calculated duration is only an approximation, because the broadcasts in the interim of programmes (e.g.,

---

8 Nordvision is a corporation between five Nordic public service stations. These stations are Danmarks Radio (DR), Norsk Rikskringkasting (NRK) (Norway), Sveriges Radio og Sveriges Television (SR/SVT) (Sweden), Yleisradio (YLE) (Finland), and Rikisutvarpid (RUV) (Iceland).
spots or trailers) are included in the calculation. In addition, for the last programme of the day (‘Godnat’) we do not have a new start time and the duration must rely upon estimation. For television programmes in a digital form, the duration is readily available, given that it is related to the size of the computer file, and it can easily be extracted. For that reason, automatic extraction of the duration of a television programmes is closely connected to the segmentation of the ongoing broadcast stream, and ensuring accuracy in segmentation entails ensuring accuracy in calculation of the duration of programmes, or features for that matter.

In the concluding remarks of Chapter 0, we state that additional physical access points might be considered when designing a future broadcast retrieval system. These access points could for instance include ‘Subtitles’ and ‘Spoken language’. Subtitles are the rule for foreign productions, e.g., ‘Inspector Rebus’, except for the children’s programmes, which are generally dubbed with Danish voices, and this is occasionally mentioned, only (e.g., ‘Shin chan’ on Tuesday in Appendix 5). For the remaining broadcasts, information about subtitles is provided for broadcasts in the schedule. That is, ‘TTV’ and ‘TH’ indicate that subtitles are available via teletext. Subtitles are provided in a standardised form, a black box with white letters in the bottom of the screen, and it is therefore easily detectable with automatic content-based techniques (Lienhart, 2003). Information about the spoken language is not explicitly provided for any of the programmes in the schedule for week 16, except for ‘Shin chan’ as mentioned above. For all non-children’s programmes, the language is generally the original language, and information can hence be derived from information about production country. Further, designers should look into the possibility of applying automatic language detection techniques (e.g., Zissman & Berkling, 2001) to automatically detect the language of broadcasts.

9.1.3 Screening access points

Five screening related access points are identified as being appropriate for broadcast retrieval. These are: 1) ‘Channel’; 2) ‘Date of transmission’; 3) ‘Time of transmission’; 4) ‘Transmission pattern’; and 5) ‘Programming’. ‘Channel’, ‘Date of transmission’, and ‘Time of transmission’ are at first hand easily obtainable from the schedule, because such information is the backbone of the television schedule, and the information is extractable in a precise and standardised form. Nonetheless, the information provided about date of a transmission is not unambiguous. For instance, the programme ‘Boogie Update’ is listed on Wednesday April 18th, though the transmission begins a quarter past midnight and consequently, the programme is actually transmitted on April 19th.
Information about transmission pattern or previous and coming transmission of the programme in question is provided in the schedule (e.g., ‘Hammerslag’). Information about previous transmission is only concerned with the first time the broadcast has been transmitted. Transmission information can be derived by establishing links from all re-runs to the first-time transmission, and it is indirectly available in the schedule. Further, the schedule provides information about programmes that previously have appeared on DR’s second channel, namely DR2. Overall, the transmission pattern is well covered in the schedule, and the information should be extractable. The schedule provides horizontal or internal station information, only. Therefore, automatic cross check of titles should be considered. In this way, the archive can derive informed about earlier horizontal as well as vertical transmission of programmes with the same title. Here it should be kept in mind that titles reappear for different programmes (see Quote 32 and Quote 33), and it is not enough merely to identify identical titles. Therefore, the automatic identification of titles and thereby transmission pattern should be considered as suggestions, only, and human intellectual interpretations are needed to ensure accurate data in establishing transmission pattern links.

‘Programming’ is a visual interconnection, condensation, and presentation of information derived from different access points, and it could be discussed whether ‘Programming’ is an access point on its own. However, for the purpose of the present work we consider it an access point, and we see it as an essential element in the design and construction of a future broadcast retrieval system. The information needed in order to construct ‘Programming’ is derived in relation to other access points, especially ‘Title’, ‘Channel’, Date of transmission’, and ‘Time of transmission’, and the value of the television schedule for construction of ‘Programming’ must consequently be judged on the basis of the schedule’s value in relation to these access points. Overall, the television schedule provides valuable information for construction of ‘Programming’.

9.1.4 Content access points

The content category of access points comprises six access points in two sub-categories. The textual sub-category encompasses ‘Summary’, ‘Keywords’, ‘Genre’, and ‘Spoken words’, while the visual sub-category encompasses ‘Clips’ and ‘Images’. This section firstly discusses the textual content access points, and secondly, the visual content access points in two separate sub-sections.
9.1.4.1 Textual content access points

Summaries are found for seven programmes in Table 9.1, but the summaries differ in the level of granularity they address, and their comprehensiveness. Four of these summaries address broadcasts at the serial level (‘Troldspejlet’, ‘Gurli Gris’, ‘Aftenshowet’, and ‘Mission integration’). In other words, they describe the serial and are hence identical for each program in the serial. An example is ‘Troldspejlet’, where the serial summary states: “Nyt fra fantasiens verden i computerspil, film, tegneserier og bøger” - [News from the imaginary world of computer games, movies, cartoons, and books]. Other serial summaries have a short description of the programmes in question. For instance, the summary for ‘Den europeiske mand’ contains the following description at the programme level: “Det første program handler om alenelivet” - [The first programme is about single life]. Two summaries (‘Inspector Rebus and ‘Hammerslag’) address the programme in question, and these summaries contain detailed descriptions. Overall, the schedule for week 16 contains 194 summaries of which approximately 60% are summaries of programmes, while 40% of the summaries mainly are concerned with serials. The summaries are of varying length, ranging from a few words to 155 words in a single summary\(^9\). Hence, the comprehensiveness of the summaries is somewhat erratic. Generally, summaries do not exist for framework programmes, current affairs magazine programmes, and live transmissions, including news broadcasts.

‘Keywords’ are not available in the schedule. However, manual as well as automatic approaches exist for generation of keywords. First of all, controlled as well as uncontrolled keywords might be constructed by manual means, e.g., by considering ofness as well as abstract attributes. Secondly, automatic approaches can be used to identify and extract keywords from text, so-called automatic text classification (Lancaster, 2003, pp. 292-298). Generally, automatic text classification approaches match text against a list of predefined categories, and the techniques show promising results for collections of textual news stories (e.g., Yang, 1999). The television counterpart of news broadcasts might be an appropriate point of departure for applying automatic text categorising techniques. However, a prerequisite for automatic text classification techniques is a textual basis. The textual basis can be constructed either by extracting text from subtitles (Lienhart, 2003), or by transcribing the spoken words (e.g., Evans, 2003; Bailer et al., 2005). Transcription of spoken words is discussed below. Thirdly, automatic classification or indexing of audiovisual content can be applied for automatic identification of content attributes as described in Section 4.2.2.

---

\(^9\) Due to the difference between the Danish and the English language (e.g., the extensive use of compounds in Danish) comparisons between the two languages should be conducted with prudence.
The automatic identification requires the establishment of a representative model, which can be matched with keywords, e.g., for objects (Mohan, Papageorgiou & Poggio, 2001). The approach matches a list of known objects (representative models) against a controlled list of keywords.

A predefined controlled list of genre categories is a prerequisite for automatic genre categorisation (e.g., Rasheed & Shah, 2003). Though the list does not exist at present it is evident that the television schedule does not hold direct information about genre categories. However, some genres may be identified from the schedule. Some programmes have informative titles that indicate the content of the programme. Examples taken from week 16 are ‘TV-Avisen’, ‘SportNyt’ and ‘Fredagsfilm: Lokkeduen’. Further, knowledge about the television schedule as a genre on its own can be applied to derive information about the most likely genre of a particular programme. For instance, knowing that the first broadcasts of the day generally are children’s programmes can be used to increase the weight of the genre ‘Children’s programmes’ in an automatic algorithmic processing of the broadcasts. Time of broadcasting and duration are two types of information that might be extracted from the schedule and provide information for automatic genre identification.

Furthermore, the fact that most of the programming is serialised means that indexing does not need to be conducted on each and every transmitted programme. For instance, each programme in the serials ‘Jersild & Spin’, ‘Horisont’ and ‘Aftenshowet’ are to be categorised within the same genre, and hence genre categorisation can be conducted on the granularity level of serials for these programmes. In this way, knowledge about the genre of television schedules can be used in the construction of algorithms for automatic identification of television broadcast genres. In automatic genre classification the characteristics of different television broadcast genres are used as cues for identification of the genre in question (e.g., Rasheed & Shah, 2003). This includes for instance the number, size, and placement of text and faces in a broadcast, which is applied by Ronfard, Garcia, and Carrive (2000) to automatically identify and categorise news broadcasts, or by Dimitrova, Agnihotri, and Wei (2000) for categorisation of news, commercials, sitcoms, and soaps. Other elements of the audiovisual data can naturally also be applied, e.g., shot change rate and motion vector length (Lienhart, Kuhmünch & Effelsberg, 1997). These multimodal automatic categorisation techniques show promising results, though false categorisation does appear, and automatic categorisations should consequently not be the solely approach applied for genre categorisation. Another channel of data to apply in identification of genre is the spoken words, for instance statements like “this is the news good evening”
or “this evening’s movie is …” could be used as cues for identification of the genre in question.

In brief, information about genre is not directly available in the television schedule, though cues available in the television schedule can be used to provide a supplementing weight in algorithms, and hereby enhance the accuracy of automatic genre categorisation techniques. However, from our point of view, the results derived with automatic genre categorisation techniques are to be treated as educated guesses that need to be confirmed by a human indexer. In order to maintain consistency in allocation of genre categories, it is necessary that human indexers are provided with extensive, clear, and unambiguous definitions of the genre categories. Essentially, the same goes for the human who is responsible for designing the algorithm for automatic categorisation techniques. Some programmes are very difficult to categorise, due to the earlier mentioned fact that the genres are constantly being challenged and developed. Therefore, semi-automatic identification of genre makes it possible for the human indexer to focus on the borderline programmes. Further, it could be considered to use democratic indexing (e.g., Brown et al., 1996) in some way or another, in order to capture new developments and new genre categories. Nonetheless, democratic indexing will be in contrast to a controlled finite list of categories.

‘Spoken words’ is not supported by information available in the television schedule. Construction of the access point must rely upon the application of automatic transcription of the words spoken in a programme. The automatic transcription can be obtained by applying automatic speech recognition techniques (e.g., Gauvin, Lamel & Adda, 2001; Barras et al., 2002; Evans, 2003; Bailer et al., 2005). Hauptmann and colleagues have shown how automatic speech recognition techniques can be enhanced by taking advantage of the closed captions transmitted along with the broadcast, in the context of news broadcasts (Hauptmann & Witbrock, 1998; Hauptmann et al., 2004). ‘Spoken words’ is mentioned by two interviewees, who primarily focus on news broadcasts, and due to the promising results obtained with automatic speech recognition in the context of news broadcasts, we recommend automatic speech recognition techniques are applied in relation to news broadcast and features.

9.1.4.2 Visual content access points
The visual content access points of ‘Clips’ and ‘Images’ are not supported by information available in the television schedule, and the construction of these access points must solely be based on automatic summarisation and skimming techniques (e.g., Mills, Cohen & Wong, 1992; Pfeiffer et al., 1996; Babaguchi, Kawai & Kitahashi, 2002; Hanjalic, 2004). The automatic approaches for construction of ‘Images’ are
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

Concerned with summarisation by extracting key frames and hereby abstracting the content of the programmes by identifying the images that best represent the essence of the content of the programme. The key images can be selected in several ways, ranging from the simple selection of images at specific time intervals (Mills, Cohen & Wong, 1992) to complex multimodal approaches incorporating the similarity and dissimilarity between frames in the programme, e.g., the content coherence principle proposed by Hanjalic (2004).

Where ‘Images’ are static representations, ‘Clips’ is a dynamic way of providing visual representations of the content of a programme. In the process of construction of the ‘Clips’ access point, the programme is condensed into a trailer-like abstract that represents the most important content of the programme, which in Section 4.2.2 is denoted skimming. Techniques for skimming identify the most important ‘events’ in a programme, extracts clips from the events, and merges these clips into a coherent audiovisual unit (e.g., Pfeiffer et al., 1996; Babaguchi, Kawai & Kitahashi, 2002; Li, Narayan & Kuo, 2003). In line with the results obtained in a user study by Agnihotri and colleagues (Agnihotri et al., 2003) Hanjalic state that the relevance of clips or dynamic abstracts is highly genre dependent (Hanjalic, 2004, p. 131). This means that the content considered to be relevant differs for different types of television genres. The example given by Hanjalic is that users are not interested in the complete plot of a motion picture, while the plot is very relevant for news broadcasts and talk shows. We agree with Hanjalic and further add that the relevance of the clips is also highly depending on contextual and situational settings. From a Media Studies perspective the complete plot of a motion picture may be extremely relevant, since the academic does not want the motion picture for entertainment or passive consumption, but rather the academic has an active viewing approach in order to analyse the broadcast in question.

The tasks of automatic identification of representative ‘Images’ and ‘Clips’ are far from trivial. The extraction is the focus in much recent research, and the approaches are pertinent for providing visual presentations of television programmes (e.g., Aner-Wolf & Kender, 2003; Hanjalic, 2004; Bailer et al., 2005).

Another aspect to consider is the possibilities to navigate these visual content access points, which is the fourth tasks emphasised by Auffret and Bachimont (1999). Different tools for navigation of these visual content access points can be constructed to facilitate navigation (e.g., Amir, Srinivasan & Efrat, 2003; Amir, Srinivasan & Ponceleon, 2003; Smeaton, 2004; Smeaton, Lee & McDonald, 2004), but these considerations are outside the scope of the present work.

In summary, the content access points are concerned with subject indexing, while the remaining categories of access points are related to descriptive indexing. The
Chapter 9: Research area 3: construction of surrogate records

‘Summary’ access point is the only textual content access point which to some extent is directly supported by information in the television schedule. For construction of the remaining four textual content access points, indexing must rely on other approaches, preferably a hybrid between manual and automatic approaches, including some cues derived from the television schedule. The two visual content access points must rely solely on automatic summarisation techniques.

9.1.5 Archival access points

The three archival access points of ‘Medium’, ‘Waiting on loan’, and ‘Borrow options’ can not be constructed on the basis of external sources, since these access points are concerned with the library function, and not the information objects or broadcasts per se. The information needed to construct the access points are concerned with rights management (borrower options), as well as an overview of the time needed in order to make broadcasts available (waiting on loan), including knowledge about the medium. This means that if broadcasts are available for download or streaming it affects the time it takes to obtain access to the broadcasts.

9.1.6 Relational access points

The three relational access points of ‘Placement within a serial’, ‘Table of content’, and ‘Audience ratings’ are discussed in separate sub-sections in line with the three sub-categories ‘Serial’, ‘Feature’, and ‘External’ relational access points.

9.1.6.1 Serial

Two types of serials can be distinguished from the broadcasts. The first type contains a finite number of programmes, while the second type is ongoing, containing an ever growing number of broadcasts. For the finite type of serials the television schedule contain information about the range of the serial as well as the placement of the present programme within the serial in question. These data are provided in a parenthesis after the title. An example from Table 9.1 is ‘Den europæiske mand’ where (1:4) denotes that it is the first in a series of four. This data is directly extractable from the television schedule.

The description in the television schedule for programmes belonging to the ongoing type of serials is provided in two ways. For some programmes e.g., ‘Rebus’, episodes are still being produced. The number of programmes in the serial is therefore unknown, and the only number provided is the number of the programme in question. For ‘Rebus’ this number is 1. This type of serials is generally an ongoing fiction serial, in which episodes or programmes are still being produced, but which is expected to find
a closure within a foreseeable future. For the more regular serials like the news, magazine programmes and children’s programmes, the television schedule does not provide any information about the placement within the serial, except the date of transmission. However, this is most likely the kind of information needed by the users for this kind of television serials. In essence, the distinction made between serials in the information provided in the television schedule is in accordance with the users’ conception of serials in a television context. Therefore, we conclude that the television schedule generally provides the information needed for construction of the ‘Placement within a serial’ access point.

In addition, it is essential that relational links are created between the programmes in a serial, as we have discussed in Section 7.1.3.

9.1.6.2 Features
The television schedule does not consider broadcasts at the granularity level of features, and consequently it does not provide any information in order to construct the ‘Table of content’ access point. Considering the amount of manual resources needed for intellectual construction of this access point, it is concluded that the access point must be conducted by focusing on automatic indexing techniques. These approaches are highly related to the techniques for segmentation of broadcasts described in Section 9.1.1. In addition, it could be considered to include textual information from the Internet as external sources for construction of the table of feature content. An example is the textual information provided with DR’s Internet transmission of their news programmes (e.g., see http://www.dr.dk/DR1/TVAVISEN/tvaindslag.htm). These sources can also be used for extracting other access points at the features’ level of granularity. For instance, feature titles can be extracted from both DR and TV2’s Internet versions of their news broadcasts. However, the Internet information does not necessarily coincide with information extracted elsewhere, e.g., directly from the broadcast by applying video OCR. For instance, the title of a news feature in the Internet transmission does no necessarily coincide with the title in the eye catching ‘title text’ provided in the broadcast (e.g., at the bottom left side of the screen for DR1’s news broadcasts and in the upper left corner for TV2’s news broadcasts). The information may therefore collide with automatically extracted information of on-screen titles, e.g., by applying automatic video OCR tools (Lienhart, 2003). Consequently, if a hybrid of approaches is applied it is important that the algorithm designer decides on the weight of the information provided by each data source.
9.1.6.3 External

‘Audience ratings’ is a relation between the programme in question and an external entity, hence the categorisation. This information is not provided in the television schedule, but must be retrieved from tns Gallup’s TV-Meter system. That is, the system is to provide a few measures of impact e.g., rating, share, and reach, and provide direct links to the TV-Meter system for further analysis of audience ratings. In constructing this linkage it is important that programmes and searches conducted in one system can be transmitted to the other system, and that this transmission is perceived seamless by the users.

This ends our reporting on the value of the information available in television schedules for indexing of television broadcasts. Prior to providing concluding remarks in Section 9.1.8 the following section compares and discusses the 24 identified access points with access points available in existing surrogate records from operating archives or indexes.

9.1.7 Access points in existing IR systems for television broadcasts

Knowledge about the access points that are available in existing surrogate records is relevant in order to learn whether exchange of surrogate records is beneficial in relation to future broadcast retrieval systems. Here we do not attempt a complete analysis and discussion of the access points available in the different operating surrogate records, but we merely indicate the similarities between the different types of surrogate records. Two surrogate records from DR (see Appendix 21), one record from Centre for Educational Service at University College Jutland (see Appendix 22), two records from tns Gallup’s TV-Meter (see Appendix 23), and one record from the Registrant (see Appendix 24) are briefly analysed. DR indexes television programmes at different levels of description, and the two records from DR corresponds to a short and a long description, respectively. Two TV-Meter records are included in order to depict the TV-Meter indexing of broadcasts in the interim of programmes as well as the indexing of regular programmes. In Table 9.2 we map each of the six surrogate records from operating archives or indexes against the 24 access points which we have identified to be significant for the test participants’ retrieval of television broadcasts.

All six records contain ‘Title’, ‘Duration’, and ‘Date of transmission’. Information about ‘Channel’ is not included in the Registrant, but the information is implicit since DR is the only channel in the historical programmes covered by the index. Information about ‘Participant’ and ‘Author’ is found in the two records from DR and the record from the Registrant, while the educational service record contains information about author, only. Information about date of production is only found in
the TV-Meter records. The two production related bibliographic access points, ‘Production/responsible department’ and ‘Production country’ are found in the DR and TV-Meter records, and information about production country is also included in the Registrant. Information about the access points are not supported elsewhere, and exchange is therefore essential for facilitating searching and retrieval with such production information.

Table 9.2: Presence of identified access points in existing television surrogate records

<table>
<thead>
<tr>
<th>Access points</th>
<th>DR Short</th>
<th>DR Long</th>
<th>Centre for Educational Service</th>
<th>TV-Meter Programme</th>
<th>TV-Meter Trailer</th>
<th>The Registrant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participant</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>Author</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>Date of production</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>Production/responsible department</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>Production country</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Duration</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Channel</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>Date of transmission</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Time of transmission</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Transmission pattern</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>(X)</td>
<td>-</td>
</tr>
<tr>
<td>Programming</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Summary</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>Genre</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Keywords</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Spoken words</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clips</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Images</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Borrow options</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Waiting on loan</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Medium</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Placement within a serial</td>
<td>(X)</td>
<td>(X)</td>
<td>X</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
</tr>
<tr>
<td>Table of content</td>
<td>-</td>
<td>(X)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Audience ratings</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Legend: X denotes that the access point is available in the surrogate record, (X) denotes that the access point is partially available in the surrogate record, and - denotes that the access point is not available in the surrogate record.
Apart from ‘Channel’ and ‘Date of transmission’, the screening access points are not covered in the DR records and the Educational service record. The registrant covers ‘Time of transmission’, while TV-Meter covers ‘Time of transmission’ and ‘Transmission pattern’. ‘Programming’ is not cover by any of the surrogate records, but as mentioned earlier ‘Programming’ is a visual condensation of access points (e.g., time of transmission), and if those access points are available, the future broadcast retrieval system can provide the ‘Programming’.

Three of the six content access points are found in the six surrogate records. ‘Summary’ is available in DR’s two records, in the record from Centre for Education Services, and in the record from the Registrant. Indication of genre is found in the two TV-Meter records and the Registrant record. However, here it is important to emphasise that the TV-Meter genre indications originate from the television stations, and because these genre indications are used for construction of the television stations public service reports, they may be questionable. The interviewees with insight in the genre categorisations find TV-Meter’s genre indication questionable from a scientific point of view. ‘Keywords’ are found in all but the records from the Registrant, and in light of possible exchange of records the indexing policy for keywords is very important, including decisions as to exhaustivity and specificity. ‘Spoken words’, ‘Clips’ and ‘Images’ requires computational processing of digital versions of the broadcasts and since neither of the archives or indexes use automatic content based indexing broadcasts, the three access points are not found in any of the six surrogate records.

The three archival access points are concerned with data that is only relevant for the archive in question. Despite information about the medium of the broadcasts is available in the two DR records and the Centre of Educational Service record, the information is not exchangeable in a future broadcast retrieval system.

The ‘Placement within a serial’ relational access point is explicitly found in the record from Centre for Educational Service, and it is implicitly part of the title in the remaining surrogate records. ‘Table of content’ is not found in any of the records, but for the long DR record the specific description of images comes close to a table of content. ‘Audience ratings’ is concerned with relations outside the programme as such, and it is consequently not part of any of the records. In the TV-Meter system the audience ratings are found via relational links to the audience ratings database.

In brief, media internal access points are found in the different search systems for television broadcasts, though access points that require automatic content based techniques are not found in any of the six surrogate records. Media external access points (e.g., ‘Borrow options’) are generally not found in the surrogate records, and if
found they are not relevant for exchange since they are only relevant for the archive in question. In summary, it appears that the future exchange of records in a future broadcast retrieval context is mostly confined to the descriptive indexing or the isness of the television broadcasts. It is however, questionable whether extraction of subject indexing or content access points is obtainable, due to a difference in indexing policy, aim, and purpose. Different operational broadcast retrieval systems are constructed in order to facilitate a particular user group in retrieval of broadcasts, and hence the same content indexing may not be appropriate for two different groups of users with diverse aims, e.g., academics and journalists.

It is important to emphasise that corporation as to the storage of broadcasts in a central archive, does not necessarily mean that institutions or archives need to use the same surrogate records, since time codes can be applied to facilitate access to (different parts of) the broadcast, as we have discussed previously in Section 9.1.1.

9.1.8 Discussion and summarising statements

In this section, we briefly summarise our results regarding research question 5 or to what extent external textual and other sources contain or provide information valuable for constructing the 24 access points identified in the previous chapters. Table 9.3 summarises the value of information from textual television schedules with respect to each of the 24 access points.

We conclude that information in the television schedule is very valuable for construction of seven of the 24 access points (‘Title’, ‘Duration’, ‘Channel’, ‘Date of transmission’, ‘Time of transmission’, ‘Transmission pattern’, and ‘Placement within a serial’). For four access points (‘Participant’, ‘Author’, ‘Date of production’, and ‘Summary’) the television schedule is of some value. ‘Programming’ is dependent on information available in other access points, and (X) denotes the difficulty in judging the value of the information in the television schedule for construction of the ‘Programming’ access point. For nine access points (‘Production/responsible department’, ‘Production country’, ‘Genre’, ‘Keywords’, ‘Spoken words’, ‘Clips’, ‘Images’, ‘Table of content’, and ‘Audience ratings’) the television schedule does not explicitly provide information valuable in the construction of the access point. Though information derived from the schedule can be exploited to derive weights for algorithms in automatic multimodal indexing techniques. The three archival access points are not supported by information in the television schedule. This is however expectable as the information needed for construction of the access points are internal archival management data, hence the (X) in Table 9.3.
Table 9.3: Value of information in the television schedule for construction of broadcast access points.

<table>
<thead>
<tr>
<th>Category</th>
<th>Access point</th>
<th>Information in television schedule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Very valuable</td>
</tr>
<tr>
<td>Bibliographic</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Title</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Author</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Date of production</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Production/responsible department</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Production country</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>X</td>
</tr>
<tr>
<td>Screening</td>
<td>Channel</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Date of transmission</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Time of transmission</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Transmission pattern</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Programming</td>
<td>(X)</td>
</tr>
<tr>
<td>Content</td>
<td>Summary</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Genre</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Keywords</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Spoken words</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Clips</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Images</td>
<td>X</td>
</tr>
<tr>
<td>Archival</td>
<td>Borrow options</td>
<td>(X)</td>
</tr>
<tr>
<td></td>
<td>Waiting on loan</td>
<td>(X)</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>(X)</td>
</tr>
<tr>
<td>Relational</td>
<td>Placement within a serial</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Table of content</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Audience ratings</td>
<td>X</td>
</tr>
</tbody>
</table>

In brief, we conclude that descriptive indexing is partially supported by information from the television schedule, while the schedule provides very limited information for subject indexing. Further, from our discussion of previous research in Chapter 4, we know that subject indexing generally is the most difficult of the two. Furthermore, subject indexing or the content access points are the access points, which are most difficult to exchange with other archives or indexes, due to the difference in the users’ information seeking behaviour. For instance, the difference between the behaviour of Media Studies scientists compared to the behaviour of television professionals, e.g., journalists (Markkula & Sormunen, 2006). Therefore, a hybrid of approaches is recommended in order to construct all of the 24 access points identified to be important.
for television broadcast retrieval. This includes extraction of information from external sources (most notably the television schedules), manual intellectual indexing, and multimodal automatic indexing techniques, where information from other sources is exploited (e.g., for deriving weights for expectable outcome). If we are to point to specific automatic techniques to focus on in a future broadcast retrieval context, automatic video OCR tools would be beneficial for extraction of the text available in the broadcasts (e.g., roller captions), as well as automatic speech recognition techniques for translating the spoken words to computer-processable text.

It is important to emphasise that indexing should not solely be based on information which is extractable from television schedules, and this is especially the case in relation to subject indexing. The reason for this is manifold. First of all, the information needed for construction of all 24 appropriate access points is not available in the schedules as shown in Table 9.3. Secondly, the information available in the schedules is erratic, and indexing solely based on this information will hence result in broadcasts being indexed at different indexing levels, which again will entail different retrievability of the broadcasts in the archive. This is related to the third issue, namely that television schedules are political statements originating from the television stations. Solely relying on television schedules for indexing means that the retrievability of broadcasts indirectly is determined by the television stations. The television broadcasts that are given most attention by the television stations are the broadcasts which are most easily retrievable from the archive. The easy retrievable broadcasts are most likely to be the focus of the users, including the researchers who are supposed to conduct critical research about the role and influence of television in our society. This is to some extent interconnected to the Matthew effect in science as described by Merton (1968). Consequently, a hybrid approach which includes other perspective than the television schedules is preferable. These alternative perspectives are for instance the derivation of author aboutness by automatically processing of television broadcasts, as accomplish with automatic content-based indexing techniques (e.g., Hanjalic, 2004; Smeaton, 2004). Further, additional external sources should be considered when indexing broadcasts for a future broadcast retrieval system, for instance external sources from the Internet, e.g., from http://danskfilmogtv.dk/.

With reference to our discussion of existing records, an exchange of records with DR should be considered especially in connection to the bibliographic access points of ‘Participant’, ‘Author’, ‘Production/responsible department’ and’ Producing country’. Further, an exchange of records with DR and the Registrant should be considered in order to cover the historical parts of Danish television.
Finally, it is important to emphasise that further research is needed in order to understand the specific characteristics of the different types or genres of television broadcasts, including knowledge about the broadcasts that should be indexed at the features level of granularity.

This ends our reporting of the empirical investigation. In the following Chapter 10, we provide a summary of the thesis, including the empirical results, we present recommendations to future work, and state our final conclusion.
10 Summary and conclusions

Each of the nine preceding chapters contributes to the unity of the present doctoral thesis with reference to the formal presentation, theoretical rationalisation, and the empirical investigation. We explore aspects of scholars’ and students’ information seeking behaviour in the context of television broadcasts. Further, we derive implications for design of a future broadcast retrieval system by identifying the appropriate access points for broadcasts. Furthermore, we investigate the value of information available in textual television schedules, and discuss other approaches for indexing of broadcasts. As such, we fulfil the overall aim of the present work, which is concerned with identification of appropriate access points in television broadcast surrogate records.

This final chapter serves multiple purposes. Firstly, the chapter draws up the main objectives and findings of the doctoral work. The chapter (re)-presents the results found with respect to aspects of the test participants’ information seeking behaviour in the context of television broadcasts. Further, the chapter discusses the implications of the results in relation to design and indexing of broadcasts in surrogate records, for the purpose of facilitating effective retrieval of television broadcasts. Secondly, the chapter discusses and brings forth recommendations for future work in connection to securing the availability of television broadcasts in future broadcast retrieval systems. Thirdly, the chapter ends with final conclusions.

10.1 Summary of thesis objectives and results

The main objective of the thesis is to improve our knowledge in order to facilitate effective retrieval of television broadcasts. To reach this objective, we contribute with research in three research areas: 1) aspects of users’ information seeking behaviour in relation to television broadcasts (including information needs, preferred search entries, and relevance criteria); 2) appropriate access points for television broadcasts; and 3) construction of surrogate records for television broadcasts. The present work provides novel research in relation to aspects of users’ information seeking behaviour in the context of Danish television broadcasts, as well as the implications for construction of surrogate records. That is, implications for indexing of television broadcasts in relation to design and construction of metadata elements in surrogate records in future broadcast retrieval systems.
The main part of the thesis generates knowledge on the nature and characteristics of users’ information needs for television broadcasts, their preferred search entries, and the relevance criteria they apply when evaluating television broadcasts. Hereby, we expand our knowledge on these aspects of users’ information seeking behaviour in the context of television broadcasts. Secondly, the thesis applies the knowledge gained about aspects of the test participants’ seeking behaviour for discussion of the implications for design and construction of a future television broadcast retrieval system. We identify 24 novel as well as previously identified access points that are essential for facilitating the recognized behaviour. Further, we discuss how the identified access points might be constructed, with primary emphasis on the value of extracting information available in television schedules. In brief, the main contributions of the thesis are:

- Knowledge about characteristics of needs for television broadcast;
- The discovery of novel, and verification of previously identified access points preferred for searching and relevance assessment of television broadcasts;
- Assessment of the applicability of the identified access points for television broadcasts in relation to needs for television broadcast; and
- Assessment of the value of information available in television schedules for constructing the identified access points in surrogate records.

The present doctoral work is explorative in nature, and along the lines of naturalistic research (e.g., Lincoln & Guba, 1985; Mellon, 1990; Foster, 2005), the results do not claim generalisability beyond the group of users under scrutiny. Instead of generalisation, the present work ensures rich description to make sure that the research theme is transferable and open for further developments. As follows, the work provides knowledge to be considered the first instance of a Danish national broadcast retrieval system which is available for all (academic) users.

In Chapter 2 the integrated cognitive framework is introduced and our work is positioned within the framework. The cognitive viewpoint is concerned with human perception, interpretation, and transfer of information, as well as how acquisition of information is related to human information behaviour. A fundamental conception underlying the framework is that a parallel consideration of research in information seeking and IR is advantageous for research within LIS. The present work pursues the connection between knowledge learned about aspects of information seeking and implications for IR. Hereby, it provides empirical research within the integrated

272
cognitive framework for information seeking and IR (Ingwersen & Järvelin, 2005). We contribute to the framework by providing:

a) Theoretical understanding of aspects of information seeking behaviour in the context of television broadcasts;

b) Description and explanation of empirical investigation of aspects of information seeking behaviour in the context of television broadcasts; and

c) Support for development of a future broadcast retrieval system by identifying 24 appropriate access points for a future broadcast retrieval system, and discussion of the construction of the identified access points.

Thus, the present work follows the call by Ingwersen and Järvelin (2005, p. 3) to consider theoretical understanding, empirical explanation and technological development in a holistic way, and by this means provide an integrated understanding of the area under scrutiny. The framework stresses that underlying any purposeful information seeking behaviour is the inclination for solving a work task, whether leisure or job related. Chapter 3 continues this line of reasoning and starts off by providing an introduction and discussion of the developing information need. The information need is instigated by an individual’s recognition of some inadequacies for solving a particular work task. We propose an expansion of the concept of information need to incorporate process or task actions, and hereby our conception of the information need is related to Kuhlthau’s model of the information search process (e.g., Kuhlthau, 1991; 1993), and Vakkari’s work on association of task performance and information actions, including the need for and use of information types (e.g., Vakkari, 1999). Further, the previous research on general theories and models introduced and discussed in Chapter 3 (e.g., Dervin, 1983b; Ellis, 1987; Kuhlthau, 1991; Byström & Järvelin, 1995; Vakkari, 1999) show that work tasks are the instigating factor for any information seeking activity, as visualised by Ingwersen and Järvelin in the nested model of information seeking behaviour (see Figure 2.1). Chapter 3 provides the theoretical groundwork for the first research area on aspects of users’ information seeking behaviour in relation to television broadcasts. Consequently, the chapter pays special attention to research on information seeking behaviour in the humanities, and to the behaviour of users who are seeking (audio)visual information objects. The previous research shows that an important element in the humanities is the need to obtain information objects that have not been analysed previously, or at least not from the same perspective.

With reference to the last part of Chapter 3, we attach importance to the concept of relevance, since relevance criteria are central in relation to research question 3. The
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

concept of relevance is closely associated to the cognitive, psychological, and personally constructed information need. Relevance is considered to be personal and connected to a particular context and situation, at a specific moment in time, as expressed by Borlund (2000; 2003a). Relevance criteria are the criteria users apply in order to assess the relevancy of an information object in relation to the information need and work task at hand. Previous research shows that the visual aspect is important for retrieval of (audio)visual information object (e.g., Yang, 2005), and hence visual aspects are important to consider when designing and constructing (audio)visual retrieval systems. In the present work, the focus is given to the metadata elements users apply when assessing the relevancy of television broadcasts in relation to their information need and underlying work task.

The research on information seeking behaviour reviewed in Chapter 3 is important because knowledge about users’ behaviour provides indications as of how retrieval systems might help users cope with their information need. Such knowledge is a prerequisite for discussion of design and construction of future IR systems, including the indexing that is needed in order to construct appropriate access points. Chapter 4 focuses exactly on this indexing. That is, Chapter 4 reviews previous research regarding indexing, and in particular indexing of (moving) images. The chapter outlines four different types of attributes which are used for description of (moving) images in surrogate records. These attributes are: 1) isness attributes, 2) primitive attributes, 3) ofness attributes, and 4) abstract attributes, as summarised in Figure 4.1. The first type of attributes is connected to descriptive indexing, while the latter three are associated with different levels of interpretive freedom in subject indexing. Subject indexing is concerned with representation and/or condensation of the aboutness of the information object. The primitive attributes are concerned with the content as it is or the so-called author aboutness, while the ofness attributes and the abstract attributes are concerned with the content as perceived by an indexing device, whether human or computer, or the so-called indexer aboutness.

Two approaches for indexing of (moving) images are discussed in Chapter 4: 1) manual or so-called concept based indexing, and 2) automatic or content based indexing. The manual concept-based approach is the customary approach for indexing of (moving) images (e.g., Hertzum, 2003, p. 169; Smeaton, 2004, p. 380), and it is the approach applied in most operational (moving) images archives (e.g., Green & Klasén, 1993). The manual approach requires that the (moving) images are examined by a human being, and hence the approach is very resource demanding (e.g., Enser, 2000, p. 202). In contrast, automatic approaches decrease the resources needed for indexing, because the approach is entirely based on computational processing of the (moving)
image in question. The automatic approaches focuses primarily on the primitive content attributes of the (moving) images, or the content as it is. In brief, the distinction between the two approaches can be described as manual intellectual interpretation of high-level concepts, versus computational processing of low-level content features. With reference to the four types of attributes listed above, the automatic approach mainly conducts low-level processing in relation to primitive attributes, whereas the human processes information at a higher level of interpretation, which is referred to as the semantic gap between computational processing and human interpretation (e.g., Rasmussen, 1997, p. 183; Bimbo, 1999, p. 12; Chen & Rasmussen, 1999, p. 293; Bachimont, 2001, p. 1). Research on automatic approaches for indexing is still in its infancy, and the prevailing portion of IR systems based on automatic indexing are experimental systems (Hollink et al., 2004). Though in its early development, the automatic approaches have provided encouraging results (e.g. Smeaton, 2004; Snoek & Worring, 2005), and several researchers have pointed out that automatic approaches are to be considered a supplement to the traditional manual approaches for indexing of television broadcasts (e.g., Bimbo, 1999, p. 12; Rui, Huang & Chang, 1999, pp. 54-55; Smeaton, 2004, p. 382).

An alternative approach is to extract and utilise the information available in external sources. One of the first researchers to utilise external sources as a supplement for indexing is Small (1978). Small exploits the referential connection between scientific articles, to identify so-called standard concept symbols. These standard concept symbols are condensed representations of the citing authors’ perception of the content of the cited article, and in this way the concept symbols can be applied as descriptive keywords in surrogate records. Work by Schneider (2004) has demonstrated how concept symbols can be identified for the use of thesaurus candidate terms. External sources have also been applied for automatic indexing of images on the Internet (e.g., Dunlop & van Rijsbergen, 1993; Harmandas, Sanderson & Dunlop, 1997). The present doctoral work is inspired by such approaches utilising the information available in external sources, in that we consider how television schedules might be utilised for indexing of television broadcasts. Television schedules contain readily available text which might be utilised, whether the programme to be indexed is stored analogically or in digital form. With reference to the collection of television broadcasts at the State and University Library, this is a strong point, since a large portion of the collection is stored in analogical form (approximately 300,000 hours). The utilisation of external sources might circumvent the cumbersome tasks of either indexing the analogical part of the collection manually, or digitising the complete collection in order to use automatic indexing techniques.
The empirical part of the thesis starts up with Chapter 5 in which the empirical context is introduced. The purpose of the chapter is to provide background information for the methodical approach outlined in Chapter 6, and the results of the empirical investigation reported on in Chapters 7, 0, and 9. Chapter 5 presents knowledge about the present users of television broadcasts, because this knowledge is used as a rationale for selecting scholars and students in Media Studies as the user target group for our investigation. Further, the chapter provides an introduction to the academic field of Media Studies.

Chapter 6 provides an overview of the methodical approaches applied for the empirical investigation. The chapter elucidates the accordance between the epistemological stance in the cognitive viewpoint and the methodological stance in grounded theory (Glaser & Strauss, 1967; Strauss & Corbin, 1998). A key aspect in grounded theory is a flexible application of methodical approaches, though maintaining that the driving force in selection of methodical approaches is the overall aim of the research project (Strauss & Corbin, 1998, p. 46). In other words, selection of methodical approaches is based upon their ability to add knowledge for answering the five research questions put forward in Chapter 1. Along these lines, we apply an adaptable and competent application of methodical approaches. We conduct explorative investigations of aspects of users’ information seeking behaviour in the context of television broadcasts, and apply a triangulation of methods for gaining knowledge about television broadcasting and archiving, the users and use of archived television broadcasts, and the field of Media Studies. In the present work triangulation means that we are inspired by several qualitative and quantitative approaches, e.g., anthropological inspired participatory observation (e.g., Bogdan & Taylor, 1975), the micro-moment time-line interview technique (Dervin, 1983b; Schamber, 2000), and web questionnaire survey (e.g., Frankfort-Nachmias & Nachmias, 1996; Buckingham & Saunders, 2004). The underlying purpose of the present empirical research is exploratory in nature, and we believe that triangulation of methodical approaches is particularly valuable when conducting explorative research.

The greater part of the empirical data is collected by employing a web questionnaire survey and conducting nine in-depth interviews. The web questionnaire survey is mainly designed to collect quantitative data, while the in-depth interviews follow up with collection of qualitative data. The subsequent analysis of the collected data naturally reflects the nature of the data, and hence the web questionnaire data has chiefly been analysed by applying statistical tests of association, Chi\textsuperscript{2} and Fisher’s Exact test to be specific (Cramer, 1998, pp. 298-306). While transcription (Kvale, 1996), inductive content analysis (e.g., Allen & Reser, 1990; Schamber, 2000;
Neuendorf, 2002), and open and axial coding (Strauss & Corbin, 1998) is used for analysis of the interview data.

Chapters 7, 0, and 9 report on and discuss the results in relation to the three research areas, respectively. Chapter 7 focuses on our results in relation to the first research area concerned with aspects of the users’ information seeking behaviour in relation to television broadcasts. Research question 1 states:

What characterises a given group of users’ information needs in the context of television broadcasts?

The identified characteristics are manifold, multifaceted and interconnected. Firstly, the test participants’ need broadcasts to serve as data objects in scientific analyses, which is in line with previous research (e.g., Wiberley & Jones, 1994; Kirkegaard & Borlund, 2006). At first glance this characteristic might seem evident. Nonetheless, it is imperative, because it signifies a particular behaviour, e.g., the fact that broadcasts are needed in order to conduct active viewing in contrast to passive consumption.

Secondly, three broadcast dimensions are identified as being important for the test participants. These dimensions are: 1) the transmission dimension; 2) the archive dimension; and 3) the reception dimension. The test participants express a need to be able to see the broadcasts in their natural context, or the programming as we denoted it. The test participants find it imperative to be informed about the programmes surrounding the particular programme in question, as a relevance criterion. This entails that a future broadcast retrieval system should be able to provide descriptions of all transmitted broadcasts, whether the broadcasts have been stored or not. In brief, the test participants need to be informed about all transmitted broadcasts in order to understand the programming context of the stored and/or retrieved broadcasts. To facilitate the users’ fulfilment of this characteristic a future broadcast retrieval system must hold descriptions or surrogate records for broadcasts which are not retrievable, because they have not been stored. The construction of surrogate records for non-stored and non-retrievable information objects is fundamentally in conflict with the very purpose of an IR system, namely enabling the user to retrieve and be provided with access to information objects. However, the test participants’ needs are not (necessarily) concerned with retrieval of the television broadcasts, but rather the retrieval of information about the television broadcasts per se. The archival dimension refers to the fact that the test participants express a need to be informed about whether the broadcast in question has been stored for posterity, as well as the terms of availability of the
Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies

broadcast. The reception dimension refers to the fact that the test participants wish to be informed about audience ratings for the programme(s) in question.

Thirdly, in relation to Ingwersen and Järvelin’s (2005, pp. 291-292) generic types of information needs, the test participants broadcast needs concern known items, known topic or content, muddled topic or content, and factual data.

Fourthly, the test participants’ needs are characterised to consist of four parts: 1) getting an overview of transmitted broadcasts; 2) identification of borderline exemplars; 3) selection of specific programmes; and 4) verification of facts. This is connected to a wish for gaining knowledge about broadcasts and a wish for gaining access to broadcasts, and as such it intermingles with the second characteristic just mentioned.

Fifth, the test participants’ needs are concerned with all types of broadcasts, access is requested at the serials, programmes and features levels of granularity, and the characteristics of the different types or genres of broadcasts determines the level of granularity needed. This entails that a future broadcast retrieval system should enable the users to access all broadcasts at the programmes’ level of granularity and provide access at the serials and features level for relevant broadcasts, only.

Research question 2 and research question 3 focuses on search entries and relevance criteria, respectively, and they state:

Which search entries are preferred for searching of television broadcasts, and why and when are they preferred?

and

Which relevance criteria are applied when evaluating television broadcasts, and why and when are they applied?

Based on the knowledge learned by examining the questionnaires and analysing the interviews we are able to identify 28 search entries and 32 relevance criteria. The relative value of each search entry, and relevance criterion is compared against the four types of information needs identified when answering research question 1, and we show whether the search entries and relevance criteria are considered to be primary, secondary, or of no or low value for fulfilling each of the four types of information needs. Thus, we expand our knowledge regarding the first research area regarding aspects of users’ information seeking behaviour in relation to television broadcasts.

In brief, Chapter 7 provides valuable knowledge about the test participants’ information seeking behaviour, by providing information about the characteristics of the test participants’ broadcast needs, the preferred search entries for searching broadcasts, and the criteria applied when judging the relevancy of broadcasts in relation to the
information need at hand. Hereby, Chapter 7 contributes to our knowledge about aspects of users’ information seeking behaviour in general, and in the context of television broadcasts in particular.

In continuation of the knowledge learned about aspects of the test participants’ seeking behaviour, Chapter 0 identifies and discusses the access points that are appropriate for facilitating the identified behaviour. As such, Chapter 0 provides answers to research question 4, which states:

Which access points in a surrogate record are appropriate for television broadcasts, and what are their functions?

Our results show that 24 access points in five categories are appropriate in relation to a future broadcast retrieval system. These are: 1) seven bibliographic access points (‘Title’, ‘Participant’, ‘Author’, ‘Date of production’, ‘Production/responsible department’, ‘Production country’, and ‘Duration’); 2) five screenings access points (‘Channel’, ‘Date of transmission’, ‘Time of transmission’, ‘Transmission pattern’, and ‘Programming’); 3) six content access points (‘Summary’, ‘Genre’, ‘Keywords’, ‘Spoken words’, ‘Clips’, and ‘Images’); 4) three archival access points (‘Borrow options’, ‘Waiting on loan’, and ‘Medium’); and 5) three relational access points (‘Placement within a serial’, ‘Table of content’, and ‘Audience ratings’. The access points are related to the three dimensions of a future broadcast retrieval system mentioned in our reporting on research question 1. The five screening access points are related to the transmission dimension. The seven bibliographic access points, the six content access points, and the two relational access points ‘Placement within a serial’ and ‘Table of content’ are related to the archival dimension, while the latter relational access point (‘Audience ratings’) is related to the reception dimension.

The value of each of the 24 access points is assessed in relation to the four generic user tasks of find, identify, select, and obtain, which have been identified in a previous study (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998, pp. 8-9), and the four types of information needs identified in the present work. This assessment gives answers regarding the function of each of the 24 access points, and it provides recommendations concerning the access points to consider in relation to the design of a future broadcast retrieval system.

In the final empirical chapter, Chapter 9, we provide recommendations concerning the construction of each of the 24 appropriate access points identified in Chapter 0. This is done with a primary focus on the use of information available in television
schedules, and as such the chapter provides answers to the fifth and final research question stating:

To what extent does external textual and other sources contain or provide information valuable for constructing the identified access points?

Our examination shows that the television schedule contains valuable information for seven of the 24 access points. These are: 1) ‘Title’; 2) ‘Duration’; 3) ‘Channel’; 4) ‘Date of transmission’; 5) ‘Time of transmission’; 6) ‘Transmission pattern’; and 7) ‘Placement within a serial’. The information is of some value for the following five access points, ‘Participant’, ‘Author’, ‘Date of production’, ‘Programming’, and ‘Summary’. There is no information available in the television schedule for construction of the remaining 12 access points. In a few words, the content access points are the access points that are least supported by information in the television schedules. This entails that information in television schedules is generally applicable for descriptive indexing, only, and that content access points that require subject indexing must be constructed by other means.

In addition to a focus on the applicability of television schedules, Chapter 9 pays attention to other means of construction of the 24 appropriate access points. We mainly discuss how automatic indexing techniques can be applied as a supplement to the indexing conducted on the basis of television schedules. These automatic techniques are for instance concerned with automatic segmentation (e.g., Hauptmann & Witbrock, 1998; Bimbo, 1999; O’Connor et al., 2001), automatic people recognition (e.g., Satoh, Nakamura & Kanada, 1999; Li et al., 2001; Mohan, Papageorgiou & Poggio, 2001), automatic video OCR (e.g., Lienhart, 2003), automatic video categorisation (e.g., Dimitrova, Agnihotri & Wei, 2000; Ronfard, Garcia & Carrive, 2000), automatic speech recognition (e.g., Gauvin, Lamel & Adda, 2001; Barras et al., 2002; Evans, 2003; Bailer et al., 2005), and automatic summarisation (e.g., Mills, Cohen & Wong, 1992; Pfeiffer et al., 1996; Babaguchi, Kawai & Kitahashi, 2002; Hanjalic, 2004). Further, Chapter 9 discusses cursorily how records from operating archives and indexes might contribute to the construction of the 24 appropriate access points. Based on these discussions, we conclude that an exchange of records in relation to a future broadcast retrieval system is mainly beneficial in relation to descriptive indexing, and to a lesser extent in relation to subject indexing.
10.2 Discussion and recommendations to future work

In continuation of the summary of the results obtained, we provide comments regarding recommendations for future work, as well as discussion of what could have been done differently in the present work.

Overall, the methodical approach has been very successful for extracting knowledge about aspects of the test participants’ seeking behaviour. However, the test participants have difficulties envisioning the technological possibilities, and how these possibilities might affect or be exploited in future searching and retrieval of broadcasts. In particular, this is related to envisioning and articulating the ideal retrieval of television broadcasts. Further, the test participants are not explicitly conscious about the relevance criteria they apply, and hence they might have difficulties expressing them in web questionnaires and interviews. Based on the results achieved in the present investigation future investigation of broadcast relevance criteria could benefit from constructing and employing mock-ups of a possible future broadcast retrieval system. Even better, the construction of a test broadcast retrieval system, and subsequent studies, e.g., by applying simulated work task situations (Borlund, 2000, 2003b), log analysis, and eye tracking, would be a significant addition to the present doctoral work. Furthermore, the experience and knowledge gained throughout the research project, made us more skilled in conducting the interviews. With the benefit of hindsight, we could have been more explicit in some of our interview questions, and especially in relation to the first interviews. In a few instances, we were not persistent enough in our questioning. This means, that we are not able to give proper explanations for the ‘when’ and ‘why’ part in relation to each and every search entry and relevance criterion.

Our results show that several categorisations are necessary in the future work towards a broadcast retrieval system. A categorisation of genres or types of broadcasts needs to be established in order for the users to be able to search or evaluate the genre of retrieved broadcasts. The typology should be well-defined, general agreeable, and easily understandable. In the construction of the typology it is important to investigate the characteristics of each broadcast genre, because these characteristics are important in relation to indexing. For instance, our results show that the characteristics of the different broadcast genres imply the level of granularity for indexing of the broadcast, e.g., broadcasts with many features should be indexed at the features level of granularity. In relation to a future broadcast retrieval context, surrogate records should be constructed for serials of programmes, programmes, and features in programmes. Therefore future work should pay attention to investigation of characteristics of different types of broadcasts, as well as the specific indexing practices to apply for the
different types of broadcasts. Another relevant categorisation in future work is the categorisation of the roles of responsible persons and participants in broadcasts. Examples of roles include actor, camera man, script writer, anchor person, and interviewee.

In future work it would also be relevant to investigate how other types of external sources might be applied for construction of the surrogate records. The external sources in mind are in particular sources available on the Internet. Apart from the extensive sources available from the television stations own web pages, sources available from independent sources such as http://danskfilmogtv.dk/, should be considered. The reason for this is that the sources contain analysis of the content which could be applied for subject analysis, as well as descriptive indexing, e.g., explicit indication of participants, authors, and transmission pattern.

One of the results found in the present work is the test participants’ need to be informed about the transmitted broadcasts, and the so-called programming. In this light it should be considered to corporate with archives and indexes from other countries, e.g., TRILT from the UK. This is not least interesting in connection to the increasing diversity of foreign television channels available for the Danish television audience. The corporation with foreign archives and indexes would enable a user to gather information about a range of television channels, and to conduct comparative vertical programming analyses.

Prior to such international exchange of knowledge and records it is imperative that a national broadcast retrieval system is constructed. The results obtained in the present work are important in relation to this future work. However, we have only cut the first sod, and much work is still to be done. For instance, the results obtained focuses on scholars and students in Media Studies, and a natural continuation for future research would be to follow Wilson’s recommendations and test the obtained results in “[…] related but different settings” (Wilson, 1981, p. 11). These settings could be based on the knowledge gained in the analysis of the present users (Kirkegaard & Borlund, 2006), and it could for instance include users from social science departments, e.g., political science.

In recent years there has been an increased focus on providing access to the cultural heritage, both nationally, and in European and international contexts (see e.g., Mulrenin, 2002). This focus should be seen in relation to the ever growing influence of television in our society. In a national context the increased awareness has for instance provided 75 million DKK for digitalisation of DR’s historical archives (Lenler, 2007). However, digitalisation requires more than signed agreements, and the recent downgrading of DR’s Archive and Research Centre by cutting off approximate 50% of
Chapter 10: Summary and conclusions

the budget (Drost, 2007), could have severe consequences for accomplishing the digitalisation plan. Further, a digital cultural heritage is not necessarily an accessible cultural heritage. As discussed in the present work, a lot of work needs to be completed in order to facilitate effective retrieval of television broadcasts, and for this work to be done a lot of resources are needed.

Based on the knowledge learned in the present work, we recommend that a future national digital broadcast archive is constructed in corporation between the major players in the Danish media and archive picture, e.g., DR, TV2, and the State and University Library, and that the central archive is accessible by the relevant institutions. It is essential that the institution in charge of such a central national archive is independent from the television stations in order to avoid antagonism from the contributing institutions (e.g., other television stations), and for the simple reason that television stations perish while archives persist, as previously illustrated with the ‘Elkjær Channel’ (TVS). From our point of view, the State and University Library is the appropriate institution for the job, and future legislators should consider providing the necessary resources in order for the State and University Library to accomplish the task.

10.3 Concluding remarks

Our ambition to contribute to an integration of information seeking and IR in the context of television broadcasts is pursued in each of the preceding nine chapters, and we are able to add to previous research. The present research has extended our knowledge about aspects of users’ information seeking behaviour, as well as our knowledge about indexing of television broadcasts. We contributed to information seeking research by expanding the concept of information need to incorporate tasks or actions. We investigate three aspects of the test participants’ information seeking behaviour, namely information need characteristics, search entry preferences, and application of relevance criteria, and we show how the behaviour affects the construction of a future broadcast retrieval system. For instance, we found that the context of broadcasts is very important, and hence metadata elements containing contextual or programming information should be available in future surrogate records for television broadcasts. Indexing is always conducted from a particular point of view, whether the point of view is explicitly or implicitly stated. For this reason, indexing implies that users behave in a given way in order to search, retrieve, and/or assess the relevancy of the indexed information objects (e.g., Hjørland & Nielsen, 2001, pp. 249-250; Lancaster, 2003, pp. 8-9; Ingwersen & Järvelin, 2005, p. 36). This means that
knowledge about the users’ information needs and preferred access points for searching and relevance assessment in relation to television broadcasts is important in order to provide statements about the indexing of television broadcasts, which in brief explains the completeness of the thesis.

In presenting the objective of the doctoral work, we stated that the State and University Library wishes to gather knowledge concerning how a user can be supported in retrieval of television broadcasts, and how this support can be created with low utilization of resources. Our work provides knowledge to begin to answer these questions. We identify access points that are appropriate for supporting academics in their retrieval of broadcasts. Further, we identify how information from external sources can be exploited for future automatic construction of each of these access points.
References


Anderson, J.D. & Pérez-Carballo, J. (2001b). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part II: machine indexing,


References


References

Brookes, B.C. (1975a). The fundamental equation of information science. In: Information science, its scope, objects of research and problems (pp. 115-130). Moscow: [VINITI].


Mooers, C.N. (1960). Mooers’ Law or why some retrieval systems are used and others are not. *American Documentation, 11*(3), II.


Metadata elements preferred in searching and assessing relevance of archived television broadcast by scholars and students in media studies


References

Acoustics, Speech, and Signal Processing (ICASSP), (pp. 3001-3004). Phoenix, AZ, USA.

