Aalborg Universitet

AALBORG UNIVERSITY

Searching for Movies

An Exploratory Analysis of Movie-related Information Needs

Bogers, Toine

Published in:
Proceedings of iConference 2015

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Searching for Movies: An Exploratory Analysis of Movie-related Information Needs

Toine Bogers[1]

[1]Department of Communication & Psychology, Aalborg University Copenhagen, Copenhagen, Denmark

Abstract

Despite a surge in popularity of work on casual leisure search, some leisure domains are still relatively underrepresented. Movies are good example of such a domain, which is peculiar given the popularity of movie-centered websites and discovery services such as IMDB, RottenTomatoes, and Netflix. In this paper, we present an exploratory analysis of IMDB movie discussion threads that contain requests for movies to watch. Through emergent coding we produce a taxonomy of relevance aspects for movie search and selection. Our analysis shows that topical aspects, such as content, metadata, and known-item search, are important for movie selection practices. Other requests focus more on recommendation and feature many subjective relevance aspects, such as the tone of a movie or its intended audience. This suggests efficient access to movies is likely to require different information access paradigms to satisfy all the movie-related information needs expressed in the threads.

Keywords: movies, every-day life information behavior, relevance, recommendation, search, information seeking

Citation: Editor will add citation.

Copyright: Copyright is held by the authors.

Contact: toine@hum.aau.dk[1]

1   Introduction

Despite the popularity of movie websites like the Internet Movie DataBase (IMDB) and movie discovery services like Netflix, our understanding of how people search for and discover which movies to watch next is still relatively underdeveloped. While there has been a fair amount of work on searching for leisure in general (e.g, Savolainen (1995); McKenzie (2003); Elsweiler, Wilson, and Harvey (2012)), there has been no significant work—to the best of our knowledge—on movie-related information needs and information seeking behavior. For instance, we know little of the variety of movie-related information needs people experience or which relevance aspects are important to them when deciding which movies to watch. Searching for other types of media, such as books, has received a great deal of attention (e.g., Ross (1999) and Reuter (2007)), but it is unclear whether and to what degree those findings transfer to the movie domain.

We believe that a better understanding of how people search for movies and what makes a movie relevant to their information needs, is essential to better support people in their everyday information behavior. Different types of information needs are likely to require different information access methods to satisfy them correctly and a better understanding of real-world movie requests will be an important contribution to this.

In this paper, we aim to take the initial steps towards such understanding by collecting and analyzing a set of 3,581 movie discussion threads from the IMDB message boards. Our contributions are two-fold. First, we produce a taxonomy of relevance aspects for movie information needs through emergent coding on these movie requests. In addition, we perform an exploratory analysis of 400 annotated movie discussion threads to better quantify the distribution of different types of relevance aspects expressed. We find that movie relevance most commonly depends on content and metadata aspects, but that known-item needs also make up a significant part of the movie requests, suggesting a continuum of needs.

The rest of this paper is organized as follows. In the next section, we discuss the relevant related work, followed by a description of our methodology in Section 3. Section 4 describes our findings in more detail. We discuss our findings and their implications in Section 5.

2   Related Work

Searching for movies is a form of everyday-life information seeking (Savolainen, 1995) Information seeking for movie selection is typically a form of casual leisure search: searching for information with the goal of enjoyment and passing the time. A growing body of research on this special type of everyday-life information seeking has shown that people's

motivations, needs, and behavior are significantly different from, for instance, Web-based information seeking behavior (Elsweiler et al., 2012).

Movie selection behavior has been studied periodically over the years, though never from an IS perspective. For example, Austin (1979, 1985) examined the movie selection behavior of high school students from media science perspective. He reported that the movie plot was the single-most important reason for watching a movie, followed by starring actors and recommendations from friends.

In contrast to movies, book selection behavior in digital and physical libraries has received more significant attention in recent years. For example, Ross (1999) examined information behavior in the context of reading for pleasure. Through a study of 194 participants she found that book selection behavior is based on different aspects, such as mood, the subject, characters, author, the physical appearance, and recommendations from different sources, such as friends, bestseller lists and promotional campaigns. Reuter (2007) studied the book selection behavior of children in a digital library setting. She identified a total of 46 different influencing factors, grouped along seven dimensions: accessibility, content, engagement, familiarity, metadata & physical entity, novelty, and socio-cultural aspects. These dimensions overlap partially with those mentioned by Ross (1999) and partly influenced our own coding scheme as described in Section 3.2.

3   Methodology

3.1   Data collection

To investigate movie-related information needs we turn to the IMDB message boards. IMDB is one of the most popular movie websites and its message boards[1] contain over 1.19 million threads. For the work described in this paper, we were specifically interested in threads where users describe their movie-related information needs. Threads that cover movie news, reviews, or where users discussed movies, actors, director or other aspects were not of interest. We therefore focused our attention on two message boards in particular: "I need to know" (INTK)[2] and "Lists & recommendations" (L&R)[3]. We crawled all threads posted to these two message boards in June 2014, which resulted in 3,355 INTK threads and 226 L&R threads for a total number of 3,581 threads.

3.2   Coding

Our aim with collecting the IMDB threads was to analyze the movie-related information needs expressed in them. To this end, we performed a content analysis of the first message in each thread using emergent coding. We coded these messages for the different relevance aspects expressed by the original poster. Because of the imbalance in thread counts between INTK and L&R boards, we selected 60 random threads from each board to develop our coding scheme. After developing our coding scheme on these 120 threads, we ended up with 30 distinct relevance aspects. We then used card-sorting to arrange these 30 aspects into eight top-level aspects; the resulting taxonomy is shown in Figure 1.

In addition, we classified each thread as containing a request for a movie or not. Requests for actors, directors or other entities were not considered as movie requests. After developing our coding scheme we annotated 280 additional threads for a total of 400 annotated threads[4].

Some lower-level aspects might not be common enough to have appeared in our random sample of 120 threads. On the four occasions that we identified a new relevance aspect in a thread, we added them to our taxonomy. However, new aspects were only added to the lower level of our taxonomy; the top-level categories were kept intact.

---

[1] http://www.imdb.com/boards/?ref_=nv_cm_bd_1, last accessed November 24, 2014
[2] http://www.imdb.com/board/bd0000001/threads/, last accessed November 24, 2014
[3] http://www.imdb.com/board/bd0000122/threads/, last accessed November 24, 2014
[4] The message board threads and annotations described in this paper are available at http://toinebogers.com/?page_id=711

COMPARISONS	**Edition(s)**	Which edition of movie is the best?
	Lists	Complete lists of movies with a certain theme
	Novelty	Identify the first movie of its type
	Sequence & series	What is the proper order to watch a set of movies in?
CONTENT	**Dialogue**	Does the dialogue in the movie have a specific style?
	Plot, topic, or event	Movies that contain a specific plot element
	Style, setting & tone	Movies in a specific tone, such as film-noir
KNOWN ITEM	**Known-item**	Looking for a movie by describing its contents when you cannot remember its name
METADATA	**Audience**	Movies with a specific intended audience
	Genre	Movies that fit in a particular (set of) genre(s)
	Language & country	Foreign-language movies
	Release year	Movies released in a certain year
	Technical properties	Movies in recorded in black-and-white
	TV show(s)	Threads discussing TV show episodes instead of movies
PEOPLE	**Actor(s)**	Does the dialogue in the movie have a specific style?
	Director(s)	Movies that contain a specific plot element
	Writer(s)	Movies in a specific tone, such as film-noir
RECOMMEN-DATION	**More like this**	Looking for movies similar to the examples
	Not like this	Looking for movies that are not like the examples
	Pure recommendation	Recommendations for movies without an explicit description
SECONDARY PROPERTIES	**Awards**	Movies that (should) receive(d) a specific award
	Book version	Discussion about book vs. movie versions
	Credits	Movies with specific type of end credits
	DVD	Requests for movies with specific DVD features
	Popularity	Movies that are popular or relatively unknown
	Promotion & marketing	Threads about the promotion material such as movie posters
	Soundtrack	Discussion about the soundtrack of a movie
MISCEL-LANEOUS	**External link**	Thread containing a link to a resource not on the message boards
	Spam	Spam thread
	Other	Any other type of discussion topic not covered by the other categories

Figure 1: Overview of the eight top-level and 30 lower-level relevance aspects in our aspect taxonomy, together with common examples of how these aspects are expressed in the thread.

4   Results & Analysis

Of the 400 IMDB annotated message board threads, 275 (68.8%) contained requests for movies to watch. Many of these requests are highly complex, providing details about required plot elements, setting, tone, as well as providing examples of similar movies or types of movies the user is not looking for. Topic #226897022 is an example of such a complex request:

> *Movies where a disaster occurs or is about to occur and strangers band together to help each other out. Especially something based on true life events would be nice. Or not even a group of strangers, just two, where one has an accident or something and a stranger passing by has to do something till help arrives. Something like that. TV movies and lesser known movies would be good. Not end of the world type movies though, something like 'World Trade Center' or 'Unstoppable' maybe.*

Here, the user requests a specific type of disaster movie (Genre) that does not depict end-of-the-world events. Instead, the user is looking for movies with the specific plot element of strangers banding together (Plot, topic or event) and provides the other message board users with examples of similar movies (More like this). It demonstrates that users can have complex movie-related information needs. Investigating such requests allows us to move beyond traditional query log analysis towards a better understanding of such complex needs.

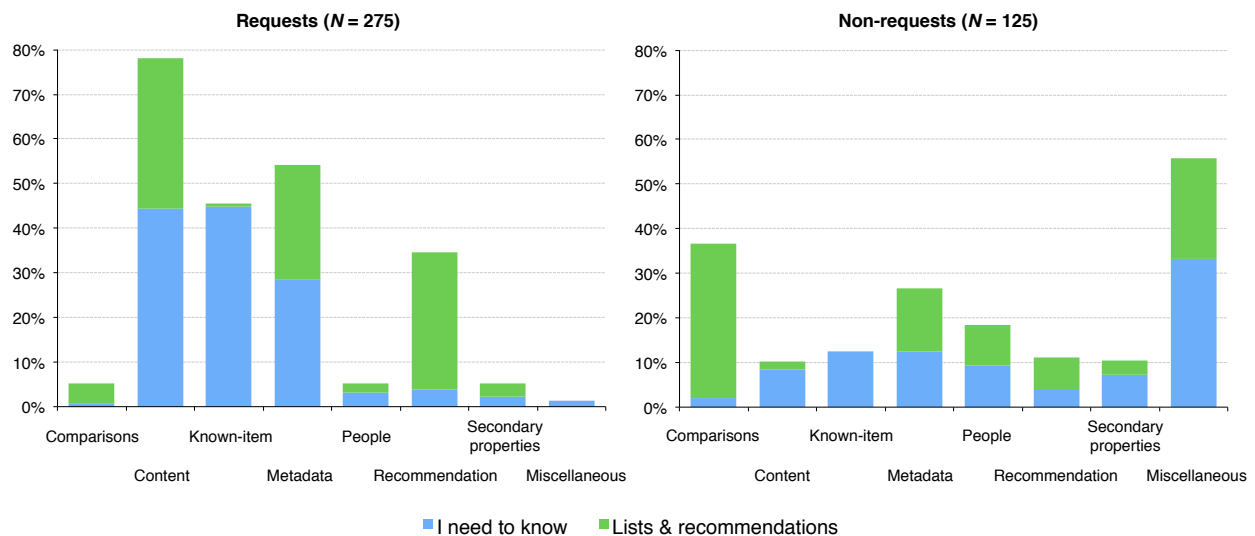**Distribution of relevance aspects**

Figure 2: Distribution of top-level relevance aspects for the two message boards, split by movie requests vs. non-requests. The height of each bar represent the percentage of all movie threads that were annotated with that aspect. Each bar is made up of the proportions of that relevance aspect's occurrence in the two message boards.

Figure 2 shows the distribution of the eight top-level relevance aspects for the 275 threads containing a request for movies and the 125 non-requests, split by message board. It shows that the Content and Metadata aspects are the most common relevance aspects expressed in the movie requests. Users most often mention Plot, topic, or event (49.5%), followed by Genre (22.5%), Release year (15.5%), and Style, setting & tone (15.0%), which is line with the findings of Austin (1979).

Another important relevance aspect is Known-item search: in 46.0% of all requests users describe a movie they could not name, but remembered specific plot details of. Finally, Recommendation is also common at 34.6% of all movie requests. Users often ask for pure recommendations based on their past preferences, and in 19.6% of all cases provide positive and/or negative examples of movies that are similar to what they are looking for. These examples move the request away from topical relevance towards a more recommendation-oriented approach.

Perhaps surprisingly, the people responsible for creating a movie, such as actors, directors, and writers, are not that important when searching for movies to watch. Only 5.5% of all requests were centered around the People relevance

aspect. Discussing movie crew members without explicitly looking for new movies to watch is more common in the non-request threads at 18.3%. This suggests that looking for movies featuring a particular actor or director is straightforward enough to be handled by IMDB's own search and browsing functionality. Presenting and discussing personal top-N lists of movies (Lists) and linking to external websites (External link) are also common in the non-request threads.

5   Discussion & Conclusions

A better understanding of how people search for movies and when they are relevant to their information needs is essential to better support people in their everyday information behavior. Different types of information needs require different information access methods to satisfy them correctly and a better understanding of real-world movie requests will contribute to this. The work described in this paper represents an essential first step.

Our exploratory analysis suggests that in movie selection Content, Metadata, and Known-item are the most common aspects. These aspects are close to the traditional concept of topical relevance, which suggests using a data collection based on the metadata and descriptions available on IMDB—such as the IMDB collection used in the 2010 INEX Data-centric Track[5]—could potentially be used to satisfy many of these needs. However, many movie requests appear to go beyond just topical relevance and include other, more subjective requirements, such as the tone or style of the movie, their popularity or intended audience. In addition, many requests have a strong focus on recommendation, asking other IMDB users to take the user's preferences or provided example movies into account as well.

This suggest that different information access paradigms are likely required to satisfy all the information needs expressed in the threads. Information retrieval algorithms seem like the preferred approach for requests that focus on topical relevance, whereas other requests may be better addressed using a recommender system. It also raises an additional question: Is there really such a clear-cut distinction between search and recommendation, or does a grey area exist in between these two approaches that is not sufficiently addressed by either? More work is needed to provide an answer to this question.

For future work, we would like to investigate the degree to which requests lean towards search or recommendation paradigms, or a combination of both. In addition, we would also like to examine the suggestions made by other users in response to these requests. Do some requests elicit more suggestions than others and how well do they appear to satisfy the user's information need? Finally, we would also like to investigate whether the suggested movies can be used as relevance judgments, similar to the use of the LibraryThing forums in the INEX Social Book Search track[6] (Koolen, Kamps, & Kazai, 2012). This would enable automatic evaluation of the performance of search and recommendation algorithms on the movie requests.

References

Austin, B. A. (1979). Motion Picture Attendance and Factors Influencing Movie Selection among High School Students. In *Proceedings of the 33rd Annual University Film Association Conference.*

Austin, B. A. (1985). Motivations for Movie Attendance. In *Proceedings of the 76th Annual Meeting of the Eastern Communication Association* (pp. 3–34).

Elsweiler, D., Wilson, M. L., & Harvey, M. (Eds.). (2012). *Proceedings of the "Searching 4 Fun!" Workshop, collocated with the annual European Conference on Information Retrieval (ECIR 2012).*

Koolen, M., Kamps, J., & Kazai, G. (2012). Social Book Search: Comparing Topical Relevance Judgements and Book Suggestions for Evaluation. In *CIKM '12: Proceedings of the 21st International Conference on Information and Knowledge Management* (pp. 185–194).

McKenzie, P. J. (2003). A Model of Information Practices in Accounts of Everyday-Life Information Seeking. *Journal of Documentation*, *59*(1), 19–40.

Reuter, K. (2007). Assessing Aesthetic Relevance: Children's Book Selection in a Digital Library. *Journal of the American Society for Information Science and Technology*, *58*(12), 1745–1763.

Ross, C. S. (1999). Finding Without Seeking: The Information Encounter in the Context of Reading for Pleasure. *Information Processing & Management*, *35*(6), 783–799.

Savolainen, R. (1995). Everyday Life Information Seeking: Approaching Information Seeking in the Context of 'Way of Life'. *Library & Information Science Research*, *17*, 259–294.

---

[5]Available at https://inex.mmci.uni-saarland.de/, last accessed November 24, 2014

[6]Available at http://social-book-search.humanities.uva.nl/, last accessed November 24, 2014.

Table of Figures