Tags on healthcare information websites

A theatre of the absurde

Ådland, Marit Kristine

Publication date:
2020

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

TAGS ON HEALTHCARE INFORMATION WEBSITES

A THEATRE OF THE ABSURD

BY
MARIT KRISTINE ÅDLAND

DISSERTATION SUBMITTED 2020

AALBORG UNIVERSITY
DENMARK

TAGS ON HEALTHCARE INFORMATION WEBSITES

A THEATRE OF THE ABSURD

by

Marit Kristine Ådland

AALBORG UNIVERSITY
DENMARK

Dissertation submitted

Dissertation submitted: July, 2020

PhD supervisor: Professor Marianne Lykke
OsloMet and Aalborg University

PhD committee: Associate Professor Tanja Svarre Jonasen
Aalborg Universitet (chair)

Professor Gunilla Widen
Åbo Academy

Associate Professor Haakon Lund
Københavns Universitet

PhD Series: Faculty of Humanities, Aalborg University

ISSN (online): 2246-123X
ISBN (online): 978-87-7210-505-5

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2020

ENGLISH SUMMARY

This thesis explores tags and tagging behaviour on health information websites using an empirical, user-oriented, exploratory case study. The purpose is to find out more about tags and tagging behaviour on a health information website.

Method: Data were collected in diverse ways, in order to obtain a view on the research questions from different angels. In a preliminary study, I analysed tags from Blogomkraeft.dk and compared them to the site structure of Cancer.dk. After the launch of the tagging feature on Cancer.dk, a study of user behaviour on Cancer.dk was conducted, with a focus on the role of tags. Participants solved tasks using the newly launched tagging feature, they filled out pre- and post-test questionnaires, and I interviewed them. When the tagging feature had been live for about a year, I also interviewed three editors about their experience and opinions on tags and tagging.

To study the tags themselves, a transaction log containing more than 25,000 tags was analysed, mainly analysed them through categorizing the tags into categories: (1) Internal and external tags, (2) Lay or professional tags, (3) Topical facets and (4) Aboutness.

Results: For taggers, the analysis indicate a connection between computer skills, an understanding of the tagging feature, and a focus on applying tags as topical descriptors.

Topical description was dominant when applying tags at Cancer.dk. Some of the taggers stated that they wanted tags to be exclusively topically descriptive. Participants who did not apply topical descriptive tags all agreed that such tags could be useful. A focus on subject description was often connected to a focus on finding information. To the participants, the topical tags did not have to describe the topic of the article; it was enough that that it described a subsection or an aspect of its topic. Thus, topical tags did not equal subject headings. Their requirements were not as strict.

Other purposes found were tags to explain the content, tags to evaluate articles, and tags to express requests for additional information. These tags represented attempts to communicate with the system, its users, or editors. All the participants agreed that topical tags were good, but they did not agree on whether other types of tags added value to Cancer.dk.

The different purposes users had when they applied tags was a challenge for the editors. From the interviews, it is my impression that the editors in a way did not want tags, but a controlled vocabulary. This would fulfil some of the purposes that both editors and users had when they applied tags. A subject language that includes

synonyms and possibly includes relations between terms (e.g. hierarchical) would give lead-in terms that users and editors need. Such a solution is, however, contradicted by the editors' view that tags are mainly the user's voice in the system. A controlled vocabulary can never replace this, which the editors were also clear about.

Internal and external taggers behaved differently. The internal taggers were from inside the organization behind Cancer.dk. It was easy to address them as a group and encourage them to apply tags. However, the crowd of external taggers were more stable.

Analysis of the log files showed how difficult it is to apply tags. The aboutness categorization reveals challenges in how tags relate to the topical content of the article. Mixed together, the tags as a whole were difficult to use and difficult to judge.

Tags from internal taggers covered categories that were more diverse and described the article content from various angles. Their tags were more evenly distributed on tag facets compared to external taggers. These results conflicted with the expectation that external users can add new viewpoints to the systems. External taggers, however, applied more tags not related to the content of articles.

The results can inform the design of tagging features; visibility is essential to attract tags, and also influence the characteristics of tags slightly. The information surrounding tagging needs testing. The disagreement between the user groups can also inform tagging features: the tags applied within the system will be influenced by who has permission to apply tags.

The communicative aspects of tags found in Cancer.dk indicate that taggers do not necessarily distinguish between tags in different systems. When looking at systems like Twitter, tags are communicative by intent; they add information to the tweet and do not necessarily cover the topical content of the tweet. However, when moved to an information website, this behaviour is unwelcome.

DANSK RESUME

Denne afhandling udforsker tags og adfærd knyttet til tagging på websider med helseinformation gennem et empirisk, brugerorienteret og eksplorativt case studie. Formålet er at finde ud af mere omkring tags og tagging-adfærd på en helse-webside.

Metode: Data er blevet indsamlet på forskellige måder for at kaste lys på forskningsspørgsmålene ud fra forskellige vinkler. I et forudgående studie analyserede jeg tags fra Blogomkraeft.dk og sammenlignede dem med site-strukturen på Cancer.dk. Efter lanceringen af tagge-funktionen på Cancer.dk gennemførte jeg en undersøgelse af brugeradfærd på Cancer.dk med fokus på betydningen af tags. Deltagere løste opgaver via brug af den ny-lancerede tagge-funktion; de udfyldte før- og efter-spørgeskemaer, og deltog i interviews. Da tagge-funktionen havde været i drift i ca. et år, interviewede jeg også redaktørerne omkring deres erfaringer og holdninger til tags og tagging.

Med henblik på at undersøge selve taggene analyserede jeg transaktionslogs indeholdende over 25.000 tags. De blev hovedsagelig analyseret via opdeling i 4 kategorier: 1) interne og eksterne tags, 2) lege- eller professionelle tags, 3) emnefacetter, og 4) tags omhandlende emnet i artiklerne.

Resultat: Analysen indikerer en forbindelse mellem computer-kompetence, forståelse af tagging-funktionen og fokus på anvendelse af tags som emnedeskriptorer.

Emnebeskrivelse var dominerende i forhold til brugen af tags på Cancer.dk. Nogle af taggerne udtrykte ønske om brug af tags udelukkende som emnedeskriptorer. Deltagere som ikke brugte emne-beskrivende tags var alle enige om at sådanne tags kunne være brugbare. Et fokus på emnebeskrivelse blev ofte forbundet med fokus på informationssøgning. Deltagerne havde ikke brug for tags til at beskrive artiklernes indhold; det var tilstrækkeligt at de beskrev en undersektion eller et aspekt af emnet. Tags var derfor ikke sammenlignelige med overskrifter – kravene til deres dækning var ikke så omfattende.

Andre formål synliggjort af analysen var: tags til forklaring af indhold, tags til evaluering af artikler, og tags til at udtrykke anmodninger om yderligere information. Disse tags repræsenterede forsøg på at kommunikere med systemet, dets brugere eller redaktører. Alle deltagere var enige om at emne-tags var brugbare, men de var uenige om hvorvidt andre typer af tags kunne tilføje værdi til Cancer.dk.

De forskellige ønsker som brugerne havde til brugen af tags var en udfordring for redaktørerne. På baggrund af de gennemførte interviews var tydeligt at redaktørerne egentlig ikke ønskede tags men nærmere et kontrolleret ordvalg. Dette ville opfylde nogle af de formål som både redaktører og brugere havde med brugen af tags. Et

emnesprog som inkluderer synonymer og eventuelt også relationer mellem termer (f.eks. hierarki) ville resultere i de "forord" som brugere og redaktører havde brug for. Den løsning bliver dog modsagt af redaktørernes opfattelse at tags primært er brugerens "stemme" i systemet. Et kontrolleret ordbrug kan aldrig erstatte dette, hvilket redaktørerne også var klar over.

Interne og eksterne taggere opførte sig forskelligt. De interne taggere var tilknyttet organisationen bag Cancer.dk. Det var let at adressere dem som gruppe og opmuntre dem til at bruge tags, hvorimod de eksterne taggere var mere stabile.

Analyse af logfilerne viste hvor svært det er at anvende tags. Kategorien 'tags omhandlende emnet i artiklerne' afslører udfordringer i forhold til hvordan tags relaterer til emneindholdet af artiklen. Sammenblandingen af tags var vanskelige at anvende og værdien vanskelig at bedømme.

Tags fra interne taggere dækkede kategorier, som var mere forskelligartede og beskrev artikelindholdet fra forskellige vinkler. Deres anvendelse af tags dækkede flere facetter sammenlignet med tags fra eksterne taggere. Disse resultater var i konflikt med forventningen at eksterne brugere kan tilføje nye vinkler til systemerne. Dog anvendte eksterne taggere flere tags, som ikke var relateret til artiklernes direkte indhold.

Resultaterne kan give indspil til design af tagging-funktioner; synlighed er essentiel for at tiltrække anvendelsen af tags, og påvirker også taggenes egenskaber Informationer omkring tagging bør testes. Uenigheden mellem vores brugergrupper kan også tilføje ny viden omkring tagging-funktioner: de tags der bruges inden for systemet vil blive influeret af hvem, der har adgang til at bruge tags.

De kommunikative aspekter af tags fundet på Cancer.dk indikerer at taggere ikke nødvendigvis skelner mellem tags i forskellige systemer. Hvis vi kigger på systemer såsom Twitter er tags anvendt kommunikativt som hensigt; de tilføjer information til tweet'et og dækker ikke nødvendigvis hele tweet'ets emne – hvorimod denne anvendelse af tags ikke er velkommen på informationswebsider.

NORSK SAMANDRAG

Denne avhandlinga utforskar taggar og oppførsel knytt til tagging av helseinformasjon knytt til nettsider. Dette er gjort i form av ein empirisk, brukarorientert, utforskande casestudie. Føremålet er å finna ut meir om taggar og oppførsel knytt til tagging på ei nettside om helseinformasjon.

Metode: Me samla inn data på ulike måtar, for å kasta lys over problemstillingane frå ulike vinklar. I ei førebuande undersøking analyserte eg taggar på Blogomkraeft.dk og samanlikna dei med site-strukturen på Cancer.dk. Etter at taggefunksjonen på Cancer.dk var lansert, gjennomførte me ei undersøking av brukaroppførselen på Cancer.dk, med fokus på kva rolle taggar har. Deltakarar løyste oppgåver med den nylanserte taggefunksjonen, dei fylde ut spørjeskjema før og etter dette, og eg intervjua dei. Når taggefunksjonen hadde vore i drift i omtrent eit år, intervjua eg også redaktørane om deira erfaringar og meiningar om taggar og tagging.

For å undersøka sjølve taggane, analyserte me transaksjonsloggen med meir enn 25.000 taggar. Me analyserte dei i hovudsak gjennom å kategorisera dei på fire måtar: (1) Interne eller eksterne taggar, (2) Leke eller profesjonelle taggar, (3) Emnefasettar og (4) Om taggane handlar om det same som artiklane.

Resultat: Tor taggarar, indikerer analyse nein samanheng mellom datakompetanse og forståing av taggefunksjonen, og eit fokus på taggar som emnebeskriving.

Emnebeskriving var dominerande når brukarane tagga på Cancer.dk. Nokre av taggarane sa at dei syntes taggar kun burde vera emnebeskrivande. Deltakarar som ikkje la inn emnebeskrivande taggar var samde i at slike taggar kunne vera nyttige. Eit fokus på emenbeskriving vart ofte knytt til eit fokus på å finna informasjon. For deltakarane trengde ikkje emnebeskrivande taggar å handla om emnet for artikkelen; det var tilstrekkeleg at den handla om ein del eller eit aspekt av artikkelen sitt emne. Emnebeskrivande taggar er altså ikkje det same som emneord. Brukarane sine krav var ikkje så strenge.

Andre føremål for tagging var: taggar for å forklara innhald, taggar for å evaluera artiklar, og taggar for å be om informasjon. Desse taggane representerer forsøk på å kommunisera med systemet, brukarane, eller redaktørane. Alle deltakarane var samnde om at emnebeskrivande taggar er gode, men dei var ikkje samde om i kva grad andre typar taggar kunne ha verdi på Cancer.dk.

Dei ulike føremåla taggarar hadde når dei tagga, var utfordrande for redaktørane. Frå intervjua, er det inntrykket at redaktørane ikkje eigentleg ikkje ville ha taggar, men eit kontrollert vokabular. Dette ville oppfila nokre av føremåla som både redaktørar

og brukarar hadde når dei tagga. Eit emnespråk som inkluderer synonym og kanskje også relasjonar mellom termar (t.d. hierarkiske) ville gjeve inngangstermar som brukarar og redaktørar har bruk for. Ei slik løysing er likevel i strid med redaktørane sitt syn på at taggar i hovudsak er ei brukarstemme inn i systemet. Eit kontrollert vokabular kan aldri erstatta dette, det var redaktørane også klare på.

Interne og eksterne taggarar oppførte seg ulikt. Dei interne taggarane har tilknytning til organisasjonen bak Cancer.dk. Det var lett å nå dei som gruppe og oppmuntra dei til å tagga. Men massen av eksterne brukarar var meir stabil.

Analysen av loggfila viser kor vanskeleg det er å tagga. Kategoriseringa av korleis taggar er knytt til emneinnhaldet av artiklane avslører utfordringar. Sidan taggane er blanda saman, er det vanskeleg å bruka og vanskeleg å bedømma verdien av taggane samla sett.

Taggar frå interne taggarar dekka kategoriar som var meir ulike og beskreiv artikkelinnhaldet frå ulike vinklar. Takkane deia var jamnare distribuert på ulike fasettar samanlikna med eksterne taggarar. Desse resultata er i konflikt med ei forventing om at eksterne brukarar kan koma med nye synspunkt inn i systement. Men eksterne taggarar la inn fleire taggar som ikkje reflekterte innhaldet i artiklane.

Resultata kan gje innspel ved utforming av taggefunksjonar; for å tiltrekka seg taggar er det essensielt at funksjonen er synleg, eigenskapane ved taggane vert også litt påverka av kor synleg funskjenen er. Informasjonen som følgjer funksjonen må testast ut. Usemja mellom brukargrupene vår kan også gje innspel om taggefunksjonar: kva taggar som vert lagt inn i eit system vert påverka av kven som har tilgang til å tagga.

Den kommunikative sida ved taggar som er funne i Cancer.dk indikerer at taggarar ikkje nødvendigvis skil mellom taggar i ulike system. Ser me til system som Twitter, er taggar ofte med hensikt kommunikative; dei er eit tillegg til twittermeldinga heller enn å dekka emneinnhaldet i meldinga. Så er det berre slik at når denne oppførselen vert flytta til ei informasjonsside, er den uønska.

TAKK

Tagging er på ein måte eit samabeidsprosjekt. Det er denne avhandlinga også, berre på ei annan måte. Eg har gjennom heile arbeidet hatt god hjelp med nyttige innspel og oppmuntring frå vegleiar Marinne Lykke. Ho var også den som trekte meg inn i samarbeid med Lois Delcambre og Jeremy Steinhauer, begge med innspel og vinklingar eg ikkje kunne funne sjølv.

Gjennom arbeidet har eg også hatt god støtte frå tidlegare og noverande intstituttleiarar, Liv Gjestrum og Tor Arne Dahl. Dei veit når dei skal spørja korleis det går, og når dei skal la det vera. Det er ein god ting.

Eg arbeider i eit kollegium ved Institutt for arkiv-, bibliotek- og informasjonfag der det alltid er nokon å drøfta noko med, ta ein pause med, eller eta lunsj med. Det er også ein god ting, takk til alle.

Takk også for hygge og forstyrringar frå dei heime, David, Leif Sebastian og Jon Arthur.

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

In this thesis, I report findings from my Ph.D. project, where I have examined the notion of tags and tagging. The project is an exploratory case study, where a tagging feature was implemented to an information website, Cancer.dk[1]. This is the official website of the Danish Cancer Society. In cooperation with the Cancer Society and especially the editors behind the website, I studied how end users and employees of the Cancer Society applied tags to documents, how they used the tags when searching and browsing, and their opinion about the tags and tagging. In this setting, documents are units identified by a URL on Cancer.dk, mainly short articles with text and illustrations. The aim was to find characteristics of tags and tagging behaviour on an information web page, and to describe tagging from the users' and editors' points of view.

The project is a longitudinal user-centred study where separate studies were conducted over time, from 2009 to 2013. The studies represent diverse aspects of the tagging and users' interactions with the tagging feature. I was a part of the Cancer.dk tagging project from the beginning and cooperated with the Cancer Society on the design of the tagging feature. I then observed the feature and the use of it from its launch and evaluated it.

1.1 MOTIVATION

Tags are words or phrases that users of a system apply to documents available in this system. The tags thus represent a users' perspective on the documents and aspects of documents. Users write tags with their own vocabulary, as opposed to the professional vocabulary of experts and information intermediaries. This give tags valuable properties when end users search and browse for information (Peters, 2009; Quintarelli, 2005). On the other hand, tags can sometimes be inaccurate and imprecise as descriptions of documents (Guy & Tonkin, 2006; Thomas et al., 2010). It is thus relevant to explore the usability of tags for searching and browsing.

Health information is a field of knowledge with a well-known gap between professionals and laypersons, also known as the vocabulary problem (Zeng et al., 2001, 2002). It is also a field where it is important for laypersons to have easy access to information, and health professionals want to reach out with information. I wanted to explore the usefulness of tags as a part of an effort to overcome the vocabulary problem in the field of health informatics.

[1] www.cancer.dk

Cancer is also a field where laypersons differ in their familiarity with the field. Newly diagnosed patients often have sparse information about their illness and its treatment, while experienced cancer patients often are experts on both their type of cancer and their own reaction to it. This diversity made it interesting to explore whether tagging could be useful for instance as a tool for experienced patients to help novice patients.

During the last 15 years or so, researchers have studied tags in various systems. In systems like Delicious[2], users apply tags as a part of a process to include a document into their collection of documents. The system therefore exposes users to the metadata of the document, like URL and title, when they apply tags. I find it interesting to study tags in a system where metadata is not as visible to the users, and where the user participation is not a main goal. Cancer.dk is an information website "in the business of providing information" (Kalbach, 2007). Research on tagging behaviour in such a setting is sparse.

My contact with the Danish Cancer Society started in 2009. I was looking for a collection of domain specific documents with engaged and dedicated users. Experience from health professionals showed that cancer patients were willing to engage in research and willing to help and inform each other. I found it therefore meaningful to use Cancer.dk as a case when studying tags and tagging.

1.2 RESEARCH QUESTIONS

The purpose of the Ph.D. project is to explore tags and tagging on an information website, as part of users' information behaviour: how the aboutness, the meanings or topical content of tags relates to document content, users' opinions on tags, and what language is present in tags. I have chosen Cancer.dk as a case and thus an information website intended for cancer patients and their relatives. With a tagging feature on Cancer.dk, both users and the editors of the information website are important players.

More specifically, I seek to answer the following research questions:

1. What characterizes tags on Cancer.dk?
    1.1. Are tags characterized by lay or professional vocabulary?
    1.2. What is the topical content of the tags and how do tags relate to the aboutness of the documents?
2. What role did the tagging feature on Cancer.dk itself play?
3. What are the users' and editors' view and behaviour towards tags?

[2] www.delicious.com

3.1. What are the purposes of users when applying tags?

The first research question (1, including sub-questions 1.1-1.2) evaluate the tags and their aboutness. The second research question (2) focus on the tagging feature itself and its influence on the tags, mirrored to theoretical views on what tags could be and on literature that show tags from other settings and systems.

The third research questions (3) explore the users' and editors' views on tags. The aim is to analyse how these views correspond to the properties of tags, and thus find out about correspondence between tags and views on tags.

The findings add to our knowledge on tags. Knowing the users' views on tags also gives knowledge on how users relate to indexing and descriptions of documents in general. In this project, I study certain tagging features on a certain website within a certain subject, cancer information. An aim for this research is to find out how tagging features on information websites in domain specific environments can be set up in a meaningful way. Cancer.dk is a site where professionals inform the public about cancer. This is a setting that differ from systems like Delicious, LibraryThing or YouTube. Thus, it is not trivial to introduce a tagging feature to such a setting.

1.3 RESEARCH TEAM AND TEAMWORK

I conducted this research as a part of the FIRE project – *Facilitating information retrieval for experts*. FIRE was formed based on a previous project where semantic sections of texts were indexed as a supplement to basic indexing methods (Price et al., 2009). The semantic components were "segments of text about a particular aspect of the main topic of the document and may not correspond to structural elements in the document" (Price et al., 2007, p. 429). This gave improved retrieval effectiveness. It was also indicated that it was easier to obtain higher indexing accuracy with semantic component indexing (Price, 2007; Price et al., 2009, 2007).

The good results, however, depended on time-consuming manual indexing. In the FIRE project, we sought methods to make the process simpler, easier, and thus more cost-effective. We found it interesting to test whether social tagging could be a way to do this. Our plan was to introduce a tagging feature where users could apply tags to both documents and parts of documents and see if it was possible to achieve good results from semantic component indexing when end users did the indexing by applying tags.

Four researchers participated in the FIRE project:

- Professor Lois Delcambre, Portland University
- Ph.D. student Jeremy Steinhauer, Portland University

- Professor Marianne Lykke, Aalborg University
- Ph.D. student Marit Kristine Ådland, Oslo Metropolitan University (previously Oslo and Akershus University College of Applied Sciences)

In the project, we worked together on data collection and overall goals, and exchanged ideas. Our cooperation with the Danish Cancer Society was also a part of our joint effort. The two Ph.D. students in the research team had separate research questions. In my work, I have focused on tags and their meaning, and on views on tags and the process of applying tags. My share of the project is anchored in information science. I have studied the nature of tags, their topics, language and facets, and how they describe and provide access to documents. The other Ph.D. student in FIRE, Jeremy Steinhauer, based his work in computer science. He focused on algorithms and retrieval (Steinhauer et al., 2013, 2011).

Whenever I use "we" in this thesis, I refer to the FIRE team or members of the FIRE team, unless something else is specified. An exceptions is in the discussion of Goffman's model on social settings, where it is natural sometimes to use "we" and mean: "all of us", "humans".

1.4 FUNDING

The tagging feature was payed for by the Danish Cancer Society and the FIRE project through NSF Grant No 0812260. It ended up being more expensive than expected, so we could not correct all errors due to lack of money. This also meant that a longer process with further changes in the feature was impossible. Oslo Metropolitan Univertity funded my share in this project. The university also paid for an assistant who helped categorize tags.

1.5 THE STRUCTURE OF THIS THESIS

The next chapter (chapter 2) gives an overall view on the research design and methods used, with methodological considerations and an emphasis on data and data collection. Chapter 3 presents the theoretical framework that I use to analyse and interpret research data. Chapter 0 is an overview of the research literature I have examined to find a starting point for what was already known, with respect to my research questions. I refer briefly to preliminary studies that was conducted before the FIRE team finalized the plan for the remaining studies in chapter 5. Then, in chapter 6, I describe the tagging feature that was set up on Cancer.dk for the preliminary study.

This feature framed the data collection for the main studies, reported in chapters 7-9. These studies are:

7 Cancer.dk tags and their usage
8 How patients apply tags to cancer information
9 Editors' view on tags and tagging

In chapter 9, I analyse and discuss all findings, and give conclusions.

2 METHODOLOGY

2.1 GENERAL RESEARCH APPROACH

When looking at tags, and on user behaviour, one can evaluate and decide what is useful, valuable, and thus correct or incorrect according to a specific use of tags. The users may or may not see or even understand this specific use of tags, and thus may or may not consider adapting to it. If users try to adapt, they may or may not succeed. This view on tags follow Fugman's basic assumption on indexing that subjects *can* be defined and described, and that indexing creates *order* (1993). Thus, one can say something verifiable about the usefulness of tags in specific practical settings. This implies a cognitive view on tags (Jensen, 2011). The tag represents the taggers thoughts about the document. Tags can then be counted and categorized, assuming that they refer to the document in a way that can be explained and understood. One can evaluate and decide what is useful, valuable, and thus correct or incorrect due to a specific use of tags.

Other researchers have tried to explain tags, based on the tags themselves. Munk and Mørk (2007) studied tags in Delicious. They observed tags that were "wrong" in the way that they did not represent the topical content or aboutness of the tagged document. They give two explanations: First, the tag is not wrong, but represent *why* the tagger is interested in the document, not the actual aboutness of the document. For example, a document about communism could have the term *capitalism* as a tag because it names why the document is interesting to the user. Reading about communism can be interesting if you want to learn about capitalism.

Their second explanation is laziness. The tag is "wrong" because the tagger did not bother to read the document or figure out its aboutness. In a system like Delicios, this makes sense. Its purpose is to give users a chance to bookmark web pages and access the bookmarks from different computers. Many users' bookmarks are lists of documents they plan to read. When Delicios ask them to apply a tag, they write something without thinking it through. Then, they plan to return to the document later, if they find time.

Each tag originates from an individual, this needs to be considered when studying tags. From a cognitive point of view, the tag represent the taggers thoughts about the document (Jensen, 2011).

On the other hand, I explore tags with Goffman's model on how people present themselves to others like actors on a theatre stage, as one way to explore tags and tagging (Goffman, 1959). This implies more of a constructive viewpoint where the tag is a result of the taggers' construction of meaning when reading or browsing the document (Jensen, 2011). From this point of view, tags are correct or at least well

intended from the taggers' point of view, whether they are correct, valuable or usable according to regular indexing standards. To see tags as correct or well intended also has a practical cause in this project. I do not intend to change the users or challenge them. Instead, I want to observe their behaviour and then explore the usability of the observed tags and the system to which they belong.

This mix of approaches catches a conflict inherent in tags: (1) Tags are metadata, or data, from the users, and there are in general no strict rules about how to formulate tags. They represent a variety of users and users' construction of meaning of documents. (2) At the same time, systems owners want to use tags as more or less controlled metadata. Implicitly, they will evaluate tags for retrieval purposes and often try to extract high quality tags. This follows Fugman's basic assumptions, and a more cognitive viewpoint on tags (Fugmann, 1993).

A mix between cognitive and constructive viewpoints is not new. Jensen refers to how this has happened in linguistics and in research on reading (Jensen, 2011). Here cognitive view focus on the structure and meaning of language, while a constructivist view focus on the utility and use of language. In these cases, one has found a need for both viewpoints in order to give a sufficient description and explanation of the empirical findings in the respective fields.

In my work, information retrieval and subject indexing hold a cognitive view, with a focus on how terms are good or bad depending on their ability to be a part of a system that provide relevant documents to the users. At the same time, the social setting and individual use of the system hold a constructive view in the sense that users have different expectations and purposes when using the system. This can be related to their role as users.

2.1.1 INDIVIDUALS AND GROUPS

To find out more about tagging, one can study the tags themselves. To understand tags, one should also study the users who apply the tags, the taggers, how they apply tags and how they use the tags. This gives a broader picture. For tagging as a social phenomenon, the completed aggregated folksonomy is interesting. To know what is really going on, the individual tags and the individual behaviour that causes tags are interesting, as well as types of tags and use of tags. The micro-level for the folksonomy is tags, taggers and tagged documents. In addition to this, the system influences the folksonomy. This system also goes back to individuals: systems owners and editors. Their individual and negotiated choices form a structure that secondly will influence how people tag.

Especially broad folksonomies (see chapter 3.4) may form digital societies of taggers. However, when explaining the folksonomy, there is a need to break it down to individual tags and individual tagging behaviour, without excluding the social aspects of tagging.

One example is that individuals find influence within the system. When applying tags, available words and information visible in the system may give them ideas. This happened in a research project where people were asked to apply tags to documents (Golub et al., 2009). The tagging interface for some documents had extra information related to the content of each document, from two different controlled vocabularies. In both cases, the tags showed that the vocabularies influenced taggers in their choice of words to apply as tags. When there was no vocabulary available, the taggers often picked words form the document itself: the title, subheadings etc.

Taggers can also browse other users' tags, and thus relate to other taggers, learn from others and get ideas from other taggers. This is a possible explanation to Golder and Huberman observation that the choice of words used as tags applied to documents in Delicious stabilized over time. When many taggers apply tags to the same document in systems with a broad folksonomy, a stable pattern emerge (Golder & Huberman, 2006). One could imagine that such patterns come from taggers learning from one another. On the other side, it is also possible that such patterns are a result of individual independent tagging. People do not always communicate before ending up with the same result.

Altogether, the individuals are always a part of a context, which I see as the reason why Goffman saw the team as a "fundamental point of reference" (see chapter 3.6.1.2) (Goffman, 1959, p. 85). Both the individual and the collective view are important when studying tags.

2.1.2 CASE STUDY

Yin states that "a case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident." (2003, p. 13, punctations left out). Hyett, Kenny and Dickson-Swift give emphasis to other features stating that: "Case study research is an investigation and analysis of a single or collective case, intended to capture the complexity of the object of study" (Hyett et al., 2014, p. 2. They credit Stake, 1995 for this statement). In the present project, I study tags and tagging, and the context is the information website Cancer.dk. The research strategy implies varied data that describe and reveal the case from different angels. The data listed in chapter 2.3 reflect this variation.

A case study of a specific tagging feature is a way to examine individual tags, users and documents in the context from where they originate. Such a case will include most of the conditions that these individual tags, taggers and documents within the system share. A tagging feature with its content include both the micro-level of tagging behaviour, but also groups of individual users, tags and documents that share common features and interact as teams or groups. Flyvbjerg states that "the advantage of the case study is that it can "close in" on real-life situations and test views directly in relation to phenomena as they unfold in practice." (2006, p. 235).

The case study also gives a limitation necessary for research: it is impossible to study an endless number of the instances of interest. But when selecting a variety of instances within a case, the case also includes "the nuanced view of reality, including the view that human behaviour cannot be meaningfully understood as simply the rule-governed acts found at the lowest levels of the learning process and in much theory."(Flyvbjerg, 2006, p. 222). In my case, all instances of tags were selected, three editors were selected, and a few other users.

Research on tags and tagging is often limited to one or a few systems, taking into consideration that systems differ. Thus, research on tags always has an aspect of a case study, meaning results are always deeply connected to the system from where the tags originate. Many studies on tags use data from Delicious[3], LibraryThing[4] and other large systems. When choosing a case that differs from these systems, I intend to broaden the view on tags: Properties that are common for tags in a variety of systems can be assumed general for tags.

An important property that separate Cancer.dk from the systems mentioned above, is the lack of focus on metadata. In for instance Delicious, users apply tags as a part of a process to include a document into their collection of documents. The system therefore exposes users to the metadata of the document, like URL and title, when they apply tags. I find it interesting to study tags in a system where metadata is not as visible to the users, and where the user participation is not a main goal.

Thus, this is a *critical case study*, meaning that the case was selected "on the grounds that it will allow a better understanding of the circumstances in which the hypothesis will and will not hold" (Bryman, 2012, p. 70). Previous research on tags and tagging constitute hypotheses about this field (see review in chapter 0), and a critical chosen case can broaden the view on tags. In addition to the fact that Cancer.dk do not highlight metadata to the users, the site itself is well organized. Tags in this setting do not play the role as structuring elements on the site. They serve as an addition to the site, in interaction with other metadata on the site.

Since tagging features differ, one cannot generalize from one or a few systems to all systems. Instead, one can describe the properties of a case, compare these properties to other cases, and then say something about what to expect from tags and taggers, based on the context of the given system.

With a case study, one can also falsify hypotheses on tags. If tags on Cancer.dk differ from tags in more frequently studied systems, this falsifies hypothesis that these properties of tags are general for tags. This brings us closer to a robust view on what tags can be. The outcome of a case study like this cannot bring us closer to statistical

[3] www.del.icio.us

[4] www.librarything.com

generalized knowledge on tags, but seeks to "expand and generalize theories (analytical generalization)" about tags (Yin, 2014, p. 21).

The results may thus not apply directly to other tagging features or other information websites. But the thoroughly description given in this thesis, gives an opportunity to compare with other settings, however with care:

- The tagging feature directs the tagging behaviour. Thus, similar tagging features may lead to similar tagging behaviour. I studied a narrow folksonomy in the preliminary study, and Cancer.dk had an extended narrow folksonomy. The tagging behaviour here is comparable to tagging behaviour on other narrow and extended narrow folksonomies.
- Cancer.dk gives information about cancer and about how to prevent and treat cancer, and thus targets users who want this information. This kind of domain specific content is typical of many web information sites. Tags applied to this content can be compared to tags applied to content in other domains.
- Cancer.dk has a group of editors, and users who in general are professionals and laypersons. This is true for many information websites. Thus, the behaviour and opinions of these user groups are comparable to other information websites that cover other topics.

I selected Cancer.dk as a case for this research project for many reasons; some of them already mentioned here and in the Introduction (chapter 1). Experience from health professionals showed that cancer patients were willing to engage in research and willing to help and inform each other. There was also a willingness among the editors to test tagging on Cancer.dk, to find out whether it would add value to the site.

2.2 RESEARCH PROCESS AND OVERVIEW OF SURVEYS

This is a user-oriented (Järvelin & Ingwersen, 2010) empirical study where I studied users' interaction with tags and tagging, and information behaviour in general on a certain website, Cancer.dk. I studied a diversified material and thus had a chance to study and describe the case from different angels. In a user-oriented study, the user is seen as a part of the system (Järvelin & Ingwersen, 2010). With tags, the user also provide metadata into the system, and thus user behaviour is crucial in understanding the whole system.

2.2.1 CONTACT AND COOPERATION WITH THE DANISH CANCER SOCIETY

Tor Øyan was our contact person in the Danish Cancer Society, as the chief editor of Cancer.dk. After the initial contact, the FIRE team had a workshop with representatives from Cancer.dk and their vendors of content management system and search engine, respectively ProActive and Ankiro. We shared our thoughts about tagging and discussed how to introduce tagging to Cancer.dk.

A period of prototype building and testing followed the workshop. This included the preliminary studies reported in chapter 5, where I analysed Blogomkraeft.dk tags and compared them with the browsing structure of Cancer.dk. I also conducted a *usability study* based on a tagging feature prototype. After this, ProActive produced a finalized tagging feature for Cancer.dk, including a logging feature.

Øyan and his co-editors wanted there to be tags in the system from the beginning. Thus, the first week only employees of the Danish Cancer Society had access to, and were encouraged to apply tags. Then Øyan reported to be live with the tagging feature available to all users. He wrote, "It looks as if the first external tag was; selleribøf = celerysteak". This tag has a timestamp valued: "2011-11-30 10:36:49.297" in the transaction log.

2.2.2 MIXED METHODS

To study the tags gives an insight into what people actually do. In this case, the activity was high, but it is not obvious what really happened: Why and how did users apply tags? The interviewed users in this study are too few to give a general view on taggers purposes when they apply tags. A single tag can also serve several purposes. Thus, one cannot always expect to make conclusions on the purposes of a single tag, even if I had a complete list of possible purposes users have when they apply tags at Cancer.dk. On the other hand, the quantitative and the qualitative data do shed light on each other's, which is why I chose these varied methods in the first place.

A combination of qualitative and quantitative methods give holistic view and thus a clearer picture of the case Cancer.dk (Bergman, 2008). For short, the quantitative data, from log files, give data on what is going on, and make it possible to find out whether a phenomenon is frequent or not. In this thesis, I have counted tags of various types, and then compared the number of tags in different categories. On the other hand, with the qualitative data I wanted to find explanations. For instance, when I see tags as "wrong" according to my own or the editors' expectations, or according to indexing standards, the qualitative interviews can give explanations to what role such tags can play.

The following chapters (2.2.3-2.2.6) give short descriptions the various studies conducted for this thesis.

2.2.3 PRELIMINARY STUDIES

The purpose of the preliminary studies was to explore whether and how social tagging could support user interaction and information retrieval on an information website like Cancer.dk, and how to implement social tagging in a way that supported this purpose.

2.2.3.1 Blogomkraeft.dk tags and Cancer.dk site structure

Blogomkraeft.dk was a blog site and a part of the Cancer.dk web. The tags applied to postings formed an extended narrow folksonomy (see chapter 3.4). There were 650 tags in total, and 344 unique tags, applied to 318 blog postings. The blog tags originated from selected users, but still users in the target group of Cancer.dk, and it covered the field of cancer and cancer treatment. Because of these similarities between Blogomkraeft.dk and a future tagging feature on Cancer.dk, I used Blogomkraeft.dk as an indication on what to expect from Cancer.dk. The content of the blog tags on Blogomkraeft.dk were analysed by categorizing all tags according to their meaning.

Cancer.dk also had a site structure, available as a sitemap on the site. The items in the structure can be compared to terms in a controlled vocabulary, where every menu item and link anchor gives information about the aboutness of the connected article. I compared the two, assuming that the result would indicate what tags could add to Cancer.dk. If it showed that tags only repeated the structure and metadata that was already there, this would indicate that there was no need for tags at all.

I did not compare individual tags and sitemap items but grouped them into categories based on the aboutness of each tag and site structure item. The number of tags in different categories gives a good picture of the important features of the collection of tags. The conclusion was that, based on experience from Blogomkraeft.dk and the sitemap of Cancer.dk, that tags have potential to support user Cancer.dk. For more details, see chapter 5.2 and Ådland & Lykke (2012).

2.2.3.2 Usability study

From the study of Blogomkraeft.dk and Cancer.dk site structure, I learned that a tagging feature could be useful on Cancer.dk. A usability study was set up to find out how to implement a tagging feature, using a prototype developed after the workshop and discussions about such a feature.

The usability test was conducted in June 2010. Five participants used the prototype. I observed them and communicated with them during the test. Pre- and post-test questionnaires gave data about personal background, Internet experience and the participants' understanding of and opinions about the prototype. With five participants, I see all data, both in the prototype sessions and the questionnaires as qualitative data. Together, this gave an impression on users' views on and opinions on tags.

Our participants were able to operate the feature and liked its design and functionality. Thus, when implementing tagging at Cancer.dk, there should not be big changes compared to this prototype. For more details, see chapter 5.3 and Ådland & Lykke (2012).

2.2.4 TAGGER STUDY

After the launch of the tagging feature on Cancer.dk, conducted an empirical study of user behaviour on Cancer.dk, with a focus on the role of tags. Like in the preliminary usability study, I collected data in diverse ways, in order to obtain a view on the research questions from different angels. This time, eight participants solved tasks using the newly launched tagging feature on Cancer.dk, they filled out pre- and post-test questionnaires, and I interviewed them when the tasks were completed. The goal was to find out more about users' thoughts when faced with a tagging feature and tags: their understanding and opinions of tags and tagging, and their purposes when applying tags. The study resulted in qualitative data, structured through questionnaires and semi-structured interviews. For details, see chapter 8.

2.2.5 EDITOR STUDY

When the tagging feature had been live for about a year, I interviewed three editors about their experience and opinions on tags and tagging. The interview was semi-structured. I asked questions about background and experience in their job, but we spent most of the time talking about their opinions about and experience with tags and the tagging feature on Cancer.dk. The goal was to find out about their opinions, and to be able to compare this to the users' opinions. Both the tagger study and the editor study gave a better understanding of the tags logged for the tag study (below). For details, see chapter 9

2.2.6 TAG STUDY

The transaction log is the largest data material in this project, and I have focused on the tags. I took account of all tags in this study. Many researchers use most frequent tags or do other types of selection before they conduct their studies. Examples are Pera, Lund and Ng (2009), Munk and Mørk (2007), and Morrison (2008), who all study selections of tags. This can be a good thing, but in this project, I found it interesting to study the whole collection of tags, to obtain a complete picture of what tags can be like.

The log file includes more than 25,000 tags, a huge amount of tags for a small system like Cancer.dk. As regard functionality and interface design the tagging feature remained the same for the whole period, but the location and visability at the cancer.dk website changed in September 2012. Thus, the collected data were produced in two

slightly different settings, with the tagging feature less visible for the users in the second setting and time period. These changes influenced the tags that users applied.

I mainly analysed these quantitative data through categorizing the tags into categories:

1. Internal and external tags – indication on who applied the tag
2. Lay or professional – do the tag belong to a lay or professional vocabulary
3. Topical facets – what is the tag about
4. Aboutness – the relationship between the aboutness of the tag and the aboutness of the article

I use these categories to explore what types of tags the users applied to Cancer.dk. Together with the qualitative data, it was also possible to explain tags and tag categories. I can also indicate how extensive phenomenas found in the qualitative data are, based on the quantitative data. For details, see chapter 7.

2.3 DATA

To sum up, the total data material includes:

- Tags:
    - Tags extracted from Blogomkraeft.dk, autumn 2011
    - Transaction log data of tags applied on Cancer.dk and internal search terms, November 2011-February 2013
- Tagging behaviour and interaction behaviour:
    - Observation of and questionnaire from participants in usability study
    - Transaction log data of tagging behaviour and interaction behaviour, November 2011-September 2012, and Septermber 2012-February 2013. This includes tags applied by the participants in the tagger study in December 2011
    - Interviews with users from tagger study
    - Interviews with editors
- Information about background and web experience:
    - Questionnaires from usability study

- Questionnaires from tagger study

All together, I observed or interviewed 16 participants: Five persons participated in the usability study, eight in the study of taggers, and I interviewed three editors. The participants gave valuable information about how to understand tagging behaviour and general behaviour on Cancer.dk. The interviews gave qualitative data and were analysed as such.

3 THEORETICAL FRAMEWORK

The aim of this chapter is to introduce central concepts needed to understand tagging, and to provide a theoretical framework for discussing the properties, use and opinions on tags. Models and theoretical considerations include theory on information retrieval, information interaction and interaction in social settings. They all shed light on the research questions and form a base to understand the data collected in this project.

I refer to indexing theory (Lancaster, 2003a; Svenonius, 2000) and use the concepts aboutness, warrant and relevance from indexing theory when analysing the tags. Aboutness is used to analyse and understand how users relate to the topical content of aritcles when applying tags. I use warrant to find what perspective the users have on articles, and which words they use to formulate their tags. And relevance is used to find the relationship between the topics of tags and the articles to which they are applied.

Indexing theory is also in the background when looking at the usefulness of tags. Indexing in general, and particularly subject indexing, have similarities with tags and tagging, as tags are descriptions from the users. Thus, indexing theory is included here in the theoretical framework.

Both applying tags and using tags is connected to information interaction, how we interact with information. Information interaction is a process where users browse or navigate from one piece of information to another, and adjust their behaviour based on what they find. If a tagging feature is a part of the information-rich environment with which they interact, users will deceide wether to use the feature or not, and how to use it, as part of their process. Thus, I have included the concept of information interaction here (Toms, 2002). In this thesis, the interaction takes place on Cancer.dk, which can be characterixed as an information website. This is also the location for the tags, and thus the concept of information websites is gives context for the tags. The aim is to see if and how users include tags in their information interaction.

The information interaction is supplemented with information retrieval, in order to analyse whether tags fit with the needs of a user that search for information.

Tags and folksonomies (Hunter, 2009; Munk & Mørk, 2007) are in the core of what this thesis is about: their usefulness, their purposes, and how to characterize them otherwise. Comparisons with subject headings is relevant, but also other ways of characeterizing tags that can include non-topical tags, such as tag funtions or purposes (see chapter 4.3.3.1).

Tags and tagging are part of an information interaction, but it also has broader social aspects that can be useful when trying to explain the tags and their purpose. Thus, I

have included the concept of performance in social settings (Goffman, 1959) as a part of the theoretical framework.

I have seeked to find models and concepts that together can characterize tags and tagging. There is no way to sum up the tags I have studied and say: "this is what you can expect from tags in any system". This is in the core of why this is a case study. But the characteristics of tags on Cancer.dk add to the knowledge on tags and bring another view on what tags can be, in tagging features with similarities to the one on Cancer.dk. The theoretical framework here gives a basis for doing this.

3.1 INFOMATION INDEXING AND RETRIEVAL – INDEXING THEORY

Lancaster defines subject indexing (and abstracting) like this:

> *Subject indexing and abstracting […] involve preparing a* representation *of the subject matter of documents. […] The indexer describes [documents'] contents by using one or several index terms, often selected from some form of controlled vocabulary*. (2003b, p. 6, italics in original)

Lancaster states that subject indexing terms can "indicate what the document is about" or "summarize its content". They also "serve as access points through which an item can be located and retrieved in a subject search".

Ingwersen and Järvelin have a similar definition of indexing in general: "Text indexing is a process that creates a short description of the content of the original text." (2005, p. 130). Like Lancaster, they continue: "The result is a representation of the text. […]." (2005, p. 130). Chowdhury uses different words but gives more or less the same definition: "The process of constructing document surrogates by assigning identifiers to text items is known as indexing. When the task of indexing is based on the conceptual analysis of the subject of the documents, it is called subject indexing." (2010, p. 77).

Subject indexing and indexing in general have a clear role: when users search a database, the index terms represent the document. The retrieval system then "match the contents of documents with users' queries" (Chowdhury, 2010, p. 77). The index terms also inform the user about the document. Thus, the user can use index terms to find out whether to retrieve and/or read the whole document or not. Tagging can be compared to indexing; tags are representations of the text. Researchers often compare tags to subject headings, the part of indexing that deals with the aboutness of documents (Heymann & Garcia-Molina, 2009; Kipp, 2005; Spiteri, 2009; Wetterstrom, 2008). This can be meaningful, but tags do not necessarily represent the topical content of the text. Tags may also represent other properties of the documents or properties of how taggers relate to the documents. Information retrieval and indexing is included here, because searching and matching between documents and

user need is important when users interact with an information website like Cancer.dk. I need to analyse whether tags fit with the needs of a user that search for information.

3.1.1 ABOUTNESS

When deceiding how to describe the topical content of a document, the concept of aboutness is useful. It can be defined as *what a document is about.* Hutchins discussed and defined this concept from 1975 (Hutchins, 1975). One of his conclusions was:

> for the purposes of information systems the 'aboutness' of a document may be defined in terms of those parts of its generalised semantic network that relate the document to the context of the assumed 'states of knowledge' (Hutchins, 1977, p. 34)

Hutchin identifies a semantic network in documents and state that the knowledge this network contains is what the document is about. This definition is open to subjective views, in that individuals can find different knowledge in the semantic network of a document.

A grammatical model for aboutness refers to aboutness as a synthesis of "the propositions of a document into macropropositions (Svenonius 2000, pa 47, referring to Kintsch and Van Dijk, 1978). Svenonius' has an example starting with this sentence: "Snow is white". The sentence is about snow. Then, if all sentences in a document are about snow, the document is also about snow. This view on aboutness is true for many types of texts. Although it is difficult, not only for some types of literature such as fiction, but for all texts. Thus, Hutchins definition cited above is more general and include all cases, without depending on a strict summary of the text.

Lancaster states: "if one accepts that indexing is most effective when oriented toward the needs of a particular group of users, the indexer's role is to predict the types of requests for which a particular document is likely to be a useful response." (2003, p 17). One of Lancasters' exampels that fit with this view, is that engineers may need documents about certain attributes or functions of materials, such as tensile strength, rather than documents about certain materials. Or, compared to Svenonius' example, documents about "white" rather than about "snow".

One can also see aboutness related to faceted classification, where subjects within a field of knowledge are separated into groups or facets based on the type of concept they represent (Ranganathan, 1967; Vickery, 1966). In this sense, the sentence "Snow is white" is about snow *and* white, but the two topics belong to different facets: matter and properties.

It is also interesting to relate aboutness to relevance. Landcaster does that in the quote above. When indexing, one must decide what the document is about and translate this

aboutness to a subject heading (Lancaster 2003). The ideal subject heading makes the document appear in a search result as a relevant hit. This is, however, not precise enough to be a definition of aboutness, because documents can be relevant for reasons other than aboutness, such as those listed in chapter 3.1.2, with reference to Taylor (2012). In addition, the description and the document must cover the same topic or subject. Lancaster use the expressions “what a document is about” as synonym with “what a document covers” and state: “These expressions may not be very precise and the terms “about” and “covers” are not easily defined. Nevertheless, they are expressions that seem acceptable to most people and to be understood by them.” (2003, p 13).

Ingwersen and Järvelin connect aboutness and relevance indirectly in their definition on aboutness: ”Fundamentally, the concept refers to ‘what’ an information object , text, image, etc. is about (i.e. the topic it discusses), and the ‘who’ deciding the ‘what’.” (Ingwersen & Järvelin, 2005, p. 381). Here, aboutness depend on the person (or machine) who deceide what a document is about.

When applying tags, there is no requirement for the tag to coin the abuoutness of the document. But if it does, the tag will be useful for search and retrieval and in other situations where a desciption of aboutness is needed. I use aboutness is this thesis in order to analyse and understand how users relate to the topical content of aritcles when applying tags

3.1.2 RELEVANCE

Ingwersen and Järvelin define relevance like this:

> The assessment of the perceived topicality, pertinence, usefulness or utility, etc., of information sources, made by cognitive actor(s) or algorithmic devices, with reference to an information situation at a given point in time. It can change dynamically over time for the same actor. Relevance can be of a low order objective nature or of higher order, i.e., of subjective multidimensional nature. (2005, p. 21)

Before them, Saracevic has dealt with relevance in several publications. He defines topical relevance as:

> relation between the subject or topic expressed in a query, and topic or subject covered by retrieved texts, or more broadly, by texts in the systems file, or even in existence. It is assumed that both queries and texts can be identified as being about a topic or subject. Aboutness is the criterion by which topicality is inferred. (Saracevic, 1996, p. 12)

From both definitions, relevance includes that the aboutness of a document is the same as the aboutness of an information need. In a broader way, relevance is when there is a match between an aspect or a characteristic of a document and a users wish or need for this aspect or characteristic.

When seeking for information, people will look for relevant documents, or documents that can give the needed information. Sometimes most people can agree on the relevance of a documents. If you are interested in cancer, there are documents on Cancer.dk that most people will agree are about cancer and are relevant for this interest in cancer. When librarians index documents, they try to find this aboutness of documents, that most of their users can agree on. However, as Ingwersen and Järvelin include in their definition, users' information needs are dynamic. When you have read one or a few basic documents about a topic, your interest will probably change. You may feel that you have all the information you need, or you may want to read documents that are more advanced. Relevance can thus be a judgement that is both objective, in the sense that most people will agree on a relevance decision, and it is subjective, in the sense that only you can find out whether a document is relevant to you in your situation.

Taylor (2012) studied relevance judgements among students in an information gathering process. He used a list of relevance criteria that can exemplify how students judge relevance when searching for literature.

- Accuracy: Document seems to have accurate information about my topic
- Advertisement: Document is an advertisement
- Affectiveness: Document is enjoyable
- Authority: The author of the document is considered an expert in this field
- Bias: Document author takes a stand and has a specific opinion (bias); the author is not neutral
- Breadth: Document covers many topics/subtopics in this subject area
- Definitions: Document contains basic and/or advanced definitions
- Depth: Document contains good depth on the topic
- Descriptions: Document contains good descriptions and explanations
- Guidelines: Document contains basic guidelines and directions
- History: Document contains a history and/or background of the topic
- Novelty: The content of the document adds new information to what I already have
- Recency: Document is up to date and contains current information
- Source: The document is from a source (website, journal) which has a good reputation in this area
- Structure: The structure of the document makes it easier to read and understand
- Time: Document is useful because of time constraints

- Tips: Document contains basic advice and instructions (tips)
- Topic: Document is on my topic and contains information about my subject area
- Understandability: Document is easy to understand; the technical information is easier to read and interpret

(Taylor 2012, p 140)

Here, topic or aboutness is one of 19 criteria for relevance. Taggers can be expected also to consider these criteria when interacting with documents at Cancer.dk, and when applying tags. In my study, I use topical relevance as a background when analysing how the aboutness of tags relate to the aboutness of articles.

3.1.3 WARRANT

When indexing, words are taken from a natural vocabulary and normalized into a controlled vocabulary used for subject indexing. Warrant is used to describe the source of term selection. For a controlled vocabulary like a thesaurus, the warrant may be the literature it is intended to describe (literary warrant). And, it may be taken from the vocabulary of the users intended to use the thesaurus (use warrant). And finally, there is usually a need for structural terms that fill in gaps in order to make the structure logical and complete (structural warrant). Usually, all three types of warrant are considered when constructing a controlled vocabulary (Svenonius, 2000).

The word warrant was first used in library and information science by Wyndham Hulme in 1911-1912. He used the term warrant to describe how actual published literature should be the source for describing literature in knowledge organization systems such as classification schemes. (Barité, 2017). This corresponds to a literary warrant. Later, Lancaster came up with ”user warrant” and stated it to be more valuable than literary warrant (Lancaster, 1977). Later, Svenonius added structural warrant.

For tagging, the warrant will be diverse and based on each individual’s choice. When taggers apply tags as they wish, they select words from wherever they want. But research on tags give some knowledge on how this work. (See also chapter 3.3 on tags and tagging.)

Warrant types can be a fruitful perspective to discuss tags and sources for formulating tags. If taggers tend to select words from the documents to which they apply tags, this is similar to a literary warrant. If they select tags from their own natural vocabulary, or words that they would use to search for document, it can be seen as use warrant. A structural warrant is harder to imagine that tags would imply, unless taggers wish to impose a structure on tags. The tag function “Refining categories”, observed by

Golder and Huberman (2006), is close, in that it indicates a relationship between certain tags.

3.1.4 SUBJECT INDEXING

Subject indexing follows a number of basic assumptions pointed out by Fugmann (1993): We assume that we *can* define and describe subjects, and that we create *order* when indexing. The order is established when those characteristics of a document that are valuable to users are identified, described and findable through search and retrieval. According to Fugman, we also assume that an accurate and predictable subject description is a key to success when interacting with information. This can also be true for indexing in general. If tagging has similarities with indexing, these assumptions may also be valid for taggers' views on tags, and the tags themselves.

On the other hand, sometimes individuals disagree on the topical content of documents and sometimes the topical content of documents is hard to name. Still, practical subject indexing follows the assumptions mentioned above, knowing (or hoping) that the subject descriptions will create *sufficient* order for information interaction, search and retrieval.

A bibliographic language is a formalized language used to describe documents or artefacts in a way that supports storage and retrieval. Bibliographic languages, including subject languages, have a normalized vocabulary that follow a set of rules. Subject languages describe topical properties or the aboutness of documents. They are simpler than natural languages and there is a one-to-one mapping between subject term and the concept it refers to.

In indexing theory, principles are "guidelines for the design of a set of rules" (Svenonius 2000, p 68). These principles make it possible to compare subject languages, and to compare subject languages to a collection of tags. The principles bibliographic languages include:

- *Principle of user convenience* Decisions taken in the making of descriptions should be made with the user in mind. A subprinciple is the
  Principle of common usage Normalized vocabulary used in descriptions should accord with that of the majority of users.
- *Principle of representation* Descriptions should be based on the way an information entity describes itself. A subprinciple is the *Principle of accuracy* Description should faithfully portray the entity described.
- *Principle of sufficiency and necessity* Descriptions should be sufficient to achieve stated objectives and should not include

> elements not required for this purpose. A subprinciple is the *Principle of significance* Descriptions should include only those elements that are bibliographically significant.
> - *Principle of standardization* Descriptions should be standardized, to the extent and level possible.
> - *Principle of integration* Descriptions for all types of materials should be based on a common set of rules, to the extent possible. (Svenonius 2000, p 68, italics in original)

Tags *can* be an answer to the first principle of *user convenience*: what is more user-friendly than using the user's own descriptions? On the other hand, this is only user friendly if the tags follow the *principle of common usage*. It is not obvious that a tagger will apply tags with other taggers in mind. An individual tagger does not know more than an editor about other users, maybe less. But if taggers share interest, tags can follow the principle of common usage simply because what they do is what other users would have done.

For the remaining principles, it is unclear whether tags and taggers would follow them. Based on what is known about tags, it is fair to expect tags to be more messy than subject headings that follow these standards.

Rules specific for subject headings include principles and mechanisms for the construction of a good controlled vocabulary such as:

- Warrant: Vocabulary should be "derived from the literature it is intended to describe" (literary warrant) and include words that real users use for searching (use warrant) (Svenonius 2000 p 135)
- Terminology: Subject headings are words and expressions from natural language, standardized grammatically
- Terms are formulated to distinguish between different meanings of homonyms and polysemic words.
- Synonyms are identified and relationships between them established.
- Compound terms are dealt with.
- In precoordinated indexing, terms are grouped in facets and ordered according to a syntax. In thesauri, terms are faceted and organized in hierarchies. For tags, none of these applications for facets are used. But still, tags can be ordered in facets, and tags can be multifaceted and thus difficult to put into only one facet.
- Subject languages often recommend using the most specific term possible, but not more specific than the topical content of the whole document.

- The number of index terms are regulated. Typically, a book where the subject can be described with one simple concept, the subject heading should be a word or expression that represent this simple concept. If the subject can be described with a compound concept, the subject term or terms should reflect this. But if a book has two or more subjects, rules indicate how many subject headings should be used in each case. For example, a book with four subjects could be described with one subject heading that is more general and thus include all four subjects.

  (The list is based on Svenonius 2000 and Hjortsæter 2009)

The detailed rules for Subject Headings can vary. The Norwegian rules for subject headings can be an example of how various needs are prioritized over time. The present rules, Emneordskatalogisering (Hjortsæter 2009), were made in understanding with a central agency, Biblioteksentralen, that deliver bibliographic records primarily for public libraries and school libraries. One consequence of this, is that the rules ignore local needs and variations. For example, these rules for subject headings do not consider the size of the local collection. This was new when the rules were launched for the first time in 1990. Before that, the rules encouraged to consider local variations and collection size when constructing subject headings. Thus, the specificity of terms for a book was decided after considering how many other books the library owned within a subject or subject area. (Olsen, & Ådland 2015).

In this project, I study properties of tags that enable me to compare tags to other indexing methods, especially subject indexing. Since there is no controlled vocabulary in this sense available for Cancer.dk articles, I will study how tags comply with the principles and mechanisms mentioned above. This includes to examine the aboutness of tags, and group tags into topical facets. I then examine how tags relate to the aboutness of articles in order to evaluate their potential as indexing terms. Thirdly, I study whether tags consist of words from a lay or professional vocabulary.

3.2 INFORMATION INTERACTION

An information website sets the stage for interaction. Goffman explains an interaction as something "which occurs […] when a given set of individuals are in one another's continuous presence" (1959, p. 15). Information interaction can be seen in a broader perspective: The interaction does not occur between persons but takes place between a person and an information source.

Toms has suggested a model of information interaction and states:

> How people interact with information-rich digital environments is directly influenced by the environment's information architecture. The quest for information is carried out through querying and browsing, but also

> represents situated action and reflects the experiences that one has in interacting with an information system. This integrated process is information interaction. (Toms, 2002, p. 855)

According to this model of information interaction, an interaction starts with a selection of meta-information such as a menu item or a search term, and more narrowly within that meta-information. Some cues guide the user further. It can be text segments that "influence the direction the user takes in scanning information." (2002, p. 858). At some point, the user may find something interesting, and extracts and integrates this into his previous knowledge. Index terms and tags are thus not only there to be matched to the users' search terms. They are also pieces of information that the user can interact with, to gain experience and cognition, and to use for further interaction. This corresponds to how index terms are not only useful in search, but also facilitate interaction through guiding and information.

To Toms (2002), user, content and systems are core concepts. Information interaction is interaction between the three: user-system, user-content, and system-content. Toms' model also includes time as an important fourth aspect of the relationship a user has to a single document. "The state of the user changes with time; a single instance is unlikely ever to be the same for the same person or for different people." (2002, p. 857) Thus, the interpretation of a document may differ from one moment to another, even if the same person views the document each time. This is because users move on in their process of information interaction and adapt to the content and its structure on the way. The process also includes constantly evaluating the information found, the system and the process itself.

Tagging features are part of "information-rich digital environments". Here, the interaction will include viewing tags previously in the system, viewing documents and applying tags to the documents. Thus, tags can play a role in both information retrieval and information interaction.

The core of text retrieval is to find a match between search terms that represent the user's need and index terms that represent relevant documents. Information interaction can include this, but it also includes the ability to browse content and its structure. Here, index terms and tags are not only there to be matched to the users' search terms. They are pieces of information that the user can interact with, to gain experience and cognition, and to use for further interaction. They can attract attention, surprise, provide preception, recognition and understanding. Index terms and structure describe the documents and thus inform the user about what he has actually found.

3.3 TAGS

Munk and Mørk describe tags as "descriptive metadata" from users (2007). A similar definition can be found in Benbunan-Finch and Koufaris' article on tagging as a part of private-collective systems, tags are "descriptive words" (2013, p. 189). Rainie

states that tags are labels, and thus "tagging is the process of creating labels for online content" (2007, p. 16).

Hunter describes tags as:

> […] a subclass of annotations that comprise simple, unstructured labels or keywords assigned to digital resources to describe and clarify the digital resource […] Tags allow users to organize resources into categories (groups of resources with the same tag) so they can be more easily retrieved later. Users can define their own tags based on the terms that seem most relevant at the time and can choose to share them with others in their social network. The tags produced by a community of users evolve organically over time […]"(2009, pp. 189–190).

This description places tags clearly in a social setting. The social aspect of tags, however, depends on the system: some systems, like Delicious, invite to more collaboration among taggers than other systems, like Flickr[5] and YouTube[6], do.

Hunter also states that tags are a kind of annotations (2009). This viewpoint is not very visible in research on tags, but it is true: Annotations are comments, viewpoint and descriptions from users, attached to documents or parts of documents. This is also true for tags; only tags are generally shorter than other annotations, and they are a result of and a part of a tagging feature.

The definitions on tags and tagging vary slightly, but they all include that tags are *words or short phrases that users of an information system apply to documents*: web pages, photos, music, videos, etc. The user may or may not also be the author of the document, or its photographer, etc. The collection of documents may vary in type from system to system, and the collection may be big or small. Tagging features have different formal requirements for tags. A tag is limited to one word or may have many words, depending on the system.

In most systems, the *taggers* are those users of a system that choose to apply tags. They are free to formulate the tags. The only limitations are the ones built into the tagging feature, such as maximum number of words or characters. These limitations are usually broad, so it is likely to believe that tags are users' genuine descriptions of the documents.

In social media like Twitter, tags are often distinguished from other words using a hash sign as the beginning of the tag: #. Accordingly, tags in these systems are frequently named hashtags. I consider hastagas as tags. However, they may originate from systems that give them properties that differ from other tags. For instance, Twitter hashtags may differ from tags applied in the Cancer.dk tagging feature in

[5] www.flickr.com

[6] www.youtube.com

terms of what they describe, and what purpose have when applying tags. But Twitter hashtags are still tags.

3.4 FOLKSONOMIES

The aggregated tags are sometimes referred to as a folksonomy (e.g. Guy & Tonkin, 2006; Hunter, 2009; Munk & Mørk, 2007; Vander Wal, 2005). Folksonomies or collections of tags can be defined in different ways and implement differently from system to system. A good way to separate between the different kinds of folksonomies was given by Thomas Vander Vaal (2005), with the concepts broad and narrow folksonomy. Isabella Peters refined this, and put tagging systems into three categories:

- Broad folksonomies: the author and other users can add tags to resources more than once (e.g. del.icio.us)
- Extended narrow folksonomies: the author and other users can add new tags to resources, which are registered only once (e.g. Flickr), and
- Narrow folksonomies: only the author can add tags to resources, but other users can search using these tags (e.g. YouTube)

(Peters, 2009)

The categorization describes the nature of the connection between tags, users and documents. It distinguishes between important differences in tagging features and thus the circumstances and constraints that influence tags in various systems. Tags are often referred to as social because they can be used to form a network between users with common interests (Hunter, 2009). All tags connected to the documents in a system constitute a *network between users, tags and documents*. This is a significant property, especially of broad folksonomies. But restrictions in the possibility for end users to be identified with their tags, like in narrow folksonomies and possibly also in extended narrow folksonomies, can reduce the social aspect of tags.

3.5 INFORMATION WEBSITES

Tags and tagging features relate to document collections. These collections are usually available to the users through a website. Kalbach has categorized websites (2007). He does this because different site genres meet different expectations and need to support different kinds of user behaviour. The categories are:

- Information sites – in the business of providing information

  Examples: Wikipedia, Cancer.dk

- E-commerce sites – web shops
- Corporate intranets – tools for communication and sharing information within an organization
- Community sites – places for people to exchange ideas and discuss topics
- Entertainment sites – such as online gaming
- Learning sites – for online learning or distance education
- Identity sites –to support a company image

These websites can also give access to document databases including a variety of document formats. Both the websites and the databases can include tags. Cancer.dk, the case in this study, is an information site.

3.6 SOCIAL SETTINGS

Websites can be social. The social aspect refers to features that allow users to interact with each other. This interaction can be live, like a chat or a video call. Users can also, immediately or later, react to information that other users previously applied to the site. This interaction is more indirect or represents a delayed social interaction. Examples are forums, debates, and even tags.

Tagging features as social. When users apply tags, other users can use them and apply their own tags. The socialness, however, depends on the properties of the tagging feature. Some features allow taggers to interact directly with each other; some do not keep track of taggers at all. The tags are, however, always social in the sense that they allow users to communicate indirectly with one another, with the system, and with the people behind the system. Users apply tags for an audience and for themselves. The intended audience may vary from user to user and from tag to tag. The users' awareness of this audience may also vary, but there is an audience. Thus, the users may consider this when they apply tags.

In studies of social media, Goffman's model on how we present ourselves to others like actors on a theatre stage, has been used as a framework in various studies (Bullingham & Vasconcelos, 2013). The model points out which roles and positions individuals, teams, locations and things have, and give suggestions for how individuals and teams need to interact for the system to work. Thus, the concepts used to describe the various parts of the model, can also be used to describe and understand users in their interaction with the system. The model gives a framework for analyzing what the roles of the actors have, and thus new ways to see how they interact with each other or fail to interact. The model includes a script, or expectations to how actors should behave, that depend on where they are and whether anyone else can see them.

An information web site like Cancer.dk is familiar to many users, this model helps to understand its features and actors in a way that not only considers the humans involved in it, but also gives the site itself an independent role as a setting or stage for information interaction.

Goffman published the model for the first time in 1956 and intended for it to cover face-to-face meetings. The model also includes a backstage, where the performers can relax from the performance (see chapter 3.6.1.1, about regions). But Goffman presented examples where the borders between the stage (front region) and back stage (back region) are hard to define: "Thus thick glass panels, such as are found in broadcasting control rooms, can isolate a region aurally but not visually, while an office bounded by beaverboard partitions is closed off in the opposite way" (Goffman, 1959, p. 106). Here, glass panels or partitions prevent some but not all impressions to pass. This reveals an understanding of how there can be gliding transitions between stage and backstage, between the front and back regions.

On the other hand, he regarded telephone conversations "as a departure from the norm", a "marginal" way to interact socially and to provide an inferior version of "the primordial real thing"" (Bullingham & Vasconcelos, 2013; Goffman, 1959, 1983, 2010). What his opinion would be on social media as stages for performances is naturally unknown, as Goffman died in 1982. However, researchers use his theory in this setting. In this thesis, the social interaction spans from traditional face-to-face meetings to anonymous tags.

It may seem artificial to use a theatre metaphor to describe tags that include only short representations of communication. However, the metaphor creates a model that will help view the data as clues of human behaviour. With tagging, and social media in general, the website interface is the stage. This is where the performance takes place. Studies can thus examine what kind of performance tagging is and to what degree the users adjust to being performers and audience. Within the frames of various tagging features, taggers may pay more or less attention to the communicative and performing aspects of tagging.

People's online information interaction is anchored in their offline self (Bullingham & Vasconcelos, 2013). Thus, it is possible to see digital footprints as traces of performance. It is not the whole picture, but it gives clues. In this project, it has been a goal to find out more about these traces, mainly through tags. They are small pieces of communication but may constitute a massive body of metadata. It is interesting to find their characteristics and to study whether tags are usable for the performer or anyone in the audience, both online and offline.

3.6.1 PERFORMANCES

Goffman saw social settings as stages where we present ourselves before an audience through performances (1959). The audience is the other person or persons who observe our behaviour or interact with us. A performance is explained as "all the

activity of an individual which occurs during a period marked by his continuous presence before a particular set of observers and which has some influence on the observers" (Goffman, 1959, p. 22).

When we use a website, we are constantly in the presence of the system. Because of the interactivity that websites often have, and other footprints that we leave behind, we can see the use of websites as a social setting. It is, however, more meaningful to restrict the definition of a social setting to the use of sites that include social interaction. Then, users are in presence of the system and in presence of other users and system editors, when they use the system.

The interaction lasts longer than the actual performance, since applied tags remains visible in the system. The interaction involves a nonsynchronous communication where users exchange information using the system as a medium: The communication is not face-to-face, and the performer and the audience are not present at the same time. Each user's communication with the system ends each time the user leaves the system.

Thus, the presence before the audience can be delayed. A user applied a tag one day. Other users can see it for as long as it is present in the system. It is like a play published as a book. The performance is still there, limited to text and its medium, the book. But it is also timeless and available to the audience for as long as the book is available. When users apply tags, the performance stays in the systems like a written play. It functions as index terms in information retrieval or as cues that guide the user when interacting with information.

3.6.1.1 Regions

In Goffman's words, the stage, or the place where the performance takes place, is the *front region*. Thus, in social media, the platform where the interaction takes place is a front region. Website editors publish their texts on the website. Taggers apply their tags through the tagging feature. Both these systems are front regions. The location where the editors and the users and taggers do their work, however, is backstage, or in the back region (see below). This is "where the suppressed facts make an appearance" (Goffman, 1959, p. 112). Goffman defines and exemplifies regions:

> A region may be defined as any place that is bounded to some degree by barriers to perception. Regions vary, of course, in the degree to which they are bounded and according to the media of communication in which the barriers to perception occur. (Goffman, 1959, p. 106)

Goffman defines the *front* region as the place where the performance takes place:

> That part of the individual's performance which regularly functions in a general and fixed fashion to define the situation for those who observe the performance. Front, then, is the expressive equipment of a standard kind

> intentionally or unwittingly employed by the individual during his performance (1959, p. 22)

In other words, the front is the collection of things used to express oneself during a performance. The front includes a *setting*,

> …involving furniture, décor, physical layout, and other background items that supply the scenery and stage props for the spate of human action played out before, within, or upon it. A setting tends to stay put, geographically speaking, so that those who would use a particular setting as part of their performance cannot begin their act until they have brought themselves to the appropriate place and must terminate their performance when they leave it (1959, p. 22).

The setting is the *stage* where the performance takes place. The systems where information interaction take place are settings or stages. This includes the information architecture and design of such systems.

The front also includes a *personal front*. This is:

> …the other items of expressive equipment, the items that we most intimately identify with the performer himself and that we naturally expect will follow the performer wherever he goes. As part of a personal front we may include: insignia of office or rank; clothing; sex, age, and racial characteristics; size and looks; posture; speech patterns; facial expressions; bodily gestures; and the like (1959, p. 24).

Goffman includes appearance and manner in the personal front. *Appearance* is "those stimuli which function at the time to tell us of the performer's social statuses" (1959, p. 24). This is not necessarily important in my study. The roles as cancer patients, editors and researchers do influence persons' social statuses, but no particular visible stimuli were present. When it comes to tags, the choice of words in a tag may reveal social status, but again the extended narrow folksonomy limits the information single tags can give.

Manners are "those stimuli which function at the time to warn us of the interaction rule the performer will expect to play in the oncoming situation" (1959, p. 24). Tags reveal an attitude towards article content, mentioned persons, or Cancer.dk. A few are negative, a few positive, many descriptive, related or even unrelated. This attitude gives a clue about which tags to expect in the future. It also gives new users a clue about how to use the tagging feature, so that they can adapt to an already established manner. In this study, taggers and their tagging behaviour reveal manners.

The back region is a region or a place where the audience has no access. Here, the performers do not have to pay attention to what the audience may see or hear. The impression fostered by the performance in the front region may be contradicted in the *back region*, where "stage props and items of personal front can be stored in a kind of

compact collapsing of whole repertoires of actions and characters." (Goffman, 1959, p. 112). The front region includes a setting or a stage. Similarly, the back region includes a backstage.

Goffman cites Simone de Beavoir to exemplify backstage activities (Goffman, 1959, pp. 112–113). She describes what women do backstage when there are no men present, in contrast to the performance they show when men are present. The example is relevant and illustrative, but it is also clear that this backstage is also a front region in a different play: Women perform for one another. But if any individual is always seen as a performer whenever there is somebody else present, then every place is a potential stage and backstage at most times. It is thus needed to define which performance to look at and who the performers and the audience are, in order to define the stage and backstage more permanently.

If a Cancer.dk user sits at home and interact with the site, this may be a performance in the home, with the family as audience, the home as a setting with or stage, and the tagging feature on Cancer.dk as a back region. This performance is, however, not in the focus of this thesis. Here Cancer.dk and its tagging feature is the setting or stage. The performers are mainly the taggers who apply tags.

Some social settings are located so that it is possible for the audience to enter the backstage. If the neighbour can hear your bedroom activities through the wall, your backstage is not fully protected. If the guests appear in a restaurant kitchen, all activities there will be visible, including the ones one may want to hide from the guests.

When the front region is a website or a social media platform, the actor can simultaneously be in the front and the back regions. Editors can publish their performance online, and at the same time get support from others backstage, and "run through their performance checking it for offending expressions when no audience is present to be affronted by them" (Goffman, 1959, p. 112). The new technologies have brought the front region and the back region closer, but they are still separate regions. In a way, people can sit backstage and publish their performance to the front region. Cancer.dk editors sit backstage when they edit or delete tags. The performers and the audience never meet face to face. They may be present in the front region at the same time, but they are still not together. The performance takes place through nonsynchronous communication.

The concept of regions is an advantage when studying interaction that has similarities across difference systems and features. Each region has its plays, with a stage and a performing team. Thus, if any of them change, the play will also change. This makes it easier to distinguish between performances and to coin what the differences are.

3.6.1.2 Idealization and impression management

Idealization is another concept that Goffman relates to performances: "a performance presents an idealized view of the situation" (1959, p. 35). His examples are of how a

rich person may show off his wealth to maintain his status as rich, or how a poor person may show off his poverty in order to receive welfare or charity. In a system with tags, there may also be idealization. One may want to show off one's interests through tags or demonstrate abilities to apply tags of a certain style or purpose.

Impression management relates to idealization but is more about how people prepare a show and put it on than the actual message they want to present. Impression management relates closely to the team (see chapter 2.1.1) and the front region. Thus, when using this model within social media, impression management can relate to user profiles and the relations between user profiles. Some tagging features include user profiles and keep track of which tag originates from which tagger. In Cancer.dk, user profiles are not a part of the tagging feature.

A small study on tags applied to descriptions of persons in a corporate internal directory reveals that 79% of the active taggers applied tags to the description of themselves, 51% applied tags to the descriptions of themselves only (Muller et al., 2006). The researchers see this as impression management (Goffman, 1959) and a rational way to promote themselves in order to show what they can bring to collaborative work. But the result could also be a consequence of modesty or respect for ones' colleagues so that taggers would avoid labelling others. However, this explanation does not rule out the first one.

On a public information website like Cancer.dk, there may also be impression management. Taggers may plan what impression they want their tags to give and use this as their tagging strategy.

Goffman states that:

> we commonly find that the definition of the situation projected by a particular participant is an integral part of a projection that is fostered and sustained by the intimate co-operation of more than one participant. […] If members of a team must co-operate to maintain a given definition of the situation before their audience, they will hardly be in a position to maintain that particular impression before one another. (Goffman, 1959, pp. 77-78 and 82-83)

When using Goffman's model on social media and tagging, it gives meaning to distinguish between different groups of persons. Taggers, users, editors, authors, all members of these groups have something in common when it comes to agenda, motivation, or at least actions. But are they teams? Goffman argues that the team should be the "fundamental point of reference" (Goffman, 1959, p. 85).

> In the case-study of particular social establishments, the co-operative activity of some of the participants seems too important to be handled merely as a variation on a previous [individual] theme. Whether the members of a team stage similar individual performances or stage dissimilar performances which fit together into a whole, an emergent team

> impression arises which can conveniently be treated as a fact in its own right, as a third level of fact located between the individual performance on one hand and the total interaction of participants on the other. It may even be said that if our special interest is the study of impression management, of the contingencies which arise in fostering an impression, and of the techniques for meeting these contingencies, then the team and the team-performance may well be the best units to take as the fundamental point of reference. (Goffman, 1959, p. 80)

Goffman further gives examples of how a team may consist of only one or even zero persons.

> "Logically speaking, one could even say that an audience which was duly impressed by a particular social setting in which no other persons were present would be an audience witnessing a team-performance in which the team was one of no members" (Goffman, 1959, p. 80).

The way many people will behave when they enter an impressive cathedral may be an example of how an audience can witness a performance without the actors being present at that time. This is an interesting view. I see the Cancer.dk editors as a team, but I also study them as individual performers. And I do not see regular taggers as a team, as they have no common back region where they can "run through their performance" together (Goffman, 1959, p. 112). I believe a common back region is necessary to form a team. Also, taggers do not interact directly with each other in the front region. With an extended narrow folksonomy as the setting, users have no way to interact and to get to know each other's. Thus, they will not form teams and probably also not feel like a part of a team when using the tagging feature. When a user applies a tag, this can be seen as a performance. But the tag remains in the system. The next user can see them as stored performances, or as just a part of the setting, depending on the users view on tags.

Tags can be seen as index terms, and they can describe documents. They can be a part of a user's information interaction. This can be examined, and the usefulness of tags in such settings can be evaluated. But tags also appear in and play a role in social settings. Goffman's model on how we present ourselves to others like actors on a theatre stage can frame the role of tags in social settings and thus add to our knowledge of tags, taggers and their purposes when tagging.

I will come back to the theories and concepts presented here later in the thesis, mainly in the discussion in chapter 9. I will use them to model, analyse and understand the collected data, and when combining results from qualitative and quantitative data.

4 REVIEW OF TAGS AND TAGGING, AND CANCER PATIENTS' INFORMATION SEEKING

4.1 INTRODUCTION

In the beginning of the project and in parallel with the preliminary studies (see chapter 5), I gathered "information to decide how useful it is to make social tagging a part of an information site concerning cancer information" (Ådland & Lykke, 2012, p. 103). Included research was about tags and tagging in general, information behaviour and information behaviour and needs among cancer patients and other users seeking for information about cancer. I did not find previous research on tags applied to cancer information in particular. Thus, I searched research literature to find what to learn when combining research on the mentioned topics.

In 2012, I published a review as part of a book chapter, co-authored with Marianne Lykke (Ådland & Lykke, 2012). The process of reviewing research literature continued during the work, in order to keep up with other researchers' work on similar fields. Thus, the purpose of this review has changed to include research that can shed light not only on whether it is worthwhile to establish a tagging feature on Cancer.dk, but also to give insight in how to interpret the data and experiences from this feature. Here, I bring updates to the review. When doing so, I also give information from the previously published review when necessary, in order to bring the needed background for this thesis.

In short, these are the themes covered in this chapter:

Information behaviour and tagging behaviour in particular: Users apply tags as part of their information interaction. This implies a variety of actions, and I consider information needs and information searcing to be most important. I focused on information behaviour within health and medicine but included research from other fields when relevant.

Seraching challenges and the counsumer vocabulary problem: From the knowledge on patients and health professionals, it is a well-known problem that patients and professionals tend to use different words and understand similar words differently. This problem be a facet of information searching, because choosing words for searh is in the core here. But the vocabulary problem can also be observed in a larger context. Patients and health professionals, and amateurs and professionals, have different view on things based on how they understand words and concepts. I have

looked at the vocabulary problem when reviewing for this project. However, I have consentrated on the context of information searching, although I paid attention to the wider problem when interviewing patients and editors.

Tags and tagging: In this project, users were asked, directly and indirectly, to include tags and tagging in their information interaction. Tags in also in the core of what I study. Thus, I also here present tags and tagging from the researht literature.

4.2 METHOD

I considered a variety of themes that could add up to a larger picture and searched for relevant research in two fields: *Library and information science*, and *Health and medicine*. The search terms differed, depending on the database. I also checked reference lists in central articles, to make sure I did not miss important references. This search for literature has been an ongoing process during the whole project period.

I examined the retrieved literature to find if it could shed light on my research questions. The amount of research literature in these fields is vast. Thus, I paid more attention to reports based on empirical results than articles with viewpoints on tags, and more attention to reports from settings that in various ways have similarities with my project. Various aspects could be similar: tagging feature, type of tags, subject field (health and medicine), types of users, or other aspects. Based on these aspects, I selected interesting articles. The procedure for reviewing literature is the same as the procedure used in Ådland & Lykke (2012)

4.3 INFORMATION BEHAVIOUR AND TAGGING BEHAVIOUR IN PARTICULAR

4.3.1 INFORMATION NEEDS

Knowledge on information needs is important when studying how cancer patiens interact with information, including tags. In the preliminary review, we identified that cancer patients want information on all aspects of their illness. They want information from health professionals and peers. And finding this information can be helpful by itself (Ådland & Lykke, 2012, which refers to ; Ankem, 2005, 2006; Leydon et al., 2000; Rasborg & Nielsen, 2008; Rozmovits & Ziebland, 2004).

Matsumaya and colleges explored the information needs of cancer patients in the US (2013). They found that gender, age, ethnicity, education and marital status influenced

the information needs. The needs for information was highest when the cancer diagnosis was new.

These user needs can be compared to studies and evaluations concerning Cancer.dk and its users. In 2008, Cancer.dk users were asked to fill out a pop-up questionnaire about themselves (Userneeds, 2008a). Results from that study showed that the users were younger than the general Danish population, 51% of the users were between 20 and 39 years old. Most users were women, only 20% were men. They came from all over Denmark. Ten percent of the users opened Cancer.dk as part of their job. Thirteen percent were students, and the remaining 77% used Cancer.dk for private purposes. The respondents were generally happy with the site, but a minority of 13% missed some information that they thought Cancer.dk should have.

Rasborg and Nielsen (2008) studied use of Cancer.dk and identified two kinds of user needs in a focus group study with 45-50 participants. (1) First, Cancer.dk needs to target those who just need specific and short information. They do not need or may want to avoid more information. Other peoples' experiences and feelings about cancer is an information type these users do not want. (2) On the other hand, Cancer.dk also needs to target those who want as much information as possible. These users want to know about other peoples' experiences, and they want possibilities for interactivity.

These two user needs may be conflicting. However, when Cancer.dk changed to meet them, the editors cut articles into smaller units and made sure that headings and links gave full information of the content. The goal was a site where "avoiders" know what to avoid, and those who want all information can find it all, through links with intuitive link anchors.

4.3.2 INFORMATION SEARCHING

Our preliminary review showed that " the Internet supplements, but does not replace traditional sources of health information." (Ådland & Lykke, 2012, p. 108) We also established that patients prefer non-commercial sources and websites with an established reputation, such as the Danish Cancer Society. However, search skills are often limited (See e.g. Morahan-Martin, 2004; Nicholas et al., 2004; Rowlands et al., 2008; Williams & Rowlands, 2007).

In a study where ten participants searched for academic articles in CiteULike and Pubmed, they had access to both controlled vocabularies and tags (Kipp & Campbell, 2010). The results showed that the participants used both tags and subject headings. Participants used the names of taggers to help select resources by relevance. The participants valued the subjective and social information from the taggers, but also the more objective subject headings. This study shows something that is also a starting point in my study: Tags does not replace existing metadata but can be a useful

supplement. This conclusion is not new (see for instance Kipp, 2011), but it has not been verified yet if and how tags on a website will be as useful.

If a system includes tags in its index, it is likely to believe that the index includes layperson vocabulary and thus contribute to result lists when taggers and other users search for lay words. (See also the consumer vocabulary problem, chapter 4.4.) There has been research showing that tags can add good terms to an index, so the potential is there, even though the reality is often more complicated (Melenhorst et al., 2008; Morrison, 2008; Pera et al., 2009). This potential of tags was a part of the motivation for the Danish Cancer Society when they decided to include tags on Cancer.dk. As a starting point, websites such as Cancer.dk are valued and used among health seekers. They can utilize metadata present in the system they use, but search skills are in general limited

4.3.3 TAGS AND TAGGING

These paragraphs include research on tags and tagging, relevant to a tagging project. Individuals apply tags to documents the way they want, but they also seem to learn from each other (Golub et al., 2009). Folksonomies or collections of tags can be defined in different ways. Also, implementation of tagging features differs from system to system.

When the tagging feature on Cancer.dk was set up, tags and tagging were established as something you could apply to documents in systems like Delicionus, LibraryThing, YouTube. Later, systems like Twitter and Instagram are more established. Probably, these systems now have a higer influence on peoples' view on tags, because these hashtags are visible to more people.

Table 1 shows the number of hits after jount searches in the two databases Library, information science & technology abstracts and Library & information science source, from june 2020. I searched for the keywords Tags (Metadata), Folksonomies, Social bookmarks and Social cataloging, in combination wiht publication either Delicious, LibraryThing, Twitter or Instagram, and with publication years 2001-2020 or 2011-2020. Numbers in brackets are the number of hits when only publications coded as peer-reviewed are included.

The table indicate a trend in research, where Twitter and Instabram has become more important in this research. Delicious is still important but relativly less important compared to Twitter. But also research involving LibraryThing has increased in number.

DE "Tags (Metadata)" OR DE "Folksonomies" OR DE "Social bookmarks" OR DE "Social cataloging"	2001-2010 Number of search results (Number of search results that are peer-reviewed)	2011-2020 Number of search results (Number of search results that are peer-reviewed)
AND delicious	21 (15)	19 (15)
AND librarything	8 (7)	14 (14)
AND twitter	2 (0)	43 (31)
AND instagram	0	8 (6)

Table 1 Number of search results in Library, information science & technology abstracts and Library & information science source, june 2020.

The table do not give a full view on status for tagging research, but it gives some indications. I see tags and tagging in social media and social tagging on an information website as two different settings for tagging. I also see systems like Delicious and LibraryThing as closer to the tagging feature as systems like Twitter and Instagram. Thus, here I have focused on research that concerne these kinds of systems.

But an end-user that apply tags may be influenced by other systems with tags, independent of their type. Therefore, I have also referred to research on Twitter and similar systems. It also became clear during my research that Cancer.dk tags had a more communicative side. This also caused an interest in tagging based in social media, where communication is more emphasized.

Twitter and Instagram are more visible and known among users now, compared to in 2011-2013 when the Cancer.dk tagging feature was live. I do not expect the influence from these and other socoal media to be as important back then as it is today. Thus, I use tags in social media not to analyse tags applied to Cancer.dk, but rather to relate them to todays media- and tagging landscape where users are more exposed to tags as they appear in social media.

4.3.3.1 Tag functions or purposes

I have already described how folksonomies can be distinguished into broad, extended narrow and narrow folksonomies (See chapter 3.4) (Peters, 2009; Vander Wal, 2005). Tags themselves can be categorized in various ways. Golder and Huberman (2006) proposed seven categories, each identifying different *functions* that can be connected

to tags in Delicious. Berendt and Hanser (2007) found another category based on their observation on blog tags, category 8 below. This category was also found in our preliminary study (Ådland & Lykke, 2012). A joint list of tag function follows:

1. Identifying what (or who) it [the resource] is about
2. Identifying what it [the resource] is, e.g. article, blog, book
3. Identifying who owns it [the resource]
4. Refining categories – a tag that is dependent of another tag
5. Identifying qualities or characteristics, e.g. scary, funny, stupid, inspirational
6. Self-reference, e.g. mystuff, mycomments
7. Task organizing, e.g. toread, jobsearch
8. Supplementing the content

The first category (1) includes topical tags. These tags can be compared to subject headings and are thus attractive for retrieval. (See chapter 4.3.3.4.) Basile and colleges try to identify such tags automatically by dividing tags into topical and non-topical tags (Basile et al., 2015). The algoritm assumes that topical tags can easily be put into clusters and non-topical tags have a more widespread distribution. This turned out to be hard. They were able to obtain fairly good results on clearly topical and clearly non-topical tags, but other tags were hard to categorize. Tests also included human categorizers, who also had a hard time categorizing tags into topical and non-topical. I believe that some taggers' tendency to apply general tags (see chapter 4.3.3.3) is one of the reasons why it was so hard to categorize tags automatically. I assume that both general topical tags and non-topical tags quite often will have a widespread distribution. Such a categorization is however useful if a certain category of tags turns out to be useful for a certain purpose but can only be utilized when the category is isolated.

Yeh and Chen (2016) may have found examples of category 5 above when they explored how sentiments can be extracted from tags and document sentences. They used a list of words assigned with sentiment information in combination with algorithms. The results indicate that tags can be helpful deciding automatically whether a sentence includes an opinion or not. However, the study uses tags produced for the purpose of the study, and it is unknown what kind of training or instructions the taggers received. So, the results do not reveal which properties social tags have per se when it comes to sentiments or opinions. With a Chinese dataset, it is also hard to evaluate the relevance of such a study in a Scandinavian setting. Still, Yeh and Chen (2016) reveal potential properties of tags. Then, if tags include various types of properties, research to distinguish between them is needed.

Other tags may have similarities with blog tags. Joorabchi, English and Mahdi (2015) studied tags on StackOverflow[7], a question-and-answer website. Unlike many sites with tags, the taggers also write definitions to the tags. Like with tagged blog postings, the authors of questions can apply tags to their own questions. But the goal of Joorabchi and colleagues was to automatically match the tags with Wikipedia concepts. This would indicate that matched tags have a value as subject descriptors. The results were good, meaning that they were able to match tag with a relevant Wikipedia concept in many cases, with scores as high as 99.6%. This study is a clear example of research where the researchers themselves assign a purpose or function to tags: they should describe the subject of the documents they are applied to.

Sometimes it is interesting to explore only a certain type of tags, like Ruocco and Ramampiaro who use Flickr tags to extract information about geographic properties of images (2015), and the above mentioned attempt to separate between topical and non-topical tags (Basile et al., 2015). In general, the variety of tag features result in a variety of tags, tag functions and tag purposes.

As reported in the preliminary study, tags have properties that differ from traditional metadata. And when categorizing the tags accoding to functions, the distribution of tags in these categories, vary. Still, there is a potential there:

> If we compare the information needs to the tag functions mentioned above, tags "Identifying what (or who) it is about" (Golder & Huberman, 2006) has the highest potential to meet these needs. As patients also want information about daily issues and the experience of other patients (Rasborg & Nielsen, 2008), metadata that reflects patients' perspective should help them find exactly what they need. Good metadata, whether professionally produced or from users or both, will also help patients find information and still maintain *hope*. (Ådland & Lykke, 2012, p. 107)

4.3.3.2 Unwanted tags

In tagging systems, it is likely to expect tags that are challenging or unwanted. Before starting this project, members of the Cancer.dk editorial staff, had concerns about this.

Researchers who compare tags with controlled vocabularies, directly or indirectly define tags that do not follow the rules of the controlled vocabulary as unwanted (Carman, 2009; Rolla, 2009; Spiteri, 2007; Yi & Chan, 2009). Guy and Tonkin

[7] http://stackoverflow.com/

observed misspellings, inconsistencies concerning plural or singular form of words, and tags with symbols as categories of challenging tags (2006). Others also add tags in foreign languages (Thomas et al., 2010). The definition of unwanted tags depends on how you expect to use the tags. The number of them will also vary: In Delicious, 1% of the tags contained non-alphabetic characters. Technocrati[8] and Furl[9] had more of these tags, 3% (Spiteri, 2007). And a study of LibraryThing tags reports an even higher number, 23% of tags applied to ten books included non-alphabetic characters (Thomas et al., 2010).

Tags have challenging sides. A small study with nine participants indicate that these challenges are not caused by the user per se (Shultz, 2011). The participants were librarians, library students and persons with no formal library science training. They applied 393 tags to 15 books in LibraryThing. The results indicate that if persons with and without library science apply tags under the same conditions, their tags do not necessarily differ. The results also indicate that the rules that taggers follow, and the taggers' subject knowledge, influence the quality of tags more than who the tagger is.

To sum up, the definition of, and the number of unwanted tags vary. Some of the definitions used above, connect to one or a few of the tag functions in the previous chapter (4.3.3.1). Thus, it was hard to define in advance, whether to expect tags on Cancer.dk that was problematic in any way.

4.3.3.3 Specificity

Munk and Mørk (2007) found that Delicious tags were less specific compared to what is normally found in subject headings.

A study on tags applied to 40 books on alternative and complementary medicine, showed very few tags on specific diseases like cancer or individual treatment (Iyer & Bungo, 2011). Instead, they covered categories that suggested spiritual or metaphysical aspects that are part of holistic health. This result is connected to the nature of these books and their topic but shows again that tags can be less specific than subject headings. Possibly, tags are also less specific than the content of documents they are applied to.

A comparison of Flickr tags applied to 2,000 images and search terminology for the same images revealed a gap between the specificity of search terms and tags (Ransom & Rafferty, 2011). The search terms were more specific than the tags, even though tags and search term often covered the same content.

---

[8] http://technorati.com/

[9] http://www.furl.net/. Furl shut down its service in 2009, according to a Wikipedia article about the site.

In my study, I intended to analyse tags based on specificity. Tags from other tagging features have been analysed in similar ways. Voorbij (2012) found that 4.4% of tags were narrower than document topic, and 25.2% are broader. They compared 600 records from the catalogue of an academic library and compared them with LibraryThing tags. Thomas with colleges found 0.3% narrower and 0.3% broader tags compared to subject headings when they compared approximately 8,000 LibraryThing tags with Library of Congress Subject Headings, all applied to ten popular books (Thomas et al., 2009). These two examples shows that even the same tagging feature, LibraryThing, can give tags with different properties when it comes to specificity.

Hajibayova and Jacob (2015) studied tags as basic categories or basic level terms in a laboratory setting where 40 participants applied tags to 36 online resources about four topics. They found that *unique* tags had a larger amount of subordinate terms, terms that refer to a lower hierarchical level than what a group of experts saw as the basic level. This contradicts the findings of Munk and Mørk (2007) and the other studies referred above. The contradiction may originate in differences beween the research settings. It is also possible that taggers in 2007 behaved differently from taggers in 2015. On the other hand, when Hajibayova and Jacob considered all tags, the distribution between tags that refer to a lower, basic or higher hierarchical level, compared to the content of the documents, was more even.

When distinguishing between topical and non-topical tags, Basile et al. (2015) identified a third category of tags, *umbrella tags*, which were hard to categorize automatically as topical or non-topical. These are high level tags, and can probably be compared to the high level tags in Hajibayova and Jacob's report (2015).

Again, tags differ. Nevertheless, if tags are used for information retrieval, it is a goal that tag and search term is similar whenever the document is seen as relevant to the searcher. This includes the level of specificity.

4.3.3.4 Tags in information retrieval

As reported in the preliminary study, tags can help improve information retrieval, when tags are a part of a system with other types of metadata as well. (See Ådland & Lykke, 2012, which refers to ; Melenhorst et al., 2008; Morrison, 2008; Pera et al., 2009)

A more recent study reports on attempts to use tags for disambiguation or to reveal facets. Zhitomirsky-Geffet and Daya (2015) used popular Google search terms for which the result lists were mainly documents already tagged in Delicious. The results showed that in this setting, Delicious tags could differentiate between topics in the result list, so that documents about various aspects or meaning of for instance *Mozart* could be separated from each other: the composer's music, the composer's life, and other entities named Mozart.

Soohyung Joo and Kipp (2015) explore how tags represent the hierarchical structure of Web space in the field of library and information science. For example, tags applied to public libraries, differ from tags applied to a library school. They point out that taggers produce these distinctions in spite of the fact that they may not be domain experts. On the other hand, some taggers who apply tags to LIS home pages may be domain experts, and you do not need to be a domain expert to be able to distinguish between a library and a library school. However, the model used suggest a possibility to cluster websites and their corresponding organizations, based on the tags applied to them. The study used data collected from Delicious in 2011, and tags that occur at least six times amongst 78 selected sites.

Vaidya and Harinarayana (2016) compared LibraryThing tags with Library of Congress subject headings in order to find similarities and explore how tags can supplement ordinary metadata such as subject headings. They used 100 titles about library and information science, published between 2000 and 2015, with 2 or more LibraryThing tags. Out of 744 unique tags and 341 subject headings, 160 terms appeared in both lists. In other words, 21.5% of LibraryThing tags appear as Library of Congress Subject Headings, and 46.9% of the subject headings appear as tags. 12% of the books in this study have tags and subject headings that do not overlap at all.

In a recent study to explore tags and subject headings in image search, test persons more often were successful when using an interface where they could utilize subject headings and tags in parallel (Lin et al., 2016). Two datasets were used in this study: First, 1,864 images came from a museums' collection, with subject headings from the museum itself and tags applied by users of the Amazon Mechanical Turk. Second, 5,281 images selected from Flickr. The images were all uploaded from the Library of Congress, and had both subject headings from the library and tags from Flickr users. For lookup tasks, the use of only tags, only subject headings or the use of both types of metadata gave fairly the same success rate. But for exploratory tasks, the use of both types of data gave a higher success rates: Test persons succeeded 54 of 54 times using both types of data, 49 times when using subject headings only and 45 times with tags only.

Both the research referred to above that were published before I started my study, and the research conducted in parallel with my work, show similar potential with tags. They can play a role in information retrieval, but most likely not as the only metadata in a given system.

4.3.3.5 Tag contribution and purposes

A tagging system is of no use if no one applies tags. This was also an issue for the Danish Cancer Association; would a tagging feature on Cancer.dk attract tags? In the preliminary study, we found examples of studies of systems with similarities to Cancer.dk, that attracted tags (Ådland & Lykke, 2012, which refers to ; Fox & Jones, 2009; Peterson, 2009; Rainie, 2007).

Benbunan-finch and Koufaris see tagging as a combined private-collective process. They conclude that "both selfish motivations associated with the need to feel competent (egoism) as well as selfless concerns for the needs of other users (altruism) drive intentional contributions to the public repository in social bookmarking systems." (2013, p. 199). The researchers asked a group of users of MyWeb[10] questions about tagging behaviour and tagging purposes. They found that users, who care for the needs of other users and want to help them by applying tags, apply less tags than others do. The researchers' explanation to this is that the concerned taggers have a more restricted tagging behaviour in order to secure quality of tags. In other words, taggers may avoid spam and only apply tags they see as useful for other users. The researchers also state that the altruistic tagging is "not a response to specific requests from other users" (Benbunan-Fich & Koufaris, 2013, p. 198), but more a motivation for the taggers' own tagging behaviour. This is an interesting view, but it is not clear whether altruism is a mechanism that directly causes fewer tags.

It is has been stated that tagging has a "low participation barrier" (Furnas et al., 2006, p. 37). However, if most taggers are "early adapters of technology", one may question whether it is easy or not to apply tags for other groups of the population, like cancer patients. Taggers' diverse motivation and purpose for applying tags will influence the tags in a given system.

4.3.3.6 Tags in social media

Over the years, social media such as Twitter has positioned as an important source of tags for research from various angels. A systematic review on Twitter as a tool for health researh found four ways in which the data was used:

> content analysis of tweets (56%; n = 77), surveillance of volume of tweets about a particular topic (26%; n = 36), engagement of Twitter users with Twitter accounts or tweets (14%; n = 19), and network analysis of Twitter users (4%; n = 137) (Sinnenberg et al., 2017, p. 5)

The review included 137 peer-reviewed articles from 2010-2015 from which 13, or 9%, used hashtags as an explicit data source. In general, the researcher found that Twitter metadata such as demograpic information was underused in the reviewed research, but it is not clear wether they include hashtags as an underused metadata or not.

When studying Twitter hashtags, researchers seem to see them as keywords that may show tendencies in tweets and thus represent the content of the tweets. But indirectly, they also see them as part of the content because their presence in the tweet itself is regarded as an expression of a trend or a viewpoint. One example is Hemsley et al,

[10] The site later merged with Delicious

who use twitter hashtags to explore the concept of digital nomadism and mobile work (Hemsley et al., 2020).

Haunschild with colleges studied hashtags in tweets about scientific articles on climate change (2019). Here, a typical tweet will include the title or a part of the title, an url, and tags. The study is based on an assumption that the tags refer to the article, not the tweet. Thus, the tags are not metadata about the tweet, but more information, more content. This corresponds to the "supplementing the content" – category discussed in chapter 4.3.3.1. Or, the whole tweet is metadata about a scientific article, both tag and text.

One study compare tweets with hashtags and responding tweets with and without hashtags. Data were collected among tweets with the hashtags #vk14 and #vk2014 as starting points. (D'heer et al., 2017) The tags refer to the Belgium election in 2014. When users published a message with one of those tags, some responds would include the same tag, others would not. The researchers found that the use of hasttags "co-occured with [a]n inclusion of additoial interactive, information elements", such as other hashtags and hyperlinks (D'heer et al., 2017, chapter 5). According to their analysis, the interactive elements were *informational* rather than *conversational* in nature. Thus the presence of tags in the ir material tend to co-occur with tags that seem to have a informational purpose, rather than a communicative. But still, all these tweets were part of communication about an election. The researchers also found that elite users " occupy central positions in Twitter networks" (2017, chapter 4.3). Elites are here defined as political parties, politicians, spokespersons, traditional media outlets and journalits. This result can not be directly compared to users on Cancer.dk, but internal and external users may have similarities to elites and non-elites on Twitter.

When tags are part of social media messages, it is likely to see them as communicative. One importan role of hashtags is to enable a conversations between users. Using the same tag is one way of saying that a message is part of a larger communication that somehow can be characterized by the tag. #metoo in a well known example of such a tag. It does imply a general topic, but its function is mainly to connect a large and faceted conversation distributed on different threads and platforms. Using a common hashtag for a local conference or seminar, is another way of communicating through hashtags. Such a tag is topical if the tweet (or other messages, if used on other social media platforms) is about the conference or seminar itself, but otherwise it is not.

4.4 SEARCHING CHALLENGES - THE CONSUMER VOCABULARY PROBLEM

In the preliminary study, we found that:

> Patients and health professionals tend to have a different understanding and model of a patients' situation (Keselman et al., 2008; LaFrance, 1989; Patel et al., 2002; Schoenfeld & Herrmann, 1982). This may be one of the reasons why many studies have observed differences in the vocabulary used by patients and health professionals. When searching for information, vocabulary is crucial. (Ådland & Lykke, 2012, p. 109)

We examined some of these studies, but also research where patients tended to adapt the professional language (Poikonen & Vakkari, 2009), and vica verca (Glenton et al., 2006; Smith et al., 2002).

Still, one conclusion was that:

> If patients' search terms lead to poor information retrieval because they choose their search terms from their own mental model, a mental model that differs from the one reflected in the database, they might also tag with terms from their own mental model. (Ådland & Lykke, 2012, p. 111)

Systems owners may find tags and tagging attractive because it is cheap. They can set up a system and have the users tag and thus do the indexing-job free. Still, the results depend on which tags the users apply to which documents, and the system that use these tags. It is interesting to know whether tags can bring a different type or style of vocabulary into a system

4.5 PERSPECTIVES TO BRING TO THE TAGGING FEATURE OF CANCER.DK

Research on tags both in live system and various laboratory settings show that tags differ in many ways. The functions they serve, and how categorize them, the specificity of tags and how they can contribute to information retrieval. In combination with users' information behaviour and the challenges lay users face when searching for information formulated by professionals, many uncertainties surrounds a tagging feature.

I intend to mirror the properties of tags identified here, with the properties of Cancer.dk tags, in order to obtain a robust view on what properties tags can have.

5 PRELIMINARY STUDIES: DESIGN OF TAGGING FEATURE

5.1 INTRODUCTION

To prepare for the design of the tagging feature, I first analysed tags applied to blog postings at Blogomkraeft.dk[11], a blog site at Cancer.dk. I compared these tags with the formal browsing structure of Cancer.dk to explore similarities and differences in topics and terminology used to represent cancer topics. Researchers in the FIRE team then developed a prototype for social tagging on Cancer.dk. I evaluated this prototype in a usability test with five Danish cancer patients. The purpose was to investigate:

- the tagger's understanding of the tagging feature
- whether the participants found the tagging feature useful.

Altogether, I used findings from these studies to define the purpose of social tagging for cancer patients at Cancer.dk and design the feature of the tagging system. In 2012, we reported from these studies (and a review mentioned in chapter 0) as part of a book chapter (Ådland & Lykke, 2012). Here, I bring a brief description of the preliminary study.

5.2 ANALYSIS OF BLOG TAGS

Tags in various systems, and applied by various users, differ. It was interesting to know more about the future taggers before participating in the design of a new tagging feature. Blogomkraeft.dk was a system that share properties with Cancer.dk, and thus could give indications of future use, for two reasons. First, the domain is the same for the two systems. Second, the users of Blokgomkraeft.dk had similarities with the future Cancer.dk taggers, although they also differed. These two features made Blogomkraeft.dk a good source of insight into how to design and what to expect from a new tagging feature. When the tags was compared to the Cancer.dk browsing structure, it was to reveal differences between the data and thus find what tags possibly can add to Cancer.dk.

Unique tags from Blogonkraeft.dk was coded into topical categories. Among our observations, I will mention two. First:

[11] www.blogomkraeft.dk. The blog does not exist on this address today.

> With Golder and Huberman's (2006) tag function categories, we found that the Blogomkraeft.dk tags were rather homogenous. One category dominated the tags, the first category Identifying what (or who) it is about. Examples are brystkræft (breast cancer), krise (crisis), and ventetid (waiting time). (Ådland & Lykke, 2012, p. 113)

Second, tags at Blogomkraeft.dk relate to content of the postings, which is promising for the design of a new tagging feature. They are inconsistent, which implies that tags do not constitute an independent search entry to the documents they are applied to. But they can supplement other metadata.

5.2.1 COMPARING BLOGONKRAEFT.DK AND CANCER.DK

Cancer.dk had a browsing structure with keywords for pages and groups of pages. When the tags were compared to the Cancer.dk browsing structure, it was to reveal differences between the data and thus find what tags possibly can add to Cancer.dk:

> There are two main differences between the tags at blogomkraeft.dk and the structure of Cancer.dk. The first and most obvious one is the size and coverage, Cancer.dk is bigger and broader. The second difference is that the origin and the goal of the two sites differ. The most frequent tags are displayed in the tag cloud on the Blogomkraeft.dk home page, but the tags in general are not utilized as directly as the Cancer.dk structure, which, by definition, is there to make browsing possible. This means that it is likely to assume that the bloggers/taggers did not have the same focus on, nor experience with how their tags would or could be used.
>
> To summarize, tags that concern cancer *treatment* and *thoughts, feelings and psychological symptoms* are more detailed and specific on Blogomkraeft.dk compared to the structure of Cancer.dk. As the blog includes information from patients to patients, their focus on these subjects may reveal a need for Cancer.dk to make these subjects more prominent on their site and more accessible to its users. (Ådland & Lykke, 2012, p. 119)

5.3 TAGGING PROTOTYPE AND USABILITY STUDY

From the study of Blogomkraeft.dk and Cancer.dk site structure, it was learned that a tagging feature could be useful on Cancer.dk. A usability study was set up to find out how to implement a tagging feature, using a prototype developed after the workshop and discussions about such a feature. The prototype was also a tool when discussing the future tagging feature with the Cancer.dk editors.

5.3.1 METHOD AND PARTICIPANTS

We set up a prototype and invited users to test it in a lab-like user test supplemented with questionnaries and interviews (Case, 2012). Details can be found in Ådland and Lykke (2012)

5.3.2 RESULTS AND CONCLUSIONS

From the usability test, we concluded:

> The usability test showed that our test persons liked the tagging feature and judged its usefulness of the tagging feature at 4.25 on a scale from 1 to 5. They were able to understand and operate its function, even with relatively little Internet experience. In general, we found that the prototype was good and clear; thus not many changes were needed before adding it to Cancer.dk.
>
> The test persons also expressed a willingness to apply tags not only for themselves, but to help others finding information. (Ådland & Lykke, 2012, p. 123)

The preliminary studies altogether gave answers and revealed new questions, for which this thesis is an attempt to give answers:

> Further studies were needed before we could conclude whether tagging actually will support user interaction and information retrieval at Cancer.dk. The next step in our project was to conduct studies after a tagging feature has been added to the site, through tag analyses, user observations, and interviews. We will test and evaluate tagging and the use of tags, for example, for information retrieval. We will study tagging behaviour and use of tags, and whether users feel that they can trust and use the tags. We will focus on whether taggers describe the subject of the articles, add content or apply other kinds of tags. (Ådland & Lykke, 2012, p. 124)

6 THE CANCER.DK TAGGING FEATURE

This chapter introduces the tagging feature produced for Cancer.dk. ProActive, who also provided Cancer.dk with its content management system, produced the tagging feature at Cancer.dk. The feature incorporated findings and results from the review (see chapter 0) and the preliminary studies (see chapter 5). The process that lead to the finished feature is described in chapter 2.2.1.

The potential of tags referred to in previous chapters was part of the motivation for the Danish Cancer Society when they decided to include tags on Cancer.dk. They were aware of differences in understanding and vocabulary between patients and health professionals (see chapter 4.4). Cancer.dk already targets patients and their relatives, so they already tried to include lay vocabulary in their articles. It is still hard to include everything because some lay terms are inaccurate, and it is difficult to think of every possible wording that any user could prefer to use. They wanted tags to see if tags covered a more complete lay vocabulary.

The tagging feature was live from December 2011, first for internal taggers, then for any user of Cancer.dk. The first logged tag was from November 8, 2011. The first tag from an external IP was from November 30, 2011. After 14 months there were 25,253 tags in the transaction log. Chief editor Tor Øyan administered the feature from the beginning. See chapter 7.1 for details on data collection from the tagging feature, and chapter 7.2 for observations and results on tagging behaviour.

Before the launch of the tagging feature, the design of Cancer.dk changed. Thus, the prototype and the tagging feature have a different design. The fundamental structure of the site remained unchanged.

6.1 AN EXTENDED NARROW FOLKSONOMY

The tagging feature on Cancer.dk constituted an extended narrow folksonomy (see chapter 3.4). Both authors and other users could apply tags to the articles. It was not possible to sign on and maintain a private collection of articles with tags. Only editors could edit or delete tags. The feature did not connect tags or articles to individual users at all, only the log file included the IP-address of users. This resulted in a feature well integrated with the structure of Cancer.dk, as this was the only place to explore tags. But it also did not allow taggers to bookmark articles privately, through the tagging feature.

This project included all three types of folksonomy, according to the categorization of Vander Wal (2005) and Peters (2009). Blogomkraeft.dk was a narrow folksonomy where only the authors could apply tags to their postings. I studied this because it was available and within the field of cancer. It also included tags from the same user group as Cancer.dk, and thus it was expected the tags to give an indication of what properties Cancer.dk tags would have. But it was never intended to introduce a tagging feature with a narrow folksonomy to Cancer.dk.

The prototype was not fully functional, but it indicated a broad folksonomy. Some of the participants saw that they were signed in as *Jeremy* when they used the prototype. Some of the questions they were asked included the possibility of a sign-in feature. However, the prototype itself did not facilitate fully functional broad folksonomy; it did not save tags at all.

It was my wish to test a broad folksonomy, but the Danish Cancer Society did not want this. They expected a sign-on feature to reduce the number of expected tags. It was also hard to set up such a feature connected to articles and tags in a way that was sustainable in the long term, based on the underlying structure of Cancer.dk and the plans for further development of the site. So, we agreed on an extended narrow folksonomy. At the time, we did not know how many tags to expect, so the solution also seemed “safe” from that point of view.

The solution however contradicted one result from the prototype test (see chapter 5.3): the test persons liked the bookmarking function that they could use tags to come back to a known article. This was still possible with the present tagging feature, but not as a personalized feature.

6.2 THE FEATURE ITSELF

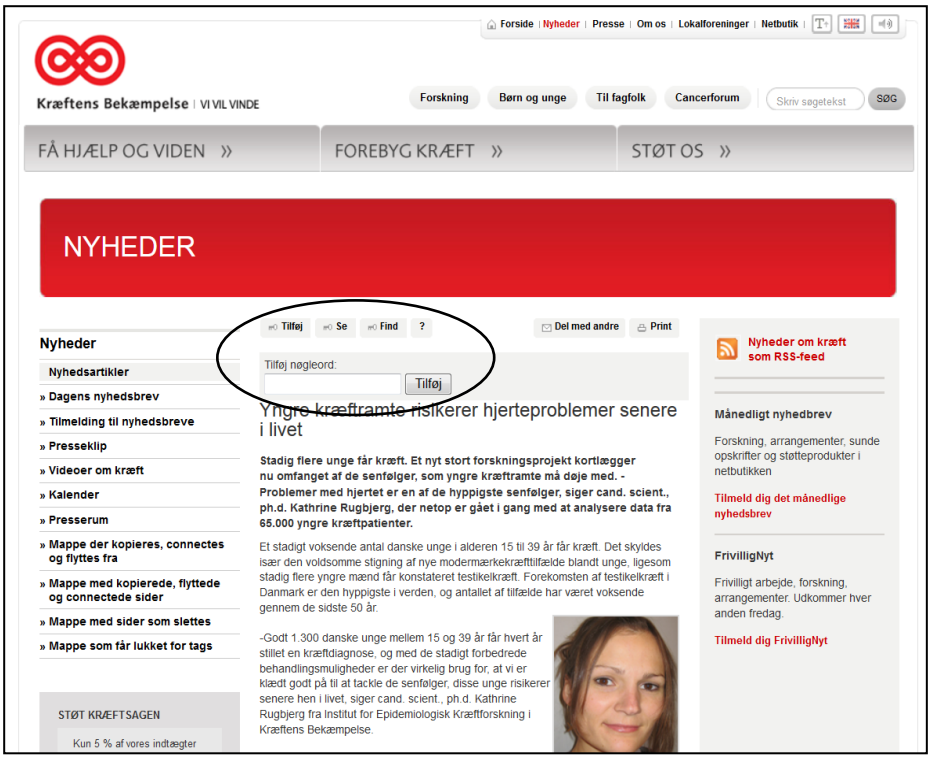


Figure 1 Article with tagging feature above heading, highlighted with an oval. Buttons "Tilføy", "Se", "Find" and "?" are above the tagging field "Tilføj nøgleord". The field for applying tags is open.

The tagging was most visible on top of the articles, above the heading. Figure 1 gives an example. The tagging feature consisted of four buttons:

- *Tilføj* (Apply) – to open a field where you could write a tag and apply it. In the beginning, this field was open by default when users opened a page, as shown in Figure 1. After the feature change in September 2012, the button was there, and the field opened when users clicked on it. Thus, the default view was as shown in Figure 2.
- *Se* (See) – to see tags already applied to the article (Figure 4).
- *Find* (Find) – link to a tag-browsing page (Figure 5).
- *?* – opens a field with information about the tagging feature.

Figure 2 Tagging feature visible on article view after feature change, highlighted with an oval. Only buttons "Tilføj", "Se", "Find"and "?" are visible. Users had to click on "Tilføy" to open the tagging field itself.

The '?'-button in the tagging feature opened a box with this text, see also Figure 5:

> Klik på 'Tilføj' for at skrive det nøgleord, som du vil tilføje til siden. Du kan tilføje ét ord eller et samlet udtryk ad gangen.
>
> Klik på 'Se' for at se, hvilke nøgleord andre brugere har tilknyttet til denne side. Klikker du på et bestemt nøgleord, vil du få en liste over de sider, hvor netop dette nøgleord findes.
>
> Klik på 'Find' for at komme til en oversigt over alle nøgleord (tagcloud).
>
> *(Click 'Apply' to write the tag that you want to apply to the page. You may apply one word or an expression each time.*

Click 'See' to see which tags other users have applied to this page. If you click on a certain tag, you will find a list of those pages where this tag was applied.

Click 'Find' to find an overview of all tags (tag cloud).)

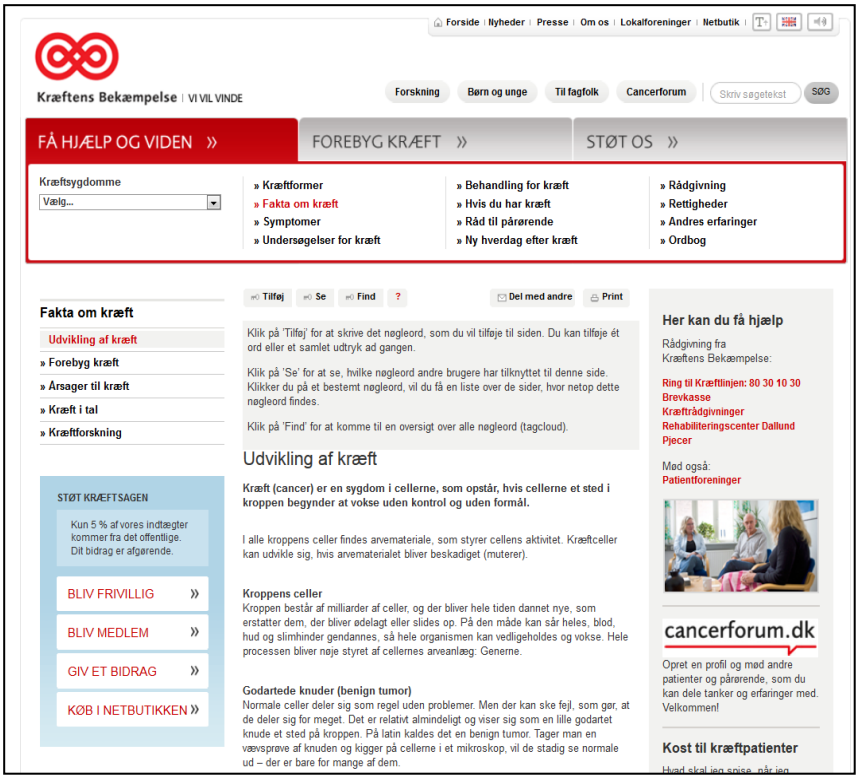


Figure 3 Information text about the tagging feature, highlighted with an oval. Users had to click on "?" to reveal the text shown here.

This text attempted to explain the tagging feature to users and potential taggers. I observed the need for this explanation among the test persons in the prototype test (see chapter 5.3). The prototype did not include this information, and I had to give it to the test persons so that they could understand both the prototype and my questions for them.

To beginners, this information could be more visible, not hidden behind the '?'-link. But with limited space, this was the solution the Danish Cancer Society ended up with.

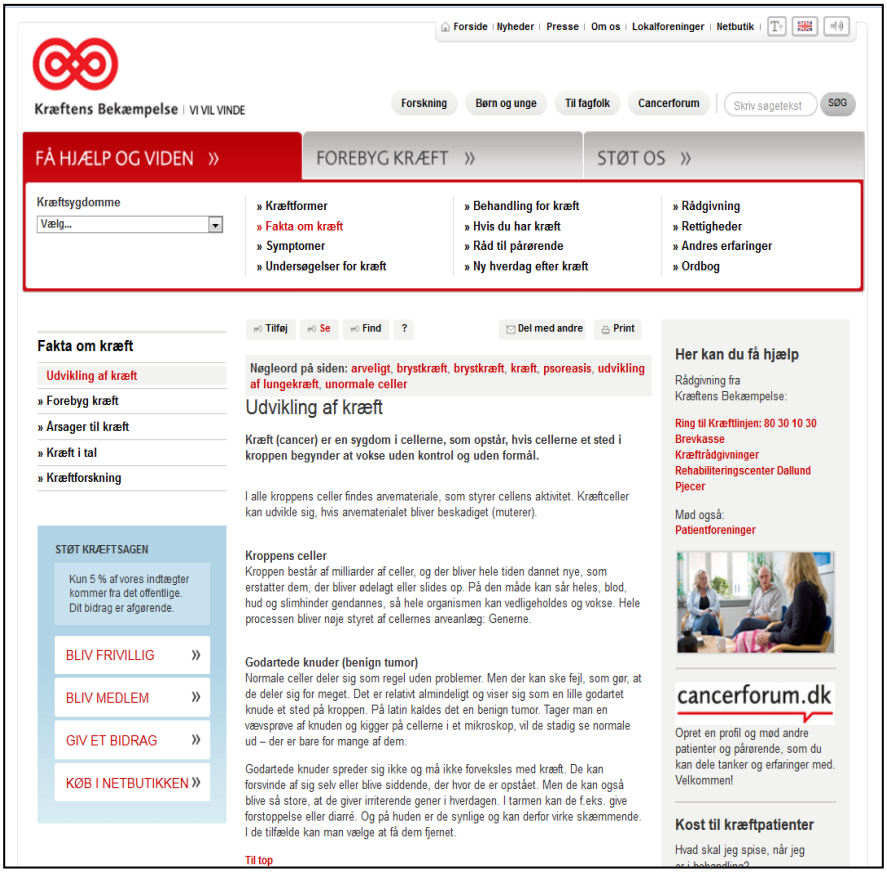


Figure 4 A page where the 'See' button was clicked, highlighted with an oval. Tags already applied to the article are shown after 'Nøgleord på siden:'.

The *tag-browsing page* included a tag cloud on the centre of the page. See Figure 5. A tag cloud is probably the most popular way to display tags visually within a website (Gerolimos, 2013). This one included the 200 most frequent tags. A click on a tag in the tag cloud resulted in a list of articles to which this tag was applied. A pane on the left side included two tag-browsing features: A scrollable list of all tags and a list of the ten most frequent tags. Their titles were *Alle* (All) and *Top 10* (Top 10). On both lists, one could click on the tags. This resulted in a list of articles on the centre of the page. The list included all articles where this particular tag was applied. If a second tag was clicked on from the left side panel, the list of articles would include all articles with *both tags* applied to them. Figure 6 shows such a list. It was possible to un-check a tag, update and expand the list of articles.

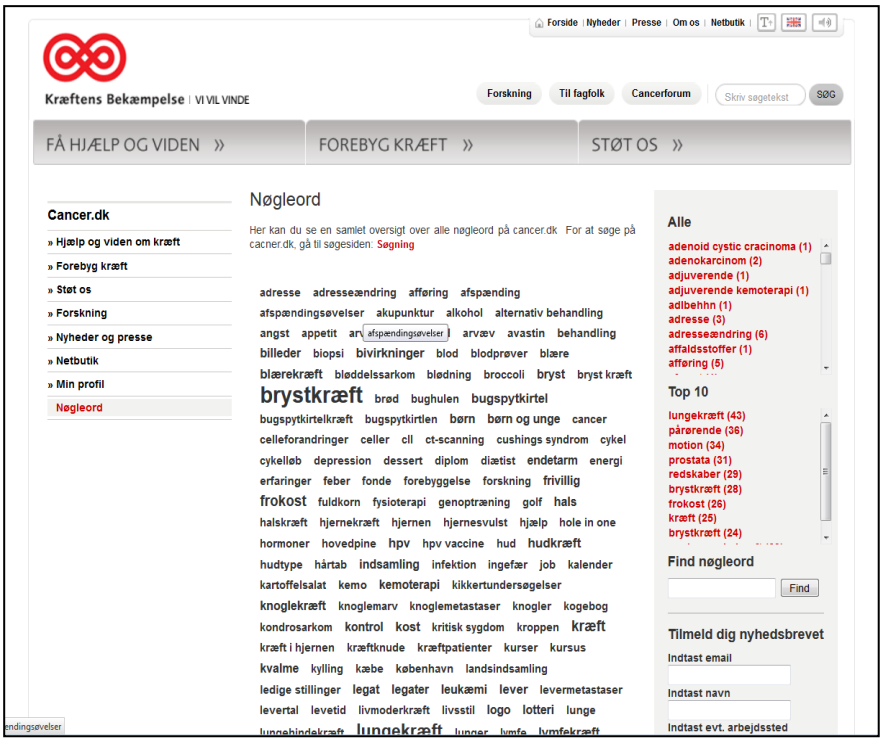


Figure 5 The tag browsing page, from the early period when not many tags were included in the system. Selected tags are listed in the main section of the site. In this figure, all tags are selected. To the right, there is a list of all tags on top. This scrollable list shows all tags independent on tag selection in the main section. Below is a top ten list.

From the beginning, tags were planned to be visible at Cancer.dk immediately after they were applied. The prototype test confirmed that this was important. The prototype did not save tags, and the participants looked for their newly applied tags without finding them. In a test situation this can be pointed out to the participants. But in real life this simply had to be in place. The editors on the other hand also saw a need to delete inappropriate tags. When tags were published immediately, inappropriate tags would remain in the system until one of them discovered and deleted them. The tagging feature included an editing tool so that this could easily be done.

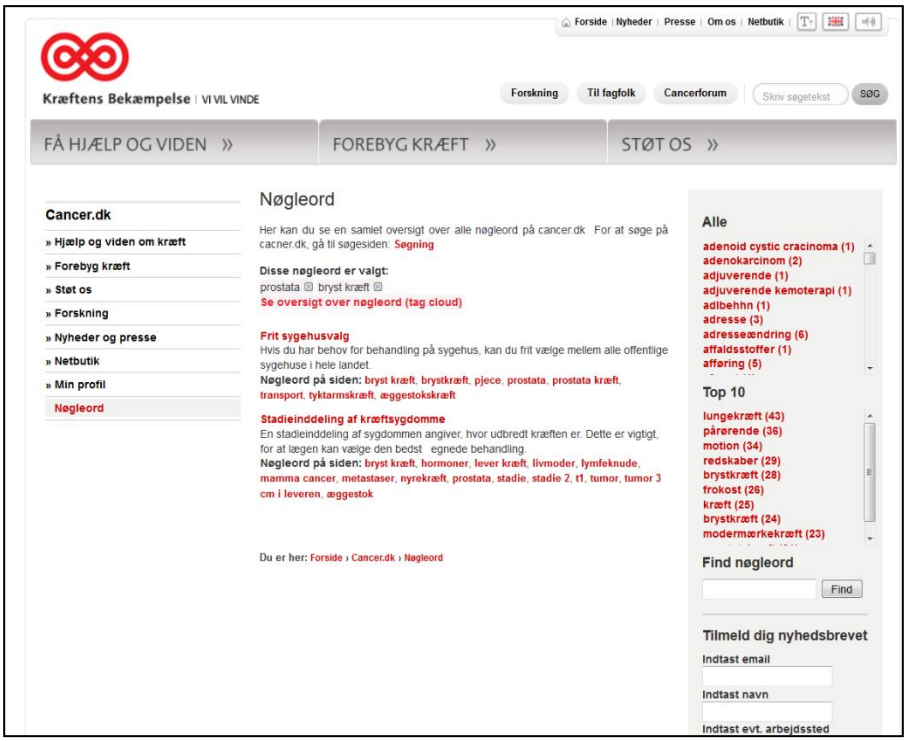


Figure 6 List of articles applied with the tags bryst kræft (breast cancer [in two words, a typing error in Danish]) and prostata (prostate). The list includes two articles.

The original plan was to leave the tagging feature unchanged for one year. However, some initial observations on the tags motivated a change. The main motivation was a high number of tags not related to the content of articles. See chapter 7.1.1 for details.

The following chapters presents tags and other data accumulated through or about the tagging feature.

7 CANCER.DK TAGS AND THEIR USAGE

7.1 METHOD

The study of the actual tags on Cancer.dk gives information about the nature of tagging, and the use of tags. The tags were categorized based on four criterias. The resulting categories, and the combination of them, give information on interests and influences of taggers. Chapter 7.1.5 has more information about how tags were categorized, and chapter 2 more considerations on methods.

7.1.1 CHANGE OF TAGGING FEATURE AND OTHER CHANGES ON CANCER.DK

The original plan was to leave the tagging feature unchanged for one year and log all activity. However, after a few months with the tagging feature live, The Danish Cancer Society worried about a high number of what they saw as irrelevant tags. They believed that the field for applying tags was used as a search field by mistake. Thus, they changed the tagging feature on September 11, 2012. The main difference before and after the change, was the visibility of the field to apply tags. Before, the field for applying tags was open by default on top of every article. Now, it was hidden, and users needed to click one of the buttons in the tagging feature to apply a tag. Figure 2 shows the tagging feature after the change. For comparison, Figure 1 shows the tagging feature before the change. In both cases, the buttons for tagging were visible on top of articles both before and after the feature change.

The change made it possible to study how this feature change influenced the tagging behaviour. The results of this study can mainly be seen in chapter 7.2.12, where data are separated into two parts, before and after the feature change.

The feature change was not planned. Thus, in interviews with patients, this change had not happened, and was not mentioned. When interviewing the editors one year later, I discussed it with them.

The tagging feature seemed to slow down Cancer.dk in the beginning. In October 2012, the site was down after a fundraising campaign, and then moved to new servers. This resulted in a site that was easier to download, an important change especially for frequent users like the editors. Their feeling that the tagging feature delayed the site, was now gone. The features connected to editing the articles in the cms system also loaded faster.

Still quite many tags were deleted, and the tagging feature were removed from the lottery pages during spring 2012. The feature was also removed from landing pages, pages where you end up after completing a form. Figure 7 shows a lottery page with no tagging feature.

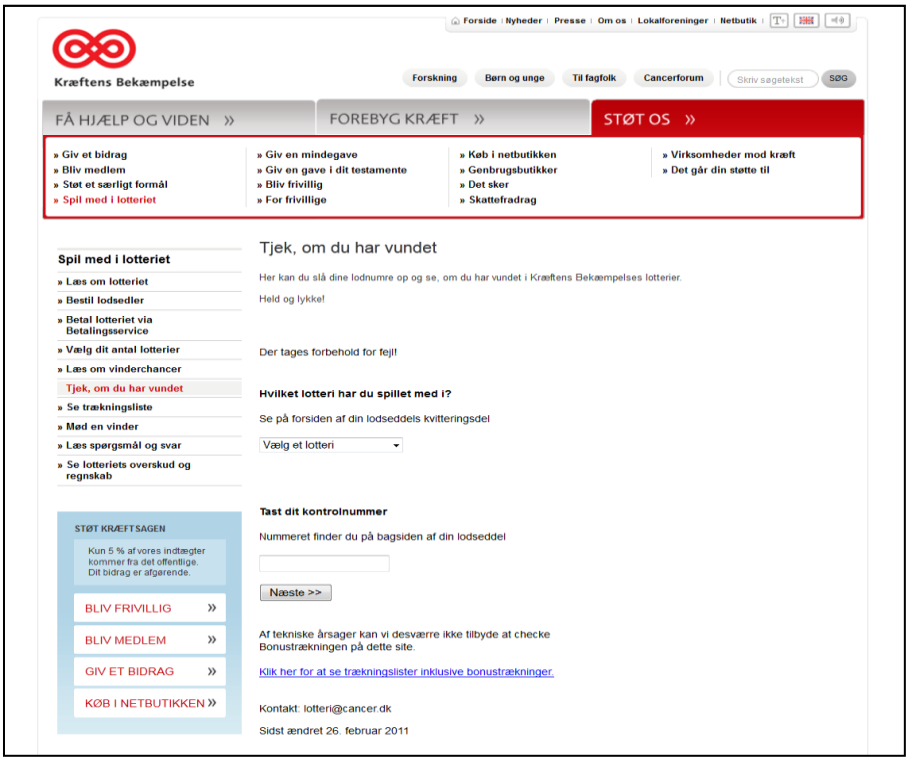


Figure 7 Lottery page to check for price. Users seemed to use the tagging field if they did not find this excact page.

7.1.2 LOG FILE

The transaction log is fundamental for this part of the study of tags at Cancer.dk. The tagging log file was produced by ProActive, who also made the tagging feature. Before that, the project team and the representatives from National Cancer Society engaged in a process to design the tagging feature and the log. To read more about this process, see Chapters 2.2.1 and 6. The transaction log was organized as a database including five tables:

TagAppliedLog includes all tags that are applied to the system:

- TimeStamp
- Tag
- SessionId
- UserId
- Url
- UserIP

The table has 25,253 rows, each representing an applied tag with accompanying data. Thus, the number of rows represents the total number of tags applied through the tagging feature.

PageViewLog includes all pageviews on Cancer.dk:

- TimeStamp
- SessionId
- UserId
- Tag
- PageUrl
- UrlReferrer
- UserIP

The table has 11,704,070 rows, each representing that a user viewed an article with an URL in a web browser. The number of rows is thus the total number of times users viewed the pages on Cancer.dk in the logging period.

SearchLog includes all searches using the local search feature at Cancer.dk:

- QueryId
- TimeStamp
- SessionId
- SearchType
- SearchString
- NumberOfResultats
- UserIP
- Category

The table has 375,718 rows, each representing a search submitted to Cancer.dk, or more specific: a click on the search button on Cancer.dk.

SearchResultLog includes all search results from the local search feature at Cancer.dk:

- QueryId
- ResultPageUrl

- ResultPageTitle
- isTagResult
- UserIP

The table has 99,111,915 rows. Each row represents an article retrieved through the search feature on Cancer.dk. Thus, each search resulted in a number of rows in this table, depending on the number of articles retrieved. The data in this table was not analysed in this thesis.

TagAdministrationLog includes editing and deleting tags from articles at Cancer.dk. Instances where a page with tags was deleted were not included, only tags that were edited or deleted from a remaining page:

- TimeStamp
- Action
- OriginalTag
- AlteredTag

The table has 8,574 rows, each representing the treatment of a single tag. Thus, the number of rows show how many tags were edited or deleted, except for those that were deleted with a page.

I sometimes refer to the tables by their names, to clarify what part of the transaction log to which I refer. Sometimes, I regard this to be clear from the context and the location in the text.

The transaction log contains more data than what was analysed for this thesis. I have focused on the tags and their meanings, and the facets they represent. This also means that the data in focus naturally falls in two categories: before and after the feature change, because the tagging behaviour could be influenced by the change.

7.1.3 LIMITATIONS

Unfortunately, the log does not have unique id-numbers for unique tags. This should be a basic property of the log, so obvious that the failure was not discovered until it was too late to fix it. Thus, some analyses were not possible or more difficult than expected. Examples are combinations of the tables above. TagAppliedLog includes all the tags with all its fundamental data. When a tag appears in PageViewLog, however, it is impossible to identify directly which tag in TagAppliedLog was clicked on to view a page. Because the same word or phrase can be applied as a tag many times, the tag itself is not enough to identify a tag. An id is needed. The timestamp that shows when the tag was applied can sometimes be a help. No users can click on a tag that had not been applied yet. But this does not help identify all tags. The same

challenge appears if you want to combine TagAppliedLog with SearchLog, SearchResultLog or TagAdministrationLog. On the other hand, as many of my analyses are based on categorization of tags from one of the tables (mainly TagAppliedLog); they were conducted with no such challenges.

The different tables contain data from slightly different periods. This is because the logging started before the tagging feature was launched. Likewise, the tagging feature was taken down just before the logging ended. Therefore, when I compare data from different tables, I regard the data collecting period as the period when the tagging feature was active. This means the time from the first to the last row in the table TagAppliedLog. This was especially important in chapter 7.2.8 about popular pages.

I accessed the transaction log through SQL Server Management Studio and searched it using SQL. When it was considered convenient, data was extracted and further manipulated in Excel. Categorization into facets and aboutness-categories are examples of such manipulation.

7.1.4 CANCER.DK AS A SITE TO STUDY

The documents I study are articles at Cancer.dk. They are units of HTML documents that appear on screen in your browser when a certain URL that equals or start with www.cancer.dk is downloaded by a user. The content is usually text concerning different topics related to cancer.

I use both article, page and URL to refer to the same unit of (mainly) text presented as a page at Cancer.dk. This is a shortcut. The article that appear on screen is what the users apply tags to. There is usually an equivalent relationship between this article and its URL. But a few times, an article is represented by two URLs. It may have changed URL during the logging period, or two URLs refer to the same article for other reasons, defined by the CMS that the Danish Cancer Society use to produce Cancer.dk.

Cancer.dk is an information site. For such sites, Kalbach states that navigation is critical for success. It is important to follow common design practices because the target groups are diverse. This reduces the energy users have to wase searching for information, and hopefully let them concentrate on the actual information (2007).

Cancer.dk is a well-organized site. A set of user studies from 2008 showed that users were satisfied with the site and its structure (Rasborg & Nielsen, 2008; Rasborg & Toft, 2008; Userneeds, 2008a, 2008b, 2008c). Some challenges were revealed, and the site was changed accordingly. Cancer.dk is thus a site where the basic structure is in place. Tags are not needed to provide this but can be applied as a supplement or an alternative.

7.1.5 VARIABLES AND TAG CATEGORIZATION

The tags in TagAppliedLog were the focus in the study of tags applied to articles on Cancer.dk. All tags that were applied during the logging period were grouped according to four different categorizations:

1. Internal or external tagger: Was the tag applied from inside or outside the Danish Cancer Society? All tags can be categorized into these two groups, based on IP numbers in the transaction log. Details and results can be found in chapter 7.2.3. The internal tags originate from internal users, employees of the Danish Cancer Society. Some of them are editors and are responsible for some of the content on Cancer.dk. A few of them, four persons, belonged to the central editing board of Cancer.dk. Other internal users do not work directly on Cancer.dk, but this is still the home page and an important information channel of their employer. They received information about the tagging feature through their job and were informed about the intentions behind the feature. External tags originate from external users of Cancer.dk. They also had access to information about the tagging feature and its purpose. But they do not share the same professional relationship to Cancer.dk as the internal users do. They are members, patient, clients. The different roles these two groups of users have towards Cancer.dk, makes it interesting to distinguish between them when analysing the tags.
2. Lay or professional vocabulary: Tags were categorized into five groups based on whether they belong to a lay or professional medical vocabulary, or neither of them. See details and results in chapter 7.2.10. Tagging has been seen as a way for ordinary people do describe documents in a way that suits them. It has also been seen as a way to include lay vocabulary into document descriptions. This was also part of the motivation for the Danish Cancer Society when they launched the tagging feature on Cancer.dk. Thus, I find it interesting to analyse the tags and see what kind of vocabulary they include. When comparing these categories with the previous internal and external tags, I can compare the external amateur taggers and the internal professional taggers with the resulting lay and professional tags and see what patterns appear.
3. Facets: Tags were categorized into 36 facets according to their topical meaning. See details and results in chapter 7.2.11. Topical facets were also used in the study of blog tags, see chapter 5.2. The same facets were used as a starting point here, with a few new facets (see Table 12). The facets reveal activity and interest in different topics.
4. Aboutness: Tags were categorized into 11 categories based on how the meanings of the tags refer to the content of the articles they are applied to. A discussion on aboutness can be found in chapter 3.1.1. Details and results are in chapter 7.2.12. Topical aboutness is crucial in traditional subject indexing, and also important to users and editors when interviewed about tags (see chapters 8 and 9). But especially users also have other purposes than

indexing when they apply tags. The aboutness categorization reveals how the tags really relate to the content of the articles. Lancaster discusses aboutness and refers to various attempts to define this concept (2003b, p. 15). He asks:

> Is it not enough to be able to recognize that a document is of interest to a particular community because it contributes to our understanding of topics X, Y, and Z? The recognition that it does contribute in this way exemplifies the process we have called "conceptual analysis", while the process of "translation" involves a decision on which of the available labels best represent X, Y and Z."

In this setting, aboutness categorizing means to evaluate whether a tag is a suitable label to represent how an article on Cancer.dk can contribute to its users. The aboutness categories are types of outcomes from such an evaluation. In other words, in what way does the tag refer to the topical content of the article?

The same categories were used for tags applied both before and after the feature change. Thus, the categories can be used to find out if and how the change influenced tags and tagging.

The categorization of internal/external tagger is based on IP-numbers and thus an easy operation. The three last categorizations required intellectual judgements. With 25,253 tags, assistance was needed. I developed the four categorizations, made examples and started the work. Then an assistant did all the lay/professional categorization, approximately $^{2}/_{3}$ of the facet categorization and approximately ½ of the aboutness categories. Especially the aboutness categorizations were time consuming, because it was necessary to open all URLs to decide whether a tag related to the content of the article or not. The assistant, Rie Bing, was a student of the bachelor programme Communication and Digital media at Aalborg University when she started to categorize in June 2013. In the autumn, she attended the master program in Communication. She finished her part of the categorization in November 2013. Oslo and Akershus University College of applied sciences paid her for the work. Bing is native in Danish and thus competent to assess Danish tags.

7.1.5.1 Procedure of categorization

The categorization was done in an Excel file with tags and all other data from TagAppliedLog. For each of the categorizations that the assistant did (lay/professional, facet, aboutness), she received the file. It included the categorization that I had already done, and a description of the categories with example tags. We also had a Skype meeting where I introduced the categories and the examples. For lay/professional and facet categorizations, tags were sorted

alphabetically, so that tags with a similar beginning would appear near each other. Thus, tags with different grammatical endings would be seen together and put into the same category when appropriate. For the aboutness

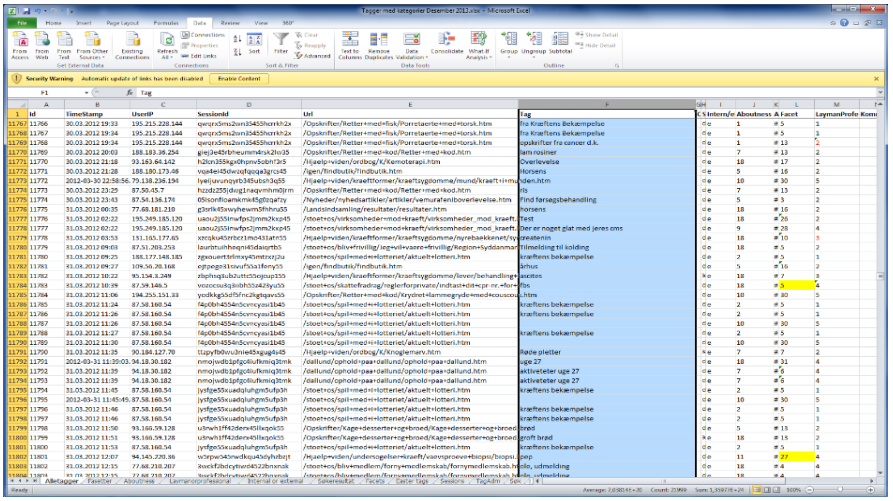


Figure 8 Tags viewed in Excel, for categorization. The column with tags in blue, the categories represented by numbers and codes in the following columns.

categories, the file was sorted based on which URL the tags were applied to. Thus, it was only necessary to look at each URL once and then categorize all tags applied to this URL. This saved time and ensured that the tags were categorized based on the relationship to the articles, not their relationship to other tags in an alphabetical list. This is also in correspondence with the categorization of Blogomkraeft.dk, where tags were interpreted based on the blog posting to which they were applied. The lists of categorized tags were later sorted in other ways for quality control.

All categories were made before the categorization started, but open to minor changes. The facets were already used for the blog tags, see chapter 5.2, but a few new facets were needed. See chapter 7.2.11 for details. I started all categorizations and chose some examples for the assistant to learn the categories. When she had categorized a number of tags and felt thet she knew the categories well, we had another Skype meeting to discuss tags and their categorization. When she finished, we had a meeting again. Professor Marianne Lykke also attended most of these meetings. In the meetings, we discussed the categorization and adapted the description and examples for each category, to make sure that our understandings were as similar as possible.

The file that we used for categorization included all instances of tags. For the internal/external and aboutness categorizations, each instance of a tag was categorized. The tag *bivirkninger* (side effects) applied to one article may coin the content of this article precisely. However, the same tag could be totally irrelevant when applied to another article. Thus, each instance of tags had to be categorized independently. The lay/professional and facet categorizations are different in this sense. The tag *bivirkninger* (side effects) is regarded as neutral in a lay/professional categorization. This is the case for all instances of this tag. In most cases, the same tag is also regarded as belonging to the same facet, *Symptoms and side effects*. Homonyms represent an exception here. We sought to categorize each meaning of homonyms independently.

In general, we attempted to neglect typing mistakes. If it was possible to identify a real word, we categorized it as if there were no typing mistake(s). The categories and the categorization were developed based on our best judgement of each word's origin and use.

When in doubt, article content, similar tags or articles with similar content were used in the categorization process. In addition, the following sources were useful in a search for definitions and explanations:

- Cancer.dk itself
- Den danske ordbog, http://ordnet.dk/ddo. This dictionary gives spelling and short definitions of Danish words.
- Danish Wikipedia, http://da.wikipedia.org/wiki/Forside and even simply searching the web (Google) gives indications as to whether the tag in question represents the name of a company, a place, a medical drug, etc.
- Store norske leksikon http://snl.no/ and Store medisinske leksikon http://snl.no/.search?e=sml gives help when the tag in question is a medical term. These encyclopedias are Norwegian, so they are sometimes used in combination with the other mentioned sources, to make sure the correct word is identified.

Unknown words were searched for in the sources, always bearing in mind that even the least meaningful term could be the name of a new drug or something unfamiliar to researchers with no medical background, or even the name of a person or a place.

If no meaningful explanation to the tag was discovered, it was categorized as *Number or code*.

With so many tags, it was necessary not to spend too much time on each tag. I think our solution made it possible to obtain a broad picture based on a high number of tags. Additionally, if the categorizations still have some inconsistencies or mistakes, this will not interrupt this broad picture.

7.1.5.2 Tag facets on Blogomkraeft.dk and Cancer.dk

The categorizing process itself revealed differences between Blogomkraeft.dk and Cancer.dk tags. When using the categories on Blogomkraeft.dk tags, the distinction between categories was clear. This changed when categorizing Cancer.dk tags, as it became harder to distinguish between some of the categories. This was partly due to the tags and their meaning, and partly due to the context given in Blogomkraeft.dk. When you know tags and postings were written by the same person, it is easier to find out what kind of word the tag really is. The tags at Cancer.dk were also applied to documents, and thus they have a context. But the tagger's view on the relationship between tag and article is not as clear and thus not as helpful when categorizing.

On Blogomkraeft.dk, the bloggers separated quite clearly between body and body parts and a cancer form located in a certain body part. They also separated between a cancer form and its treatment. On Cancer.dk, these subjects are closely related in the articles, and in the tags. An example is that within a cancer information context, *bryst* (breast) and *brystkræft* (breast cancer) may represent the same topic or are very closely related. The same kind of relationship occurs between treatment and the method or substance used for this treatment. Examples are *stamceller* (stem cells) and *stamcelletransplantation* (hematopoietic stem cell transplantation). The reason is that the articles all have a focus on the diseases. At Cancer.dk, cancer diseases are the cause of all information and activity, and this is very visible on the site. On Blogomkraeft.dk the focus is more on the individual stories. The bloggers could thus write a posting about treatment and side-effects, without mentioning the cancer form or its location at all.

There are many differences between the tags applied to postings in Blogomkraeft and tags applied to articles at Cancer.dk. See also Chapter 5.2. As mentioned above, the tagger at Blogomkraeft.dk was always the same person as the one who wrote the posting. At Cancer.dk this is not the case, with the exception that some editors probably applied tags to their own articles. It is not known how many times this happened, but the number of internal tags is limited.

Another major difference is the selection of taggers. The bloggers were selected by the Danish Cancer Society. Most of them posted several postings each and gained experience with the system they used, including how to apply tags. The users of Cancer.dk are not selected; anyone who accessed the site could apply or attempt to apply tags.

7.2 RESULTS

7.2.1 GENERAL ACTIVITY

The Danish Cancer Society uses Google analytics to acquire statistics on the use of Cancer.dk. Figure 9 shows these numbers.

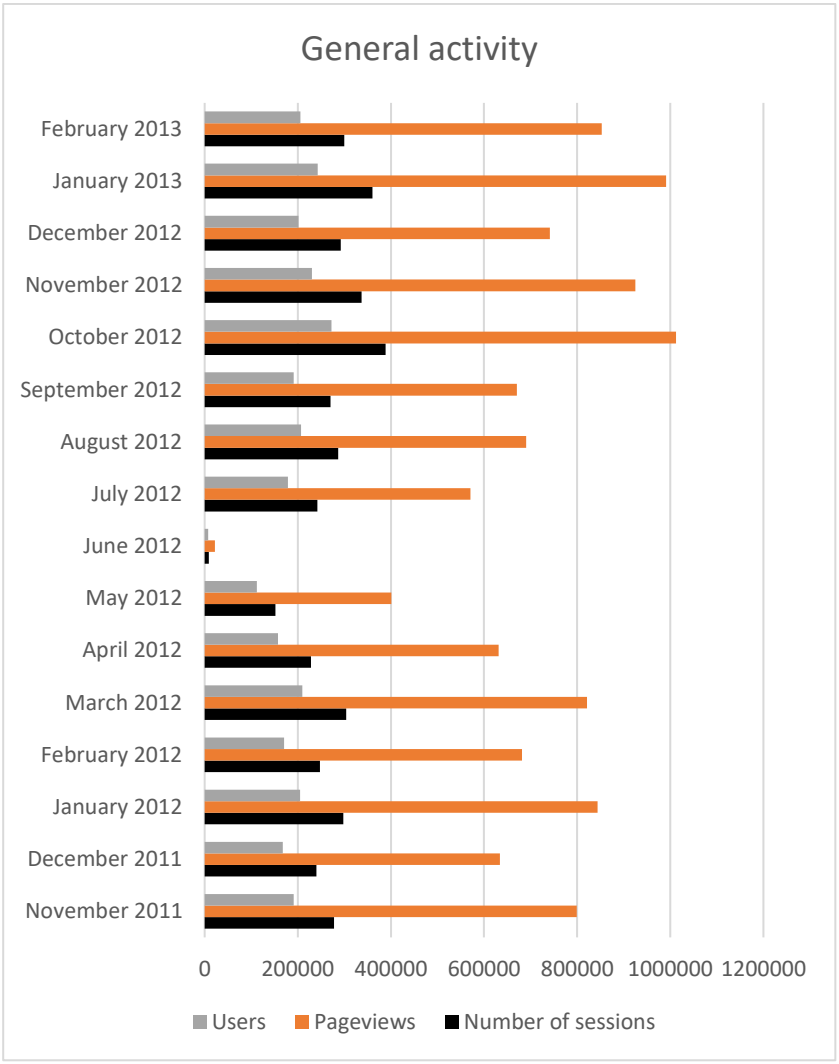


Figure 9 Use of Cancer.dk during the logging period, with the number of users, pageviews and user sessions for each month.

7.2.2 TAGGING ACTIVITY

Month	Tags	Sessions	Tags per session	Empty tags	Unique tags	Tagged URLs	Unique tag/url combinations	Tags per Url
Nov.11	992	150	6,61	16	590	300	911	3,31
Des.11	2502	1777	1,41	315	1323	599	1753	4,18
Jan.12	2740	2033	1,35	376	1460	727	1981	3,77
Feb.12	2319	1638	1,42	331	1279	645	1667	3,6
Mar.12	3287	2158	1,52	393	1622	729	2184	4,51
Apr.12	2719	1454	1,87	253	1466	853	2156	3,19
Mai.12	3354	2423	1,38	452	1973	660	2387	5,08
Jun.12	2412	1679	1,44	357	1254	654	1630	3,69
Jul.12	1655	1206	1,37	275	868	518	1185	3,19
Aug.12	2025	1410	1,44	324	1032	577	1343	3,51
Sep.12	721	499	1,44	109	436	310	529	2,33
Okt.12	38	29	1,31	5	27	29	32	1,31
Nov.12	134	113	1,19	16	110	91	126	1,47
Des.12	200	75	2,67	24	135	61	155	3,28
Jan.13	94	82	1,15	12	79	73	90	1,29
Feb.13	59	52	1,13	2	57	40	59	1,48
Average, Jan-Aug 2012	2557	1753,11	1,47	341,78	1364,11	662,44	1809,56	3,86
Average, Oct-Dec 2012	105	70,2	1,49	11,8	81,6	58,8	92,4	1,76
Average	1578,19	1048,63	1,47	203,75	856,94	429,13	1136,75	3,07
All year	25253	16774	1,51	3260	9989	3024	15902	8,35

Table 2 Tagging activity per month. Averages of tagging activity per month from the tagging feature was well established and to the feature change, and for the months after the feature change. Also average for all year, including those months excluded from the other averages.

The first tag was applied on November 8, 2011. The first tag from an external IP was applied on November 30, 2011. After 13 months, there were 25,253 tags in the transaction log. This number includes all tags, empty tags, repeated tags, and possible errors that the user seems to correct by adding another tag. In other words, it includes every time a user hit the *tilføj*-(submit)-button in the feature for applying tags. Table

2 summarises general tagging activity per month and in total. It also gives averages for the year 2012 before the feature change, and all months after the tagging feature change. The months are selected to give an impression on the activity after the first start. See Chapter 7.1.1 for details about the change.

Columns in Table 2 include the number of tags applied each month (Tags) and the number of user sessions where users applied tags (Sessions). The number of tags applied per session are calculated from Tags and Sessions (Tags per session). The number of empty tags is counted for each month (Empty tags). Empty tags occur when users hit the *tilføy* (submit)-button without writing anything in the tagging field first. Further, the table gives the number of unique tags (Unique tags). This refers to how many distinct words or phrases are applied as tags, independent of how many times. The number of URLs with tags are counted for each month (Tagged URLs), as is the number of unique combinations of tag and URL (Unique tag/URL combinations). This number each month is higher than the number of unique tags, because the same tag can be applied to different URLs. The number of tags per URL is calculated for each month, based on the number of tags and the number of URLs with one or more tags applied to them (Tags per URL).

An average URL with tags had 3.86 tags before the feature change. In total, 3,024 URLs were tagged with 25,253 tags, including the empty tags. There were 9,989 unique tags, and 15,902 unique pairs of tags and URLs.

7.2.2.1 2011 November-2012 August

After the first start, the number of tags stabilized to an average of 2,557 tags per month through January to August 2012. There were 84 tags per day, 11 of them empty ones. There were on average 39 pairs consisting of tag and URL that were unique each day.

During the summer, especially in July, the activity at Cancer.dk was lower than usual. Each day has an average of 53 tags per day. This corresponds to a general drop in Cancer.dk usage; this month has the lowest number of viewed pages in 2012, see Figure 9.

7.2.2.2 2012 October-December

After the feature change in September 2012, the tagging feature was less visible. The number of tags applied to the system now went down to 105 per month, a dramatic change, and the number of tags per URL per month was down to under two. At the same time, the average number of tags per session where tags were applied, remained stable: 1.47 before and 1.49 after feature change. Only one tag was applied from internal taggers after the change, see Table 4.

7.2.3 USER GROUPS – TAGS FROM INTERNAL AND EXTERNAL USERS

The Danish Cancer Society have a central group of editors who have the daily responsibility for Cancer.dk. In this thesis, they are generally referred to as editors (for reports from the interviews with them, see chapter 9). There are also decentralized editors in different divisions of the organization, who are responsible for the content on a limited number of URLs at Cancer.dk. Both editors in the central editorial office and the decentralized editors applied tags to articles at cancer.dk. Also, other employees of the Danish Cancer Society had access to and were encouraged to apply tags.

The transaction log does not separate the central and decentralised editors from other users. But IP-addresses can be used to distinguish between internal (from within Danish Cancer Society) and external users. A clear weakness is that if internal users work from home, they appear in the log as external users. When I asked the editors about their use of Cancer.dk, they reported that they did not work from home, so this is a potential problem that concerns the decentralised editors and other employees of the Dansih Cancer Society. If they applied tags from home, their tags are categorized as external in this study.

To my knowledge, nobody from outside the Danish Cancer Society had access to use Cancer.dk from IP-addresses that belong to the organization. One exception is users who participated in the usability study (see chapter 8). Two of them used Cancer.dk from the Danish Cancer Society's building and network in Copenhagen, during the test. I identified their activities manually in the log and categorized them as external users. The other participants accessed Cancer.dk through an Internet connection that belonged to Aarhus University. They are thus automatically categorized as external users.

The researchers only applied a few tags for testing purposes, less than five. These are not used as examples or analysed as individual tags, but are a part of the general material, and I found it most correct to categorize them as internal tags.

In sum, there were two groups of users and tags: Internal users are users that had access to use Cancer.dk from IP-adresses that belong to the Danish Cancer Society, or its search engine and CMS vendors, or IP-addresses that belong to the researchers. Internal tags are tags applied by internal users. Accordingly, external users and tags are the remaining users and tags after identifying the internal ones.

I do not know how many users applied the tags in the transaction log. But I do know the number of distinct IP addresses. Each IP address can be used by one or many users, but this is the case for both internal and external addresses. Thus, I find it fair to use the percentage of distinct IP addresses as an indication of the number of users. Measured this way, the internal users are a very small minority. Still they applied 9.14% of the tags. The overall numbers of internal and external tags can be found in Table 3. It shows how many tags were applied (Tags), and the distribution on internal

and external tags in per cent (Tags %). The table further gives the number of distinct IP-addresses in the log (IP) and their distribution on internal and external addresses in percentage (IP %)

	Tags	**Tags %**	**IP-adresses**	**IP-adresses %**
External users	22.946	90,86	13.978	99.96%
Internal users	2.307	9,14	5	0.035%
Sum	25.253	100	13.983	100

Table 3 Internal and external tags and users

Only one internal user applied one tag after the feature change. Thus, internal taggers share of the tags go from 9.14% (see Table 4) to almost zero.

Table 4 shows the number of tags from internal and external users (External tags and Internal tag) before and after the feature change on 2012, September 11. On this date, 37 tags were applied. This number is between the average number of applied tags per day before and after the feature change. I do not have the exact time for the change. Thus, the tags are put in a separate category in Table 4, because it is unknown which version of the feature they were applied to.

	External tags		**Internal tags**		**Sum**
	Number	Percent	Number	Percent	Number
Before feature change	22334	88.44	2306	9.13	24640
Day of feature change	37	0.15	0	0	37
After feature change	575	2.28	1	0.01	576
Sum	22946	90.86	2307	9.14	25253

Table 4 Number of external and internal tags before and after the feature change. Tags from the day of the feature change are separated from the other tags.

7.2.4 APPLIED TAGS PER SESSION

The average number of tags per session in the transaction log was 1.51, all clicks on the button to submit tags included. Overall, the number of sessions where users

applied tags is stable in the period before the feature change, with a decrease in the summer months, July and August. See Table 2.

Month	Applied tags per external session	Used tags per external pageview	Applied tags per internal session	Used tags per internal pageview
Nov 2011	1.43	0.06	10.80	0.24
Dec 2011	1.35	1.37	4.79	0.54
Jan 2012	1.34	1.81	2.31	0.47
Feb 2012	1.38	1.84	4.04	0.23
Mar 2012	1.40	1.33	11.63	0.16
Apr 2012	1.39	2.01	24.47	0.26
May 2012	1.38	2.20	1.25	0.28
Jun 2012	1.44	2.57	1.75	0.13
Jul 2012	1.37	3.18	1.17	0.39
Aug 2012	1.44	2.81	1.59	0.18
Sep 2012	1.45	3.23	1.00	5.26
Oct 2012	1.31	1.30	0	0.43
Nov 2012	1.19	3.13	0	0.20
Dec 2012	2.67	2.70	0	0.20
Jan 2013	1.15	2,56	0	0.13
Feb 2013	1.13	2,00	0	0.11
Average Nov 2011-Aug 2012	1.39	1.92	6.38	0.29
Average Oct 2012-Feb 2013	1.49	2.34	0	0.21
Average	1.42	2.13	6.03[12]	0.58

Table 5 Applied tags per session, and used tags per page view, from internal and external users.

[12] The number is the average number of applied internal tags per session with tags, from November to September 2011. October 2011-February 2012 are not included, because no internal tags were applied those months.

Distributed on internal and external users, the number of tags per session per month vary; see Table 5 and Figure 10. For each month, they show the number of internal and external tags applied, as a percentage of the total number of internal and external sessions (Applied tags per external session and Applied tags per internal session). Table 5 also shows external and internal page views, see chapter 7.2.5 for details.

The external taggers showed a stable number of tags per session, 1.42 during the whole period. There was a peak in December 2012, with 2.67 tags per session. This month is an exception from a general trend that the number of tags per session decreased a little after the feature change.

Internal taggers applied on average 6.03 tags per session during November 2011-September 2012, before the feature change[12]. The first month had almost 11 internal tags per session. Employers of the Danish Cancer Society were encouraged to apply tags from the beginning. This probably caused the peak in number of internal tags per session in the beginning.

Then the number dropped for a few months and raised again to a peak in April 2012. This peak turns out to originate from two internal sessions where somebody applied 290 and 56 tags. This happened during Tuesday and Wednesday before the public holidays during Easter, on respectively April 3 and 4. If these sessions are excluded from the remaining log of tags, the number of tags per session is 13.85 in April 2012. This is not far from the previous month March with 11.63 tags per session, but still a high activity compared to other months. This is most likely as a part of a process of getting to know the tool for administering the tags. At one point, somebody deleted tags by accident, and later the Danish Cancer Society's staff tried to replace them as well as they could. It is not known how this was done, for example what sources were used to know exactly what tags applied to what URLs that needed to be re-applied. When asked, the editors did not know what had happened. So, either they forgot, or the replacement tags were applied by one of the decentral editors.

From May 2012, the internal taggers behaved more like the external ones, with a stable number of applied tags per session, until a drop again in September 2012 and then no more internal tags. Only one internal user applied one tag after the feature change.

It was probably easier to encourage internal users to apply tags than external users. It was easier to communicate directly to them; they were more likely to apply tags as part of their job. On the opposite site, when there was no recent encouragement to apply tags, they applied fewer tags.

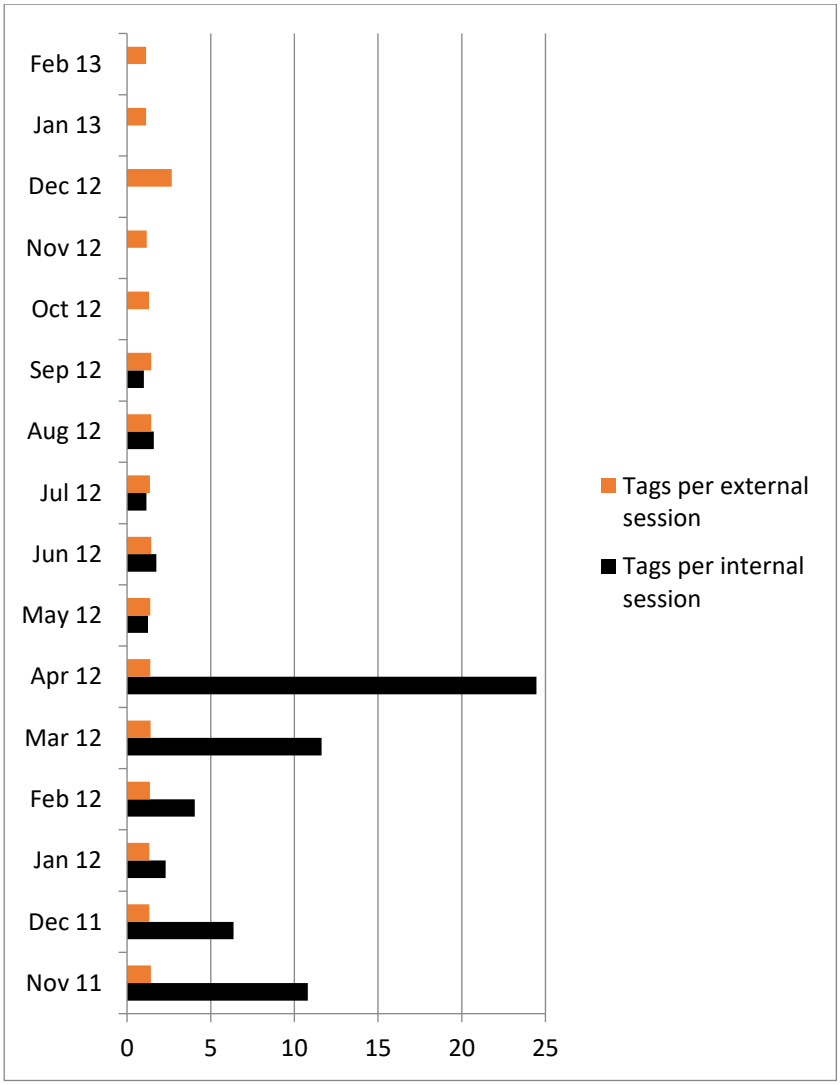


Figure 10 Number of applied tags per session, from internal and external users, distributed on months.

7.2.5 THE USE OF TAGS

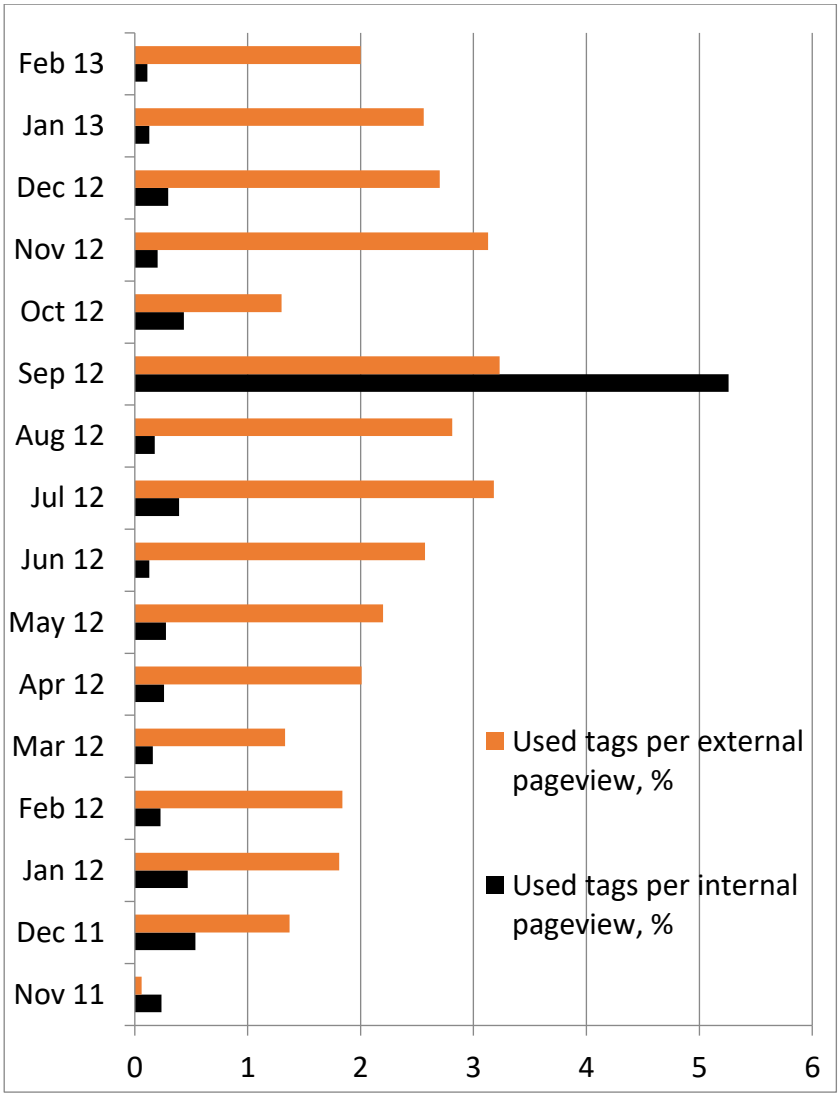


Figure 11 Number of used tags per page view, by internal and external users.

Most users did not apply or click on tags, and most user sessions do not include applying tags. Figure 11 and Table 5 show the numbers of external and internal page views where the users clicked on one or more tags, in percentage of the total number of internal and external page views (Used tags per external page view and Used tags

per internal page view). Table 5 also shows the number of applied tags per session, see chapter 7.2.4 for details.

In the logging period there were 4,816,300 sessions with one or more visited URLs (pageviews in PageViewLog). There are 16,774 sessions where one or more tags were applied. This means that in 0.35% of all sessions, the user applied one or more tags.

The use of tags, in Table 5 measured as the share of viewed pages where a tag was clicked on, shows a growth. This corresponds to an increase in the total number of tags in the system. The drops in tag usage in March and October may be results of a high number of deleted tags. See chapter 7.2.13 about tag editing. But there was also a high number of deleted tags in May, which did not result in a tag usage drop. Thus, the number of tags available in the system does not correspond directly to the use of tags; other factors also influence the use of tags.

7.2.6 NUMBER OF URLS TO WITHCH USERS APPLIED TAGS

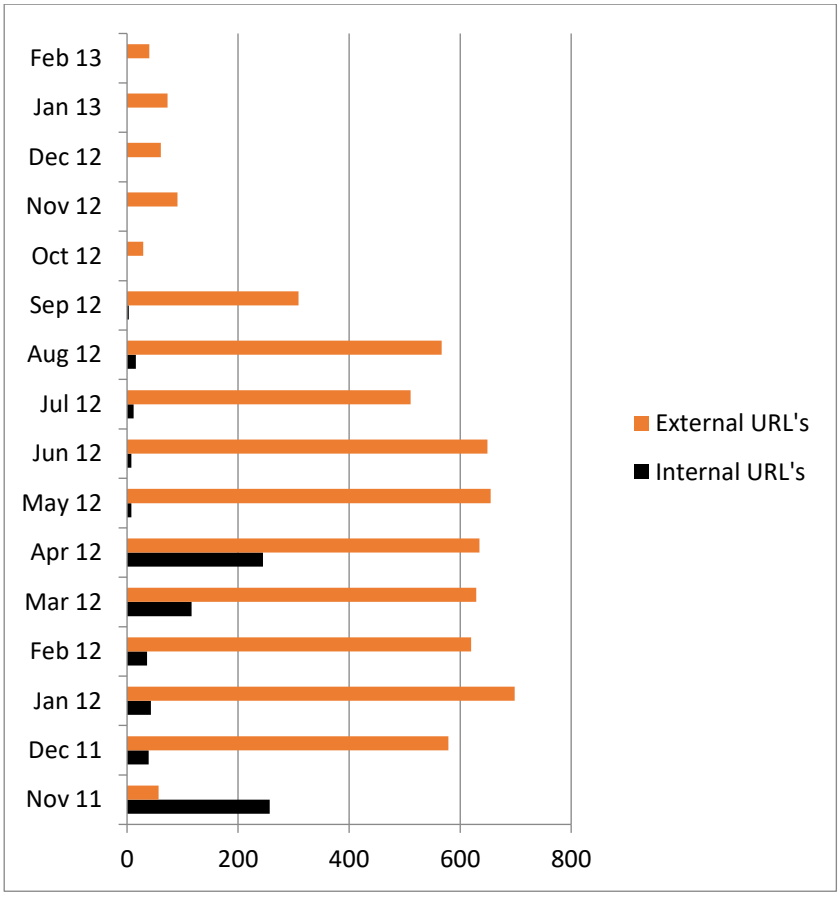


Figure 12 Number of distinct URLs with taggs applied to them

Figure 12 shows the number of URLs to which users applied tags. For external users, this number was quite stable before the feature change. After the feature change, it is not easy to characterize the stability, due to lower numbers of tagged URLs. Chance may therefore play a role in this period. For internal users the numbers vary. In Nobember 2011, there was a peak, as well as in March 2012 and April 2012.

7.2.7 APPLIED TAGS PER ARTICLE

In total, there were 8.35 tags per URL with at least one tag. Because the number of articles is relatively stable and the number of tags increased, the average number of

applied tags per article per month was lower, 3.07. Figure 13 shows the average number of applied tags per distinct URL per month, applied by external and internal users. In other words, the number of tags per month and the number of distinct URLs to which these tags are applied to within the month, are used for the calculation in Figure 13. In general, the number vary from 1 to 5.11 from each group of taggers (note that URLs with zero tags are excluded) per month. Approximately 1/3 of tags were deleted (see chapter 7.2.13 about tag edititn), but the remaining 2/3 of tags accumulated in the system. Thus, the number of tags visible for the users on Cancer.dk was higher than what is shown in the figure.

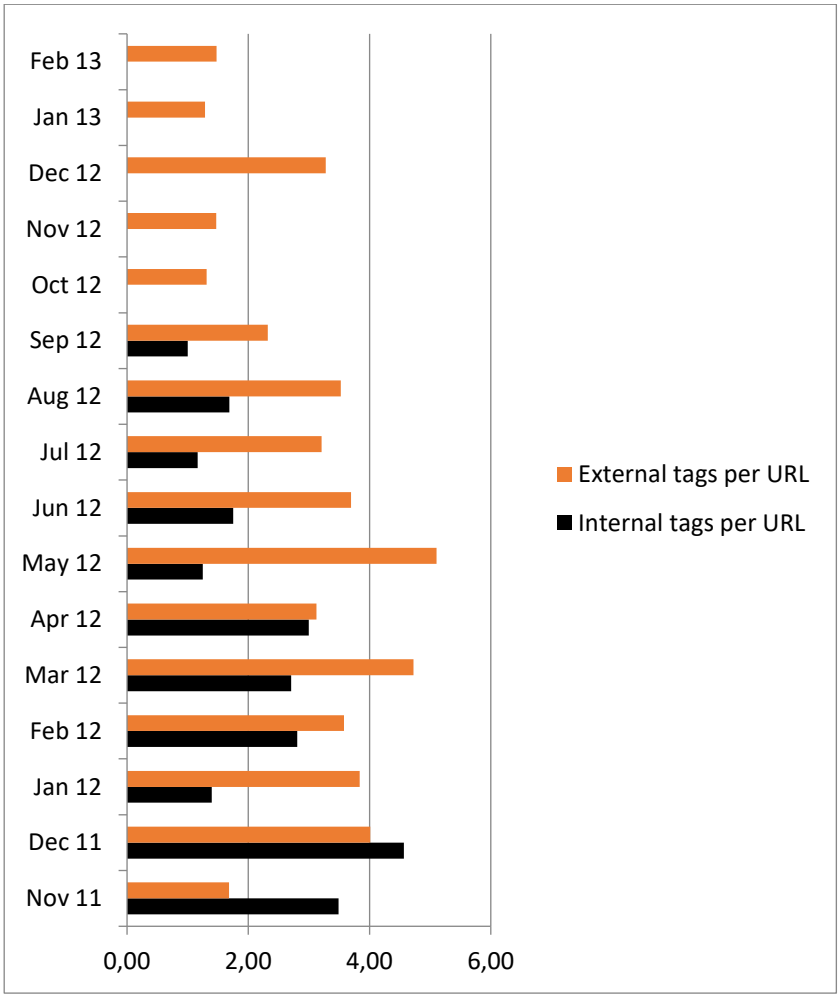


Figure 13 Applied tags per URL, per month.

7.2.8 POPULAR PAGES

The figures in this chapter (Figure 14, Figure 15, Figure 16 and Figure 17) show applied tags and visits on Cancer.dk sections based on the URL of pages. Numbers are taken from TagAppliedLog and PageViewLog. To form sections, the URLs are grouped to sections based on the beginning of the URL. For example, the URL: www.cancer.dk/Hjaelp+viden/fakta+om+kraeft/udvikling+af+kraeft/udvikling+af+kraeft.htm belongs to the hjaelp+viden-section of the site.
And the URL: www.cancer.dk/skrunedforsolen/UV-Index/Frameld/Frameld+dig+Dagens+UV-index.htm belongs to the skrunedforsolen-section of the site. Thus, the sections are based on the beginning of the URL, the first part of the path in the address.

The number of pages in each section varies. The number of unique URLs in the transaction log tables also vary. TagAppliedLog has 3,024 distinct URLs, and PageViewLog has 620,922 distinct URLs. This last number is substantially higher, because the table includes individual visits where unique URLs were generated, including search results pages and landing pages. Those pages were not available for tagging and are thus not in the TagAppliedLog. To overcome the variation in the number of pages that are available for viewing and tagging, tags and visits are shared on the total number of URLs in each logging table and URL section. This is the numbers that can be seen in Figure 14 - Figure 17. But when comparing the figures, it is important to also pay a little attention to the scales of each figure.

Figure 14 and Figure 15 show that internal and external taggers applied tags to different sections of Cancer.dk. External taggers tend to apply tags to pages on sections related to organizational content. The five sections with most tags represent such sections (Figure 15):

- Landsindsamling (A nation-wide fund raising), 20 tags per page in section
- Stoet+os (Support us), 19.7 tags per page in section
- Pinkcup (A breast cancer fund raising campaign), 19.6 tags per page in section
- Igen (Name of the Danish Cancer Society's secondhand stores), 15.4 tags per page in section
- Om+os (About), 14.3 tags per page in section

The five sections with most internal tags show a different profile (Figure 14):

- Omsorg (care), 2.6 tags per page in section
- Ditliv (your life), 2.5 tags per page in section
- NR, 1.8 tags per page in section
- Rensluften (clean the air), 1.2 tags per page in section
- Hjaelp+viden (help and knowledge), 1.2 tags per page in section

These sections have more information on knowledge about cancer, how to prevent it or treat it. It seems like the internal taggers applied tags to provide information to the users, while the external ones apply tags to pages about to relate to the Danish Cancer Society. If tags are a way to communicate, the external taggers communicate with the system and with the Danish Cancer Society, while internal taggers communicate with the users by tagging information for them.

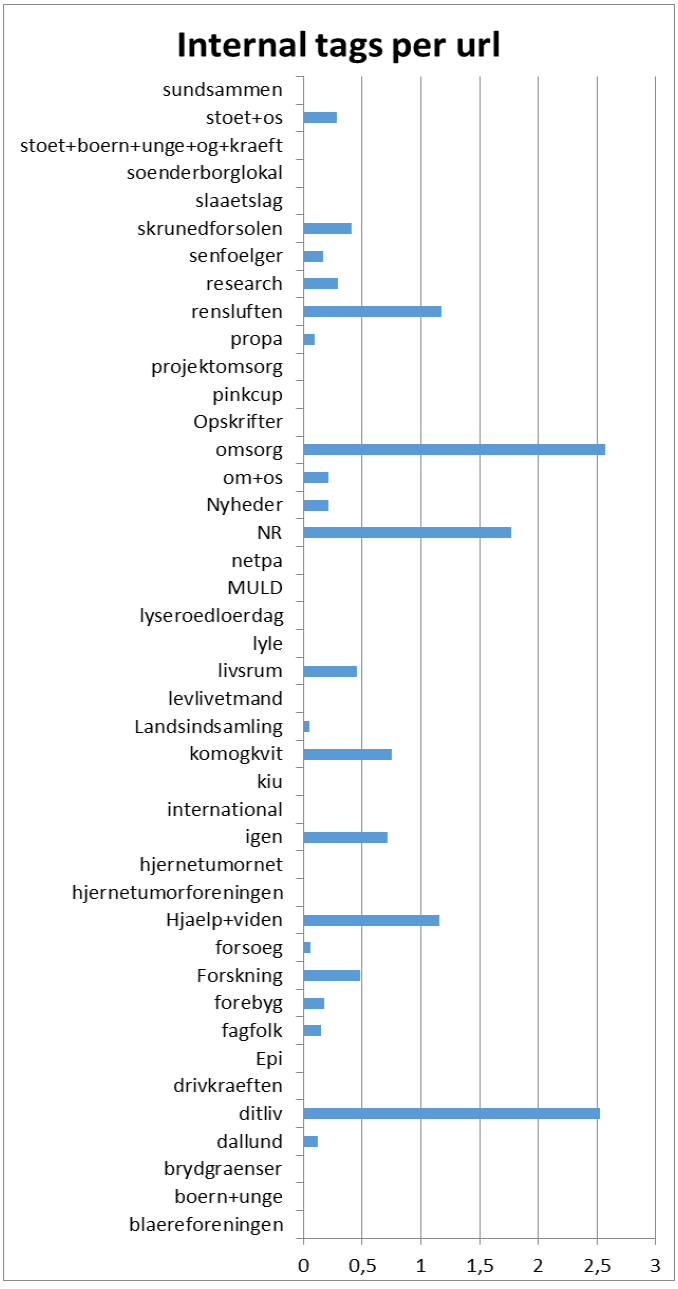


Figure 14 Internal tags per distinct URL in TagAppliedLog

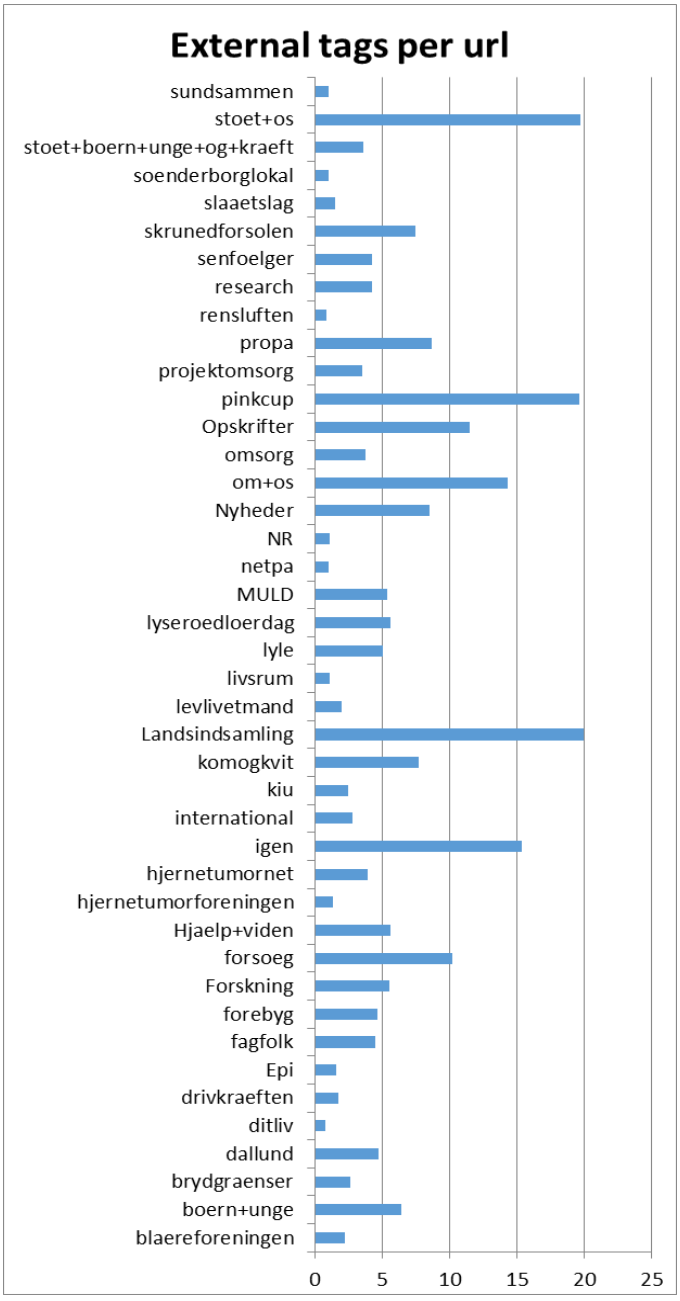


Figure 15 External tags per distinct URL in TagAppliedLog

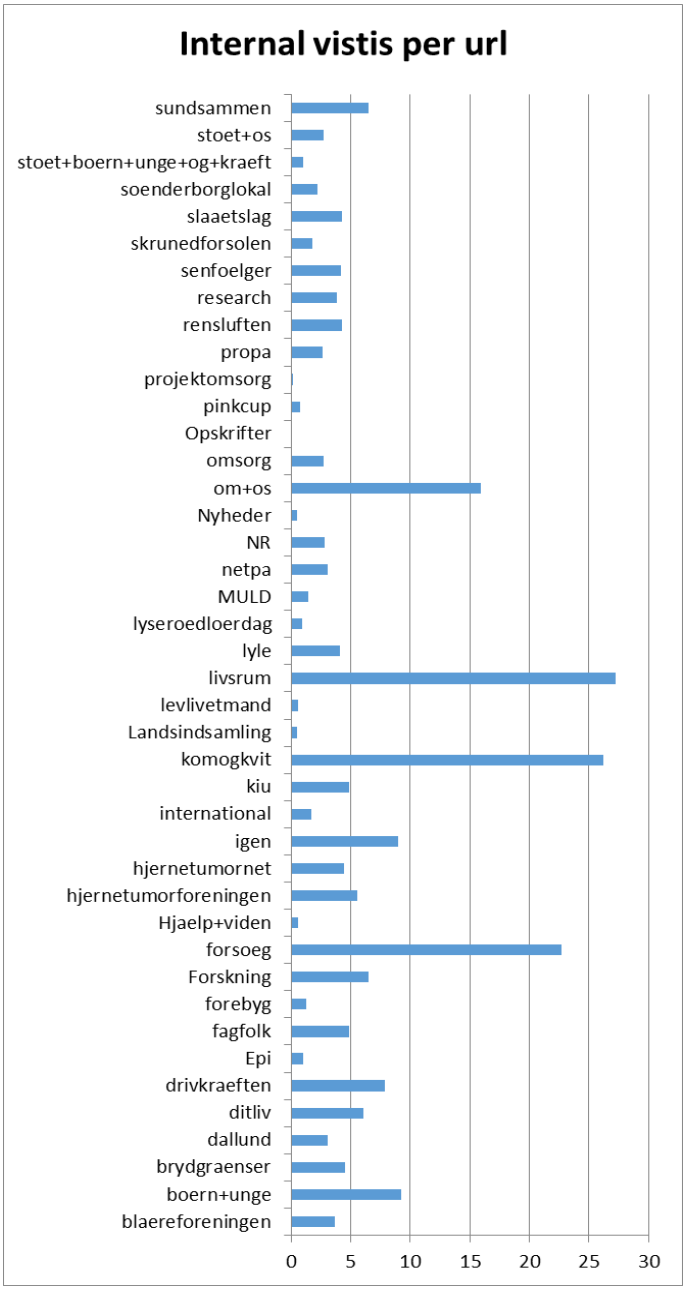


Figure 16 Internal visits per distinct URL in PageViewLog

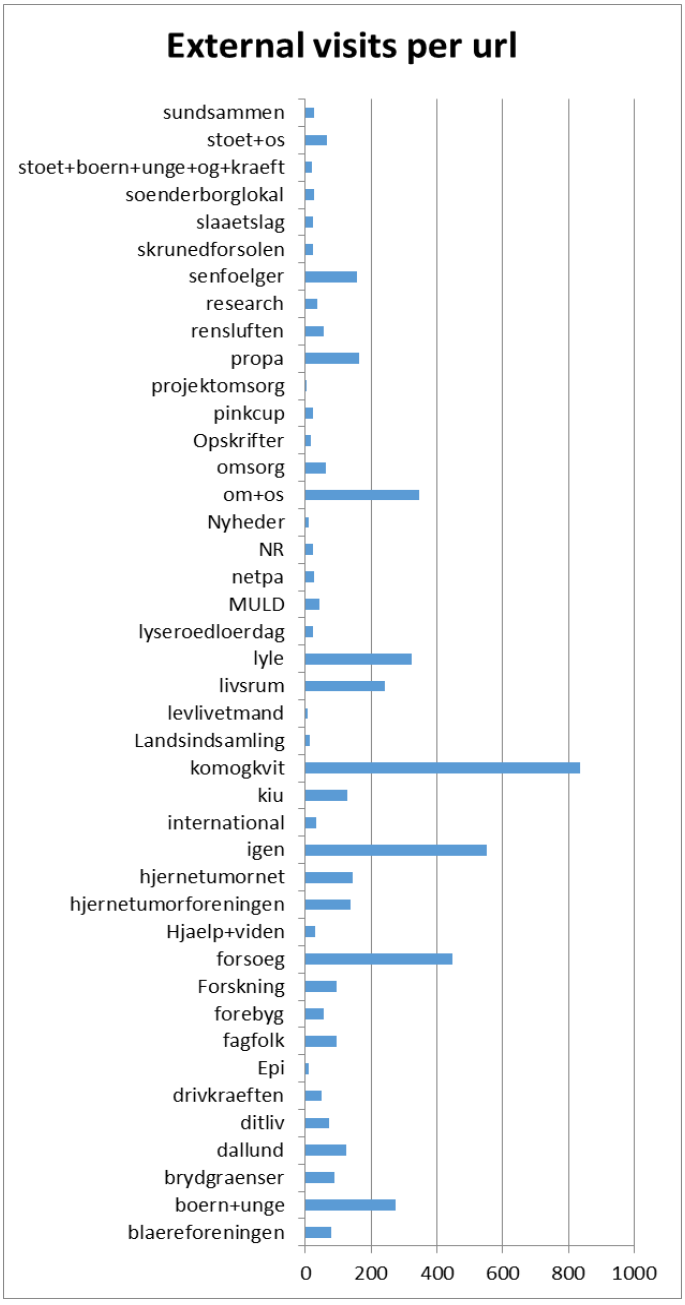


Figure 17 External visits per distinct URL in PageViewLog

7.2.9 EMPTY AND MEANINGLESS TAGS

Month	Total number of tags	Internal meaningless tags	Internal meaningless tags, share of total internal tags each month	External meaningless tags	External meaningless tags, share of total external tagseach month	Internal empty tags	Internal empty tags, share of total internal tags each month	External empty tags	External empty tags, share of total external tags each month
feb.13	59	0	0,0	26	44,1	0	0,0	2	3,4
jan.13	94	0	0,0	24	25,5	0	0,0	12	12,8
des.12	200	0	0,0	38	19,0	0	0,0	24	12,0
nov.12	134	0	0,0	55	41,0	0	0,0	16	11,9
okt.12	38	0	0,0	4	10,5	0	0,0	5	13,2
sep.12	721	0	0,0	54	7,5	0	0,0	109	15,2
aug.12	2025	3	11,1	117	5,9	1	3,7	323	16,2
jul.12	1655	0	0,0	119	7,3	3	21,4	272	16,6
jun.12	2412	1	7,1	189	7,9	1	7,1	356	14,8
mai.12	3354	0	0,0	1223	36,6	0	0,0	452	13,5
apr.12	2719	0	0,0	155	7,8	1	0,1	252	12,7
mar.12	3287	1	0,3	242	8,1	2	0,6	390	13,1
feb.12	2319	0	0,0	150	6,8	3	3,0	328	14,8
jan.12	2740	0	0,0	195	7,3	3	5,0	373	13,9
des.11	2502	0	0,0	152	6,4	1	0,7	314	13,3
nov.11	992	4	0,4	5	5,2	4	0,4	12	12,5
Sum	25253	9	0,39	2748	11,98	19	0,8	3240	14,1

Table 6 Empty and meaningless tags from internal and external users, per month.

2,757 of the tags, 10.9%, were numbers, codes or words with no identified meaning. 3,260 tags, 12.9%, were empty (see Table 2). All together, these tags represent 23.8% of the total number of tags. From an overall perspective, these tags are useless. They caused a need to edit tags on a daily or at least weekly basis. Of course, they were also easy to identify and thus usually easy to delete.

Internal and external taggers differ when it comes to empty tags and tags that consists of numbers or codes. External users applied 99.36% of the empty tags and 99.75% of the numbers, codes and meaningless words. For comparison, Table 3 shows that external users applied 90.86% of the total number of tags, a slightly lower share of tags. Thus, internal taggers applied less empty and meaningless tags.

Figure 18 shows the share of empty and meaningless tags applied by internal users. The numbers are very low, in total internal users applied only 9 meaningless and 19 empty tags. Table 6 show both the number of tags and the share of empty and meaningless tags, for each month, for comparison. The highest number of such tags per month was in November 2011, when the internal users first learnt to know and use the tagging feature. That month internal users applied four empty and four meaningless tags. But their total number of tags was high this month. Thus, the share as shown in Figure 18 is low. During the entire tagging period, 0.39% of the internal tags were empty or meaningless. Since the numbers are low, there may be no reason to look for other explanations on these tags than fortuitous accidents by the internal users.

Figure 19 shows the share of meaningless and empty tags from external users. Again, Table 6 includes both the number of tags and the share of empty and meaningless tags for each month. Here the total number of tags is high, external taggers applied 2,748 meaningless and 3,240 empty tags. There is a peak in the number of meaningless tags in May 2012. There is also an increase in the number of meaningless tags after the feature change in September 2012, but the increase does not start until November, so it is not clear wether the feature change caused this increase or not.

When the empty and meaningless tags were categorized, they were usually put in one or a few categories that were meant for these tags. This pattern simplified the process of categorization these tags. When it was done in one type of categorization, it was partly copied to the other ones. Exceptions can be found for instance when a tag includes a meaningless and a meaningful word. I paid a little more attention to those tags, in order to make sure they were categorized in the best way for each categorization.

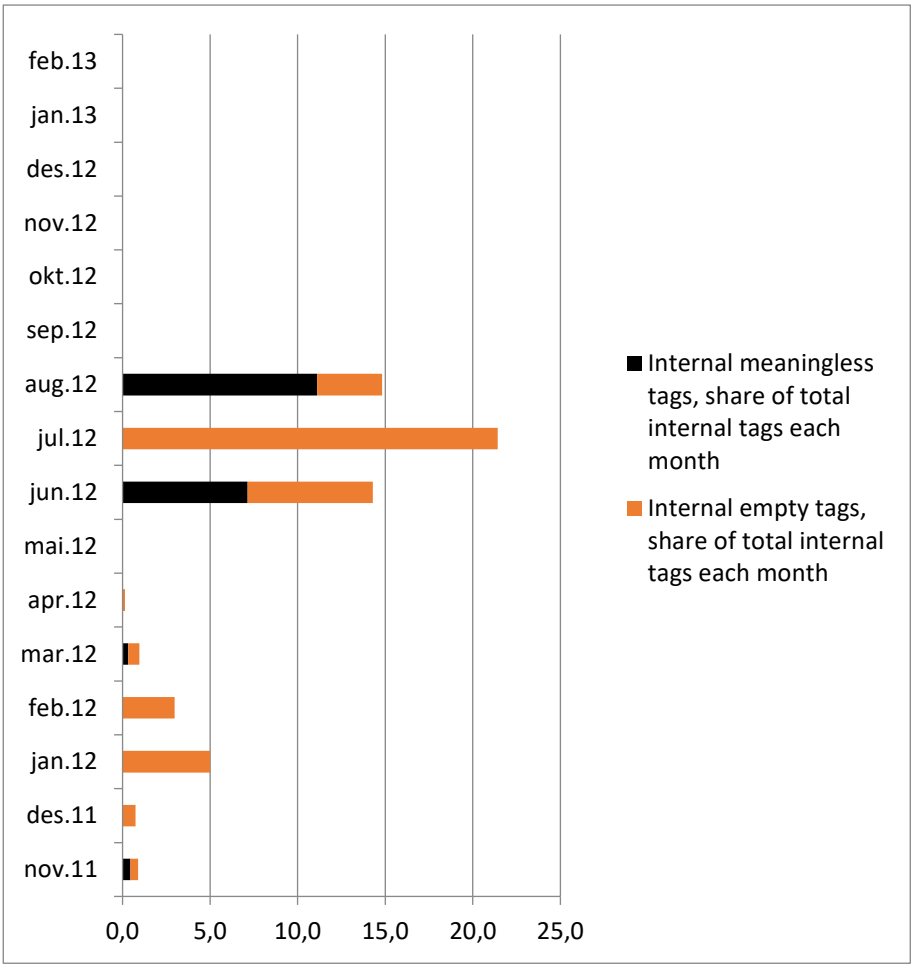


Figure 18 Empty and meaninless tags applied by internal users, in percent for each month. The total number of empty and meaningless tags applied during the logging period is 28.

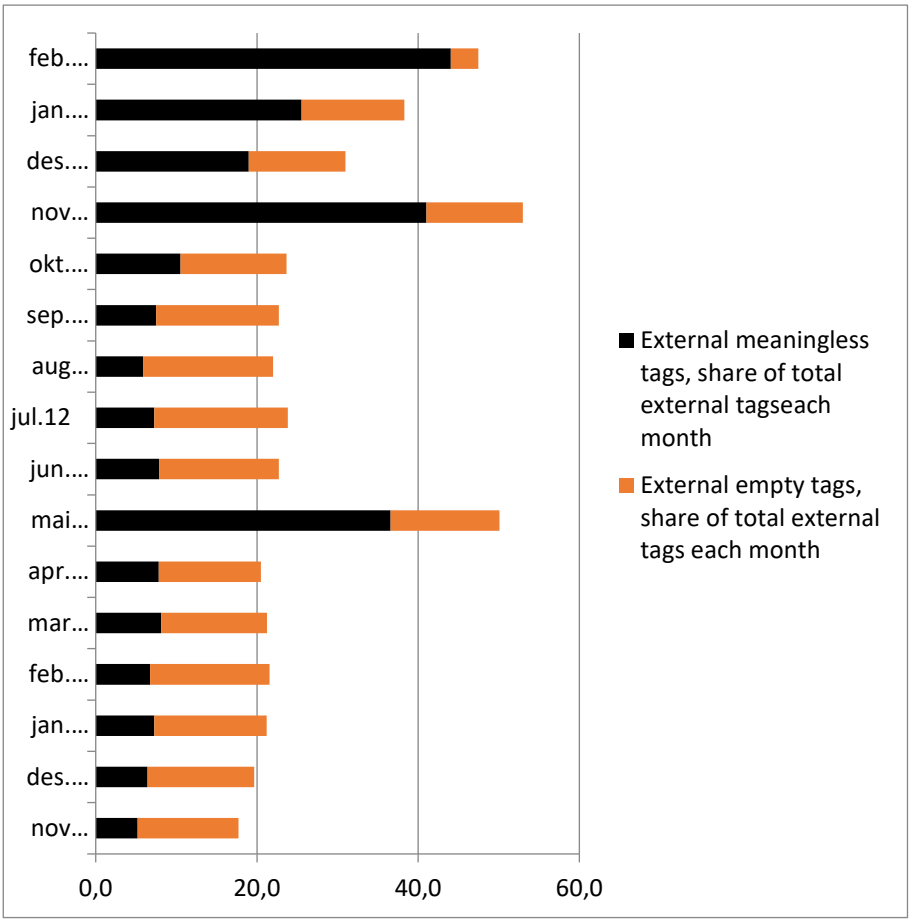


Figure 19 Empty and meaninless tags applied by external users, in percent for each month

7.2.10 LAY AND PROFESSIONAL VOCABULARY

All tags were put in categories according to how they can be characterized as lay, professional or neutral. The professional vocabulary is cancer medicine. The categorization is based on familiarity with Danish language and its use. No statistics of word usage among laypersons and health professionals were used. This is due to the fact that laypersons and health professionals tend to use each others' vocabulary (see also chapter 8.3.5 and Glenton (2006)). It was a goal to categorize the tags based on a combination of their meaning, origin and general familiarity, more than their use.

Five categories were used:

1. Words that belong to a lay vocabulary:

   Words concerning cancer or medicine in general that are Danish and easily understood for the general Danish public belong to this category. Compound words that are combinations of Danish and Latin/English were usually put here. Not only because they are lay compound words in their origin, but also because the Cancer.dk editors stated that they could not use these combinations in their texts. They regard them as wrong, or at least not optimal. These words may also be used by professionals, but there are often one or a few professional synonoms belonging in category 3.

   Examples: *Kræft* (cancer), *Brystkræft* (breast cancer), *Brystcancer* (breast cancer [a combination of Danish and Latin/English]), *Mammakræft* (breastcancer [a danification of the latin Cancer Mammae, or a way to describe breastcancer]).

2. Words that are neutral and common to both laypersons and professionals:

   Words covering concepts that may be used in connection with cancer, but also words related to other topics belong to this category. These are often concepts where professionals and laypersons generally use the same concepts. The category includes common names for symptoms and parts of the human body. It also includes names of fundraising events.

   Examples: *Rygning* (smoking), *Arvelighed* (inheritability), *Medisinsk behandling* (medical treatment), *Transport* (transportation), *Strålebehandling* (ray therapy), *Kalk* (calcium), *Årsager* (causes), *Pårørende* (next of kin), *50 år* (50 years), *Stress* (stress), *Appelsinkage* (orange cake), *Armbånd* (bracelet), *Bryst* (breast), *Diætist* (dietitian), *Klinisk forsøg* (clinicial trials), *København* (Copenhagen), *Pinkcup* [name of a combined women's golf tournament and fund raising campaign, in Denmark], *Mindfullness* (mindfulness).

3. Words that belong to a professional's vocabulary:

   This category includes Latin or English words concerning cancer or medicine in general. Danish words concerning cancer or medicine in general that are not well known by the general Danish public are also included.

   Examples: *Herceptin* [a medicine], *Breast cancer*, *cancer mammae* [Latin for breast cancer], *Antihormoner* (antihormones), *Statiner* [a medicine], *Triple cancer* [short for Triple-negative breast cancer], *Bløddelssarkom* (soft tissue sarcoma).

4. Words that are hard to put in the previous three categories:

This category is used for names, tags related to membership in the Danish cancer society and the use of forms at Cancer.dk. Many of them have a meaning not related to medicine at all.

Examples: *afmelding nyhedsbrev* (newsletter unsubscription), *adresseændring* (change of address), *alt afmeldes* (unsubscribe everything), *anja olsen* [a woman's name], *arrivederci franz* [the name of a music album], *sommerlotteri* (summer lottery).

5. Words that were not categorized at all:

   Tags that do not contain real words, empty tags and tags with no meaning are put in this category. Tags that looks like usernames are also in this category.

When categorizing, it was sometimes hard to decide which category to choose. With names in category 4 and usernames in category 5, it was necessary to decide wether a halfway-finished name was a username or a name. I ended up putting e-mail addresses in category 4, because they usually identify a person in the same way as a name does. And usernames were kept in category 5 because there were so many tags where it was hard to say whether it was a username or something meaningless. However, in the facet categorization, which Bing and I did after the lay/professional categorization, we did try to separate between usernames and meaningless words. Thus, we covered this distincion and made it possible to discover if the taggers turned out to have a separate behaviour concerning this.

It was also sometimes hard to decide wether some tags chould be in category 2 or 4. One example is the fundraising projects. I see them as neutral: they are related to cancer but are generally not a part of the cancer medicine field. On the other hand, I have put words concerning membership in the Danish Cancer Society in category 4. This is because I see these words as more peripheral to the field of cancer. But the distinctions leave some terms to fall between two stools, like *trækningsliste* [list of winning numbers in a lottery] or *engangsbeløb* [lump sum]. Usually, we put these tags in category 2. In general, we restricted category four and five to the kinds of tags mentioned in the description and used category two when in doubt.

	Number of tags	Share	Number of tags before feature change	Share of tags before feature change	Tags on day of feature change	Share of tags on day of feature change	Number of tags after feature change	Share of tags after feature change
1 Lay	3322	13,15	3253	13,20	8	21,62	61	10,59
2 Neutral	11777	46,64	11601	47,08	21	56,76	155	26,91
3 Professional	1988	7,87	1955	7,93	0	0,00	33	5,73
4 Hard to categoreize	2075	8,22	1968	7,99	1	2,70	106	18,40
5 Empty tags, meaningless tags, etc.	6091	24,12	5863	23,79	7	18,92	221	38,37
	25253	**100**	**24640**	**100**	**37**	**100**	**576**	**100**

Table 7 Number of tags in lay/professional categories in total and before feature change, on day of feature change and after feature change

Table 7 shows the number of tags in each of the five categories, and the percentage share of each category. The first colomns with numbers give total numbers, then numbers are given for whether the tag was applied before the feature change, on the day of feature change, or after the feature change. Numbers for the day of the feature change are small, but the status of the feature for these tags is unknown. Thus, they are separated from the other tags.

All the tags, whether they belonged to a lay or a professional vocabulary, constituted the total amount of tags available on Cancer.dk. Thus, this was what users could see. But with internal and external tags separated, one can see who applied which tags. Table 8 show how internal and external users applied words from a lay and professional vocabulary as tags. The table include both numbers and shares. 37 tags from the day of the feature change and 1 tag from an internal tagger after the feature change are excluded. The most visible differences in tagging pattern is between internal and external taggers. But there are also differences in tagging behaviour for external taggers before and after the feature change.

	Tags before feature change				Tags after feature change		
	Internal tags	Share of internal tags	External tags	Share of external tags	External tags	Share of external tags	Sum
1 Lay	347	15,05	2906	13,01	61	10,61	3314
2 Neutral	1737	75,33	9864	44,17	154	26,78	11755
3 Professional	146	6,33	1809	8,10	33	5,74	1988
Sum, categories 1-3	2230	96,71	14579	65,28	248	43,13	17357
4 Hard to categoreize	48	2,08	1920	8,60	106	18,43	2074
5 Empty tags, meaningles tags, etc.	28	1,21	5835	26,13	221	38,43	6084
Sum	2306	100,00	22334	100,00	575	100,00	25215

Table 8 Tags for internal and external users with lay and professional vocabulary, before and after feature change. 1 tag from an internal user after the feature change and 37 tags from the day of the feature change are excluded.

Figure 20 shows the numbers of lay, neutral and professional tags without categories 4 and 5, empty and meaningless tags and tags that are hard to categorize, from Table 8. For external taggers, the most visible change is a higher share of lay tags and lower share of neutral tags after the feature change. For internal tags, the share of neutral tags is higer than external tags. And internal tags applied before the feature change have relatively more lay tags than professional tags, compared to external tags applied both before and after the feature change.

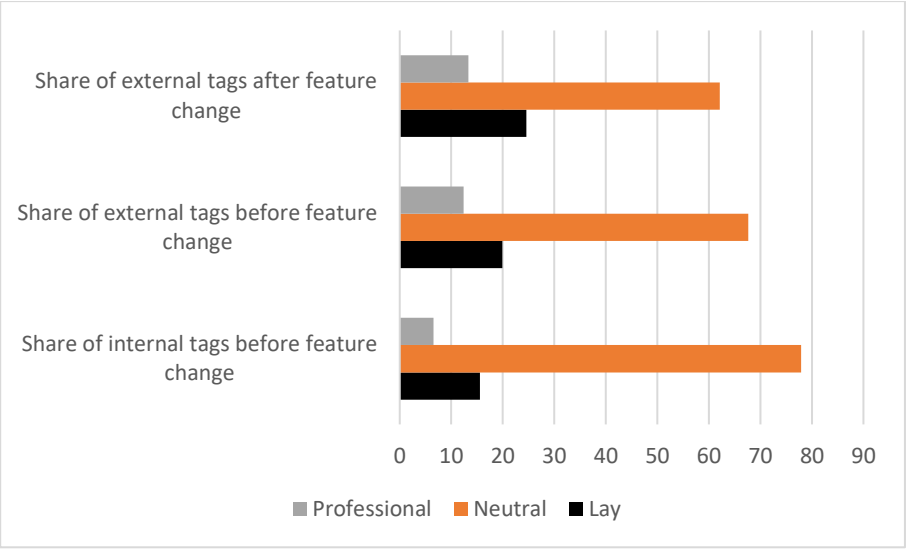


Figure 20 Share of lay and professjonal tags from internal and external users before and after feature change. 38 tags are excluded.

Both groups applied a high share of neutral terms. Both groups also applied more lay terms than professional ones. I assume that most of the external taggers were laypersons when it came to medicine in general and particularly cancer medicine. But the external taggers applied the highest share of the professional terms Before the feature change, 6.5% of tags from internal taggers were professional, and 12.4% from external taggers, a share almost twice as high.

After the feature change, external taggers applied fewer neutral tags and more lay tags.

7.2.11 FACETS

When working with the topical content of the tags, 26 blog tag categories made for Blogomkraeft.dk was as a starting point (see chapter 5.2). This would give an opportunity to compare the two sets of tags. When categorizing Cancer.dk tags according to these categories, new categories was added when needed, like *Statistics*, *Time, season* and *Other persons* [other than patients and professionals]. Some of the new categories were not topical, like *Statements, questions* and *Form, genre, document*. Among the new facets are also gropus for empty and meaningless tags, these tags are impossible to put in other facets even if the taggers who applied them had something meaningful in mind. See chapter 0 for details on new facets.

Some of the Blogomkraeft.dk categories needed evaluation and clarification before they culd be used on Cancer.dk tags, since the tags and the context differ sligthly. Examples are the categories *Doing something else*, *Health in general*, and *Life in general*. Table 9 shows the facets, with example tags and explanations.

Facet	**Example**	**Comment**
Cancer types and cancer in general	*Brystkræft* (breast cancer), *hudkræft* (skin cancer)	Types of cancer, stadium, short names for cancer, ondartede svulster.
Thoughts, feelings, and psychological symptoms	*sorg* (grief)	Includes all kinds of feelings and paychology connected to cancer and life. Activities to deal with these thoughts and feelings are also included, but no medical treatment
Treatment	*Kemo* (chemo[therapy]), *operation* (surgery), *behandlingsgaranti* (guaratnee of treatment), *pakkeforløb* (pathway [for treatment]), *redskaber* (tools [for a good life during treatment]), s*orggruppe* (grief support group),	Includes experimantal treatment, but not alternative treatment. Most tags in this facet are treatment of cancer or cancer related symptoms. But other kinds of treatment is also included.
Communication	*Afmeld* (unsubscribe), *har ikke bedt om* (did not ask for)	Communication or attemts to communicate, and tags about communication
Organizations and organized activities	*Testament* (will), *åbningstider* (opening hours), *sommerlotteri* (summer lottery)	Tags concerning funraising, meetings, economy, also organized activities and economy for patients, and administration.
Doing something else	*Golf* (golf), *ferie* (vacation), *Kunstkonkurrence* (art competition), *Løb* (race), *makeup* (makeup), *music* (music), *nyd livet* (enjoy life), *sport i solen* (sports in the sun), *Zumba* (Zumba)	Activities that can make me forget illnes, or give other impulses.
Symptoms and side effects	*Dissemineret* (multippel sklerosis)	Includes illnesses that are not cancer, unless they can cause cancer, then they are facet 20
Examination	*Scanning* (scanning)	Examination to discover cancer, what type of cancer, examination after cancer treatment is finished.

Facet	Example	Comment
		Includes examination for other illnesses as well.
Family, friends and others	*Pårørende* (next of kin)	Any person who are friends or family or in other ways close to a cancer patient. If the role is unclear, use facet 11.
Body and biology	*Tarm* (intestine), *bryst* (breast), *hud* (skin), *celler* (cells), *menstruation* (menstruation), *blod* (blood), *hår* (hair)	Body parts, organs, cells, also processes in the body that are not a part of an illness.
Patients and professionals	*Lege* (physician), *patient* (patient), *sorggruppe leder* (grief support group leader), *barn* (children), *kvinne* (woman)	Cancer patients and professionals who work to cure, comfort or in other ways help cancer patients.
Drugs to treat cancer	*Bcg* (bcg)	Cancer madicine, and other drugs too.
Food	*Mad* (food), *broccoli* (broccoli), *c-vitamin* (vitamin C)	Includes food, their recipes, ingredients, and notritous substances
Work	*Arbejde* (work)	Mainly patients' work. Healt professionals' work is usually treatment or examination etc.
Places, including places of treatment	*Udlandet* (abroad), *beliggenhed* (location), *venterom* (waiting room), *Operation udland* (surgery abroad), *København* (Copenhagen)	All places mentioned as tags. This includes generic names and buildings when they are seen as a location.
Cancer outcome	*Død* (death), *Rask* (well), *Overlevelse* (survival), *helbredelse* (recovery), *dø* (die), *dø i eget hjem* (die at home), *prognose* (prognosis), *uhelbredelig lungekræft* (incurable lung cancer), *terminal pleje* (terminal care)	Cancer outcome concerning life or death. Includes activities directly connected to cancer outcome, like funeral.
Preventing cancer	*HPV vaksine* (HPV vaccine), *rygestop* (quit smoking)	Drugs, vaccines, organizational activities, information campaigns, advices on food, anything that is said to prevent or may prevent cancer.

Facet	**Example**	**Comment**
Health in general	*Energi* (energy), *graviditet* (pregnancy), *graviditet og hpv* (pregnancy and HPV), *helbred* (health), *stress* (stress), *sund* (healthy), *sundhed og overvægt* (health and overweight)	Terms that describe health and health issues in general, independent of wether you have an illness or not.
Causes of cancer	*Radioaktivitet* (radioactivity), *Røyking* (smoking)	Anything that causes or may cause cancer, including suspicion or speculations that something may cause cancer.
Politics	*valg* (election)	Political activities, election
Life with cancer	*Egenomsorg* (self-care), *Erfaringer* (experiences), *Et liv med kræft* (a life with cancer), *hverdag* (everyday life), *Håret kommer igjen* (The hair comes back), *Kræftpatientens verden* (the world from a cancer patient's perspective, *leve med stomi* (live with stomy), *rehabilitering* (rehabilitation), *Samleje* (intercourse), *seksualitet* (sexuality), *Gravid og HPV* (pregnant and HPV) *sol efter strålebehandling* (sun after radiation treatment), *ventetid* (waiting time)	Tags that relate to life as it is with cancer. Daily normal things, but also how to adapt to symptoms or concequeses of cancer or its treatment.
Life in general	*60 års fødselsdag* (60 years' bithday), *arv* (legacy), *berøring* (touch), *middagslur* (afternoon nap), *orlov* (leave), *rejse* (travel), *skole* (school), *bryllup* (wedding)	Tags about general life, undisturbed by cancer.
Alternative treatment	*Auramassage* (aura massage), *alternativ behandling* (alternative treatment)	Treatment outside academic medicine. Experimental treatment is not in this facet.
Voluntary work	*frivillig arbeide* (voluntary work)	

Facet	Example	Comment
Other tags	*Ekstra* (extra), *fremragende* (outstandeing)	
Numbers, codes, words with no meaning		
Statements, questions	*Hej* (hello), *Hva med 1986* (What about 1986?), *Hjertelig til lykke Kirsten* (many congratulations, Kirsten)	Statements about anything, questions and greetings.
Name, e-mail address		Names, e-mail addresses and usernames that have similarities with names or e-mail addresses.
Empty tags		Nothing was typed at all before hitting the Submit-button
Time, season	*Vår* (spring), *årshjul* (annual cycle)	Time and time-related tags. Some dates may also be "lost" in facet 27 because they are numbers.
Things, goods	*Perler* (beads)	All kinds of things. Most examples are things for sale in the cancer.dk webshop, but also other things.
URL		Addresses to websites, or parts of such addresses.
From, genre, document	*Bilder* (pictures), *årsrapport* (annual report)	Both what the article is, and documents that are mentioned
Other persons	*Arkitekt* (architect)	Persons who are not patients, not friends and family, not health professionals.
Statistics	*Nordcan* [name of a souce of cancer statistics from the Nordic contries] *Kræfttal* (statistics concerning cancer, *Infektionstal* (number of infections)	Tags that relate to statistics how many times a phenomenon occurs.

Table 9 Facets, with examples and explanations

In many cases, the same word can have different meanings. This can also be the case with tag words. One example is the tag *aggressive* (aggressive) which was applied to an article titled *Inflammatorisk brystkræft* (Inflammatory breast cancer)
[Url:

/Hjaelp+viden/kraeftformer/kraeftsygdomme/inflammatorisk+brystkraeft/inflammatorisk+brystkraeft.htm]. Isolated, the tag could fit to at least these categories on Cancer.dk:

- Thoughts, feelings and psychological symptoms – Aggression as an emotional reaction to diagnosis or treatment.
- Cancer types and cancer in general – Aggression as a characteristic of the cancer itself.
- Communication – Aggression as a characteristic of communication.

In this example, the article used *aggressive* to describe inflammatory breast cancer and thus the tag was categorized as *Cancer types and cancer in general*. The word, in another grammatical form, *aggression* (aggression), was applied to an article titled *Råd til pårørende* (Advice for those next of kin). Here it is more likely that it is connected to feelings. Thus, the tag was categorized as *Thoughts, feelings, and psychological symptoms*.

When subject indexing, tearms are formulated to distinguish between different meanings of homonyms and polysemic words. Organizing tags into facets is not the same, but I have distinguished the different meanings in a similar way. And it is worth noting that it is needed because tags and tagging features normally do not include such distinctions.

The external taggers applied most tags in the following categories:

1. Cancer types and cancer in general
2. Places, including places of treatment
3. Organizations and organized activities
4. Food
5. Body and biology
6. Symptoms and side effects

The internal taggers applied most tags to in the following categories:

1. Cancer types and cancer in general
2. Drugs to treat cancer
3. Treatment
4. Causes of cancer
5. Organizations and organized activities
6. Symptoms and side effects
7. Thoughts, feelings, and psychological symptoms

Table 10 shows facets where internal users dominate with relatively more tags compared to external users.

Nr	Facet	Internal tags	Share of tags in facet (%)
24	Alternative treatment	84	60,43
8	Examination	108	44,81
19	Health in general	31	40,79
22	Life with cancer	87	35,51
20	Causes of cancer	174	33,53
2	Thoughts, feelingd, and psychological symptoms	133	29,62
9	Family, friends and others	53	29,12
36	Statistics	23	27,71
3	Treatment	178	25,72
14	Work	26	24,53
18	Preventing cancer	75	22,06
11	Patients and professionals	81	20,10

Table 10 Facets where internal users dominate with relatively more tags compared to external users

The facets listed in Table 10 have in common that they include tags about what The Danish Cancer Society really wants to inform about. 60% of tags about alternative treatment are from internal taggers. This could correspond to an interest of the Danish Cancer Society to inform about alternative treatment. They fund and conduct cancer-related research and have an interest to inform patients about what does and what does not work. The Danish Cancer Society also probably have contact with patients who have tried different kinds of alternative treatment with diverse results. They understand how desperate cancer patients can be and their willingness to spend money, time and energy on alternative treatment, simply because they have a hope that it might save one's life. The articles about this topic have a rather neutral and friendly tone, but also includes warnings and facts about what kinds of treatment does not work, and what works under what conditions. The relatively high number of tags on alternative treatment from internal taggers can be result of their experience with this as an important topic for many of the users.

The other facets in Table 10 can also mainly be explained with a wish to inform users.

Blogomkraeft.dk tags	Internal Cancer.dk tags	External Cancer.dk tags
Cancer types and cancer in general	Cancer types and cancer in general	Cancer types and cancer in general
Thoughts, feelings and psychological symptoms	Food	Places, including places of treatment
Treatment	Treatment	Organizations and organized activities
Communication	Causes of cancer	Food
Organizations and organized activities	Body and biology	Body and biology
Doing something else	Organizations and organized activities	Symptoms and side effects
Symptoms and side effects	Symptoms and side effects	Other tags
Examination	Thoughts, feelings, and psychological symptoms	Treatment
Family, friends and others	Examination	Cancer outcome
Body and biology	Life with cancer	Communication
Patients and professionals	Alternative treatment	Causes of cancer
Drugs to treat cancer	Patients and professionals	Patients and professionals
Food	Other tags	Thoughts, feelings, and psychological symptoms
Work	Preventing cancer	Preventing cancer
Places, including places of treatment	Places, including places of treatment	Drugs to treat cancer
Cancer outcome	Family, friends and others	Life with cancer
Preventing cancer	Cancer outcome	Examination
Health in general	Health in general	Family, friends and others
Causes of cancer	Drugs to treat cancer	Doing something else
Politics	Work	Voluntary work
Life with cancer	Doing something else	Work
Life in general	Life in general	Life in general
Alternative treatment	Voluntary work	Alternative treatment
Voluntary work	Politics	Health in general
	Communication	Politics

Table 11 Facets sorted on number of tags applied by each user group: Blogomkraeft.dk taggers, Cancer.dk internal taggers and Cancer.dk external taggers. Facets with the most percentage of tags in each user group have an individual colour, to make it easier to compare the columns visually

7.2.11.1 Facets before and after feature change

Figure 21 show the share of tags categorized by their topical content. Tags from internal taggers after the feature change, and tags from the day of feature change are left out, leaving three categories: Tags from internal taggers before the feature change, and tags from external taggers before and after the feature change.

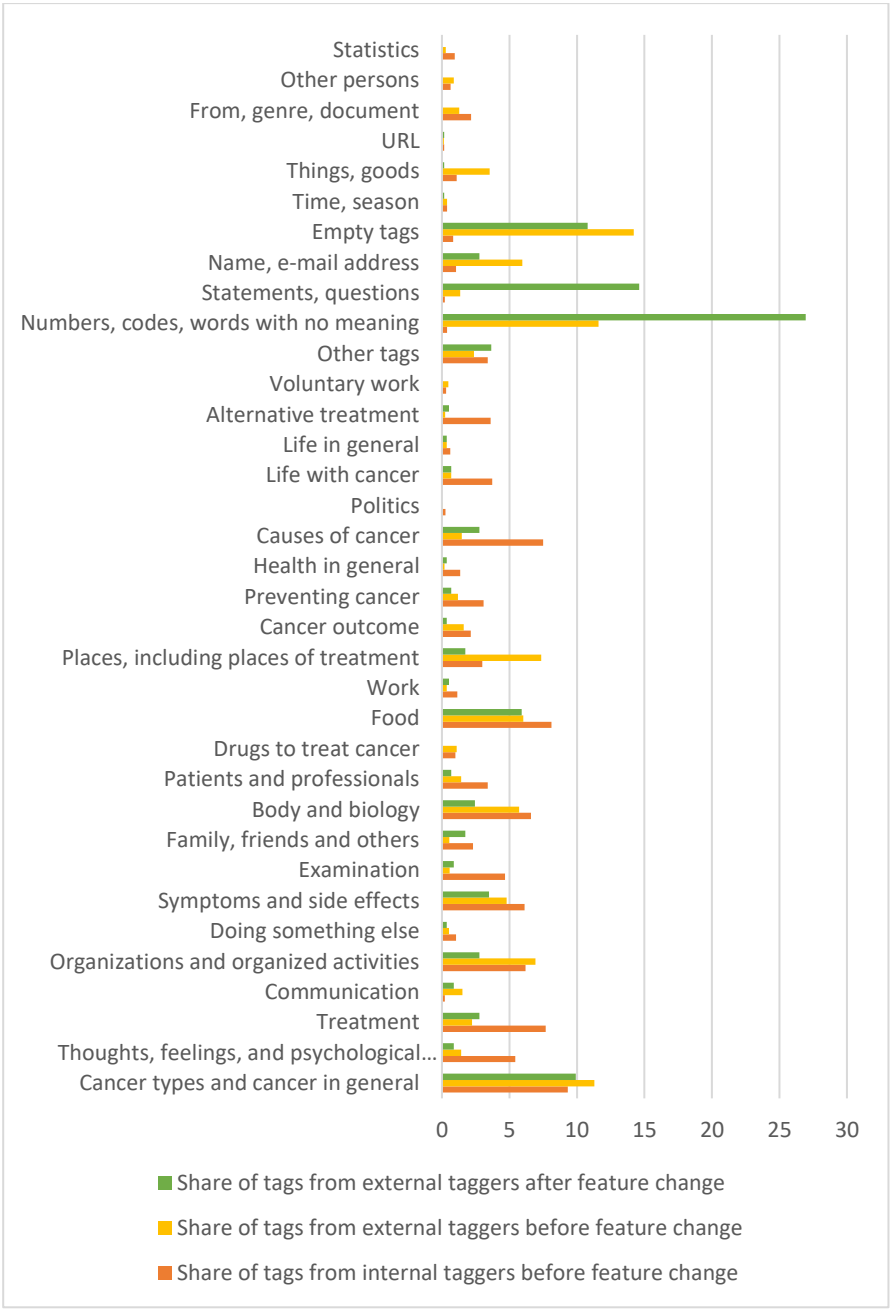


Figure 21 Tags in facets: Share of applied tags from external and internal taggers before feature change, and from external taggers after feature change. Tags form the day of the feature change and from internal taggers after the feature change are not included

7.2.11.2 New facets on Cancer.dk

Facet	External tags	Share of external tags	Internal tags	Share of internal tags
Statistics	61	0,27	22	0,95
Time, season	85	0,37	9	0,39
Statements, questions	387	1,69	5	0,22
Things, goods	795	3,46	25	1,08
Other persons	194	0,85	15	0,65
From, genre, document	288	1,26	50	2,17
Name, e-mail address	1344	5,86	24	1,04
Empty tags	3240	14,12	19	0,82
URL	30	0,13	4	0,17
Numbers, codes, words with no meaning	2748	11,98	9	0,33
Sum	9172	39,99	182	76,8

Table 12 Cancer.dk tag facets that are new compared to Blogomkraeft.dk tags

Table 12 shows new facets on Cancer.dk, compared to Blogomkraeft.dk. Examples and comments on these facets can be found in Table 9. The five first of the facets in Table 12 contain tags that describe the content of articles. The five last facets include tags that most often represent mistaken use of the tagging field. They almost never represent the content of the article, at least not in a way that is visible to other users.

The percentages in Table 12 refer to the total numbers of external and internal tags. Thus, most of the categories have a low share of tags. The exceptions are facets with tags from external users: the empty tags and tags with *numbers, codes and meaningless words*. Also, *names and e-mail addresses* from external users has a relatively high number of tags.

7.2.11.3 Statements, questions and tags with unappropriate content

When the tagging feature was planned, the possibility for unappropriate tags was one reason why the editors wanted a tool for editing and deleting tags. When a feature like this is set up and available to anyone, it is not easy to predict whether somebody will misuse it in any way. All unappropriate tags were categorized into the *Statements, questions*-facet. 113 of the tags are clearly inappropriate. See Table 12, and Table 9 for general examples of tags in this facet. It contains 1.38% of the tags and only a few of them are unappropriate, 0.24% of the tags. Most of them have a sexual content, and many include a name, like this example: *A er homo* (A is gay). Another example, *sundhedsfascisme* (health facism), applied to an article about changes in the Danish smoking law, may be unwanted because of the plump formulation.

All tags that can be considered unappropriate, were applied by external taggers. Only five of 383 (1.31%) tags in this facet, Statements and questions, were applied by internal taggers.

Some of the tags with statements and questions have a content that is fear and ok. There are questons that other users also may want an answer to, like: *Hvornår er man terminal?* (When is one terminal?), applied to a short article on terminal care (www.cancer.dk/Hjaelp+viden/ordbog/T/Terminalpleje.htm). 141 of the external tags are such questions. Other examples are *hvad siger statistikken* (what can statistics tell[?]) and *hvilke andre skader har uv-stråling* (what other damages can ultraviolet radiation give[?]).

7.2.11.4 Tag facets distributed on lay or professional categories

The distribution of lay and professional tags on facets can be found in Table 13 and Table 14. This distribution can mostly be explained with how some facets are naturally connected to a lay vocabulary, while some naturally include more professional words. Lay and professional vocabulary are thus not evenly distributet among the facets. Some facets are closely related to health and medicine, some are not. This can explain why the *Drugs to treat cancer*-facet have the highest number of tags with professional words, 74%. The drugs are a part of the professional sphere and normaly do not have a lay synonym. Only 2% of the tags abuot drugs have lay words. This may lead to an explanation to why drug tags are relatively more often applied by external users, compared to internal users, 90% of the tags about drugs are applied by external users: Patients need to nail these words, to learn them and to find information about them. They want to make them searchable.

Facet	1 lay	%	2 neutral	%	3 prof.	%
Cancer types..	1898	67,71	54	1,93	841	30,00
Thoughts, feelings..	12	2,67	429	95,55	5	1,11
Treatment	200	28,90	397	57,37	90	13,01
Communication	2	0,56	75	21,19	0	0,00
Organizations..	275	16,10	1124	65,81	93	5,44
Doing something..	1	0,70	131	92,25	0	0,00
Symptoms..	255	20,75	759	61,76	211	17,17
Examination	63	26,14	110	45,64	67	27,80
Family, friends..	6	3,30	167	91,76	1	0,55
Body and..	163	11,27	1094	75,66	184	12,72
Patients and ..	59	14,68	330	82,09	10	2,49
Drugs..	7	2,62	60	22,47	198	74,16
Food	12	0,77	1457	93,04	29	1,85
Work	2	1,89	103	97,17	0	0,00
Places..	14	0,81	1651	95,88	5	0,29
Cancer outcome	94	22,93	303	73,90	12	2,93
Preventing cancer	39	11,47	293	86,18	3	0,88
Health in general	1	1,32	59	77,63	11	14,47
Causes of cancer	64	12,33	437	84,20	13	2,50
Politics	0		17		0	
Life with cancer	42	17,14	202	82,45	1	0,41
Life in general	6	6,45	86	92,47	0	0,00
Alternative..	3	2,16	118	84,89	14	10,07
Voluntary..	0	0,00	108	94,74	0	0,00
Other tags	5	0,79	511	80,73	50	7,90
Numbers, codes..	4	0,14	18	0,65	66	2,39
Statements..	63	16,15	147	37,69	23	5,90
Name, e-mail..	0	0,00	106	7,73	37	2,70
Empty tags	0	0,00	0	0,00	0	0,00
Time, season	1	1,08	87	93,55	0	0,00
Things, goods	3	0,37	778	94,88	4	0,49
URL	0	0,00	9	26,47	1	2,94
From, genre..	20	5,92	295	87,28	2	0,59
Other persons	2	0,96	200	95,69	3	1,44
Statistics	6	7,23	61	73,49	14	16,87
Average	**92,28**	**8,98**	**327,11**	**66,89**	**55,22**	**7,62**
Sum						

Table 13 Facets combined with categories for lay/professional vocabulary, categories 1-3

Facet	4 hard	%	5 no cat.	%	sum
Cancer types..	10	0,36	0	0,00	2803
Thoughts, feelings..	3	0,67	0	0,00	449
Treatment	5	0,72	0	0,00	692
Communication	277	78,25	0	0,00	354
Organizations..	216	12,65	0	0,00	1708
Doing something..	10	7,04	0	0,00	142
Symptoms..	4	0,33	0	0,00	1229
Examination	1	0,41	0	0,00	241
Family, friends..	8	4,40	0	0,00	182
Body and..	5	0,35	0	0,00	1446
Patients and ..	3	0,75	0	0,00	402
Drugs..	2	0,75	0	0,00	267
Food	68	4,34	0	0,00	1566
Work	1	0,94	0	0,00	106
Places..	52	3,02	0	0,00	1722
Cancer outcome	1	0,24	0	0,00	410
Preventing cancer	5	1,47	0	0,00	340
Health in general	5	6,58	0	0,00	76
Causes of cancer	5	0,96	0	0,00	519
Politics	0		0		17
Life with cancer	0	0,00	0	0,00	245
Life in general	1	1,08	0	0,00	93
Alternative..	4	2,88	0	0,00	139
Voluntary..	6	5,26	0	0,00	114
Other tags	66	10,43	1	0,16	633
Numbers, codes..	37	1,34	2634	95,47	2759
Statements..	155	39,74	2	0,51	390
Name, e-mail..	1056	76,97	173	12,61	1372
Empty tags	0	0,00	3259	100,00	3259
Time, season	5	5,38	0	0,00	93
Things, goods	35	4,27	0	0,00	820
URL	2	5,88	22	64,71	34
From, genre..	21	6,21	0	0,00	338
Other persons	4	1,91	0	0,00	209
Statistics	2	2,41	0	0,00	83
Average	**57,64**	**8,47**	**169,19**	**8,04**	**701,44**
Sum					**25252**

Table 14 Facets combined with categories for lay/professional vocabulary, categories 4-5 and total

The facet for *Cancer types and cancer in general* has a different distribution. Here, 67% of the tags are with lay words and 30% with professional ones. The low number of neutral words here can be explained with the fact that this is a facet in the core of cancer medicine. The high number of lay words is because there are lay alternatives for most of the cancer diagnosis terms. The different types of cancer are named with body part. Usually this body part has a well known Danish name. Cancer.dk use these Danish names, and taggers use them. Thus, there is not the same need to nail the professional terms.

Nine facets are totally dominated by neutral vocabulary. They are (see also Table 13 and Table 14):

- Thoughts, feelings and psychological sumptoms
- Doing something else
- Family, friends and others (close to patients or their situations)
- Food
- Work (most work tags concern the jobs of cancer patients)
- Places, including places of treatment
- Life in general
- Voluntary work
- Time, season
- Other persons (other than health professionals)

All these facets belong to an everyday sphere where it is not natural to use a separate professional vocabulary. For some facets, a professional vocabulary hardly exixts. For a facet like *Food*, there is a professional vocabulary with loanwords from e.g. French. But it does not seem natural to use them in this setting of health and medicine, for neither professional nor laypersons.

7.2.12 ABOUTNESS

The aboutness categorization concerns wether tags are suitable labels to represent the topical content of an article. See chapters 3.1.1 and 7.1.5 for definistions on aboutness. Both the editors that I interviewed, and to a cartain degree the interviewed taggers, saw it as important that there was a relationship between the aboutness of tags and articles they were applied to. To evaluate this relationship, all tags were categorzied according to whether they referred to the same topics as the articles they were applied to. The importance of this categorization was underlined by an early observation: A relatively large amount of the tags do not relate to the aboutness of the articles they were applied to.

It is sometimes hard and sometimes easy to decide whether a tag is about the same topic as the article or not. Tags that belong to facets not represented in the content of the article are often easier to categorize. If an article on skin cancer has the tag

solcreme (suntan lotion) applied to it, it is quite easy to decide that this tag do not represent the content of the article. On the other hand, the tag is relatet to the content, because the use of suntan lotion can prevent skin cancer.

Do the tag *morphea-typen* (the morphea type) concern the content of an article about the treatment of Squamous-cell carcinoma? *Morphea type* probably refers to the skin desiese Morphea, or localized scleroderma. Squamous-cell carcinoma is a type of skin cancer. On a more detailed medical level, the two may have something to do with each other. At least it is possible that one patient has both desieses. But the nature of the relation between tag and article content is only available to those who know both concepts. For the rest of us, it is hard to know. If you search for 'Morphea' and retrieve this article, is it relevant? To a layperson, I would say no. I see this question: "If you search for …" as a help to decide whether the tag and article are related or not.

Another example shows another relationship between tag and article. The tag *marginalzone* (marginal zone) was applied to an article about the Lymphatic system. The marginal zone is a specific part of the lymphatic system, so the tag is about the same topic as the article, but more specific. This means that if you are interested in the marginal zone (within medicin), this article gives you information on a general level.

A third example illustrate a third relation between tag and article. The tag *kolangiokarcinom* (cholangiocarcinoma) was applied to an article about the gall bladder and the bile duct. It is a synonym for cancer in bile duct, and thus a good term that adds relevant words to the article, without adding content. This is a type of tag that both taggers and editors agree to be good or ideal tags.

The following categories was used to characterice how tags related to the aboutness of the articles:

Too general:	The tag and the article refer to the same topic, but the tag is too general and thus not a good description of the article topic. Examples: *Kokebok* (cookery book) applied to an article with only one recipe. *Redskaber* (tools) applied to an article about teqcniques to help yourself to sleep well.
General topic:	The tag and the article refer to the same topic, the tag is more general than the article content, but still a fairly good description of the article topic. Examples: *Psykiske reaktioner* (psychological reactions) applied to an article on anxiety trigged by a cancer digagnosis.

	Tilbehør (trimmings) applied to a remoulade recipe.
Exact topic:	The tag and the article refer to exactly or nearly exactly the same topic. Examples: *Fysisk aktivitet* (physical activities) applied to an article which gives good reasons to be physically active. *Kræft i hjernen* (cancer in the brain) applied to an article on brain metastasis.
Exact aspect of topic:	The tag refers exactly or nearly exactly to a part of the main topic of the article. Examples: *Testikel* (testicle) applied to an article on how to examine yourself and find symptoms of cancer in the testicles, if any. *Uv-varsel* (uv forecast) applied to a page where you can subscribe to such a forecast.
Topic mentioned:	The tag refer to an aspect of the topic of the article, for instance a subtopic mentioned in a paragraph in the article. Examples: *Aalborg* [a Danish city] applied to a page with an overview on secondhand shops, one of them in Aalborg. *And* (duck) applied to a list of recipies with poultry, some of them with duck as an ingredient.
Too specific:	The tag refers to an aspect of the topic of the article, but is too specific. The article does not mention this subtopic. Examples: *Knoglekræft* (bone cancer) applied to an article about types of cancer in general. *Lotteri* (lottery) applied to a general article about how to support the Danish Cancer Society. Lottery is one of the ways, but not mentioned directly here.

Indirectly related to article:	Tags that relate or may relate to the article, but not the topical content. Examples: *God nyhed* (good news) applied to an article about decrease in the number of patients with cancer. *Udmeldelse* (sign out) applied to an article about membership in the Danish Prostate Cancer Organization.
No relation:	The tag does not refer to the content of the article at all. Examples: *Blodlegemer* (blood corpuscle) applied to a definition about chemotherapy. *Brystben* (breastbone) applied to a general article about preventing cancer.
No access:	The article was deleted when I attempted to categorize it, or access was denied for other unknown reasons.

The categories have similarities with relationships between index terms in a thesaurus (International Organization for Standardization, 2011). They also have similarities whith a requirement often seen in controlles vocabularies, that the subject heading should express the aboutness of the document directly, not indirectly (Hjortsæter, 2005). Only tags in the categories *Exact topic* and *Exct aspect of topic* would qualify as subject headings, as subject languages often recommend using the most specific term possible.

Empty and meaningless tags were left out from these categories. Instead, they were included as separate categories in figures and tables.

7.2.12.1 General numbers on aboutness

The number of tags that does not relate to the aboutness is disturbingly high, 48.42% of tags are not related to the topic of the articles to which they are applied. See Table 15, the table is further described in chapter 7.2.12.2.) The number of tags related to the aboutness of articles is lower, 19.96%. This number includes the categories General topic, Excact topic, Excact aspect of topic, Topic mentioned and Relate to article. If tags that are related but too specific are added, the number is 30.24%, which is still lower than the number of not related tags.

7.2.12.2 Aboutness before and after feature change

The feature change in September 2012 brought changes to the satisfaction of editors of Cancer.dk. They observed that the number of tags decreased dramatically. Also, the number of irrelevant tags decreased. But the percentage of tags referring to the topic of the articles they were applied to, was slightly higher before the feature change. In August 2012, 297 tags were related to the aboutness of the article. In October 2012, after the change, the number was 13. Thus, the satisfaction of the editors was connected to the amount of tags *not* related to the aboutness of articles, a decrease from 873 non-related tags in August 2012 to 10 in October. To the editors, this loss meant less work and less annoying tags.

Table 15 show how this looks when all tags are taken into account. It also shows the number of tags when the categories are grouped into summarized categories: Related, not related and empty/meaningless tags. The shade of grey in Table 15 show which aboutness categories belong to which summarized aboutness category. Each category is shown with both numbers and percentage shares. Figure 22 shows these summarized aboutness categories visually. The share of tags related to the aboutness of the articles is 19.99% before the feature change, and 18.58% after. In numbers, this is an unwanted decrease, but the share is almost the same and could be accidental due to the low number of tags applied after the feature change.

The difference between the shares of tags not related to the aboutness for the article is higher, down from 48.55% to 42.88%. For the share of empty and meaningless tags it is opposite, the share is up from 31.46% to 38.54%. Thus, these two categories have the larges changes before and after the feature change.

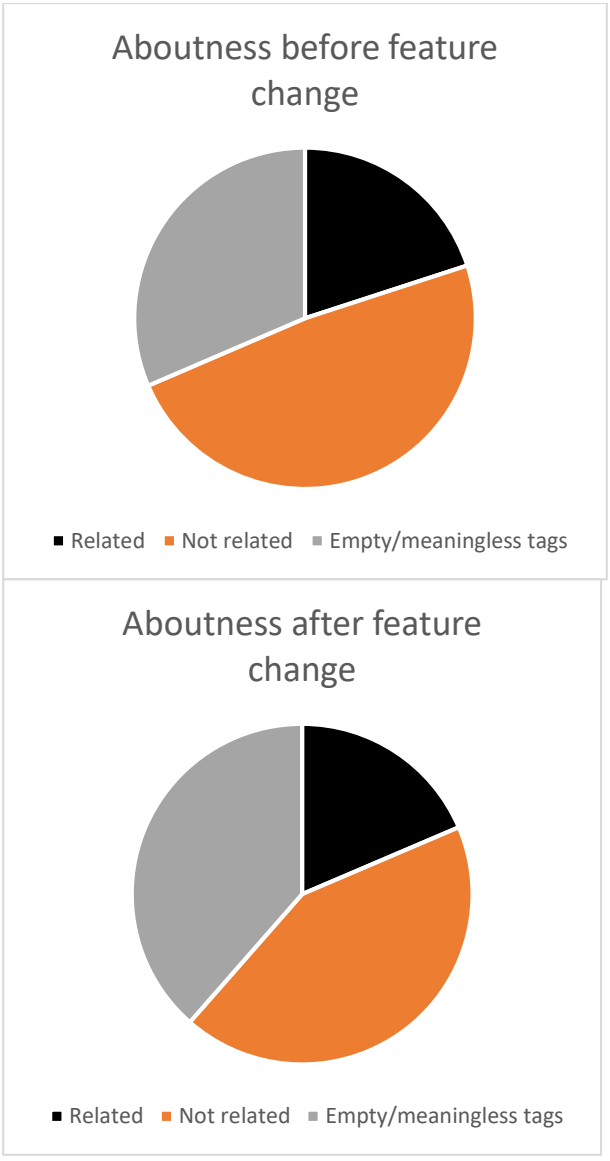


Figure 22 Share of tags in summarized aboutness categories, before feature change

Aboutness categories	Total number of tags		Tags before feature change		Tags before feature change	
	Number	%	Number	%	Number	%
General topic	310	1,23	308	1,25	2	0,35
Exact topic	855	3,39	838	3,40	17	2,95
Exact aspect of topic	622	2,46	605	2,45	17	2,95
Topic mentioned	3042	12,05	2983	12,09	60	10,42
Indirectly related	210	0,83	199	0,81	11	1,91
Too general	149	0,59	142	0,58	7	1,22
Too specific	2597	10,28	2554	10,35	43	7,47
No relation	9481	37,54	9284	37,62	197	34,20
Empty tags	3260	12,91	3198	12,96	62	10,76
Meaningless tags	2673	10,58	2515	10,19	158	27,43
Article was deleted	2053	8,13	2051	8,31	2	0,35
Sum	**25253**	**100**	**24677**	**100**	**576**	**100**
Summarized abountess categories						
Related	5040	19,96	4933	19,99	107	18,58
Not related	12227	48,42	11980	48,55	247	42,88
Empty/meaningless tags	7986	31,62	7764	31,46	222	38,54
Sum	**25253**	**100**	**24677**	**100**	**576**	**100**

Table 15 How tags relate to the abuotness of articles before and after feature change, and in total. The summarized aboutness categories are based on rows with similar shade of grey.

7.2.12.3 Aboutness in internal and external tags

Table 16 shows the number of tags in each aboutness category from internal and external taggers before and after feature change. Figure 23, Figure 24 and Figure 25 visualize the grouped aboutness categories and show the number of related and not related tags, and the empty or meaningless tags. There is no figure for internal tags after the feature change, because there is only one (see Table 16).

Note that external taggers applied more related tags after the feature change compared to before, up from 15.17% to 18.43%.

Aboutness categories	**Internal tags before feature change**		**Internal tags after feature change**		**External tags before feature change**		**External tags after feature change**	
	Number	%	Number	%	Number	%	Number	%
General topic	75	3.25	0	0	233	1,04	2	0,35
Exact topic	263	11.40	0	0	575	2,57	17	2,96
Exact aspect of	231	10.01	0	0	374	1,67	17	2,96
Topic mentioned	956	41.46	1	100	2024	9,06	59	10,26
Indirectly related	15	0.65	0	0	184	0,82	11	1,91
Too general	40	1.73	0	0	102	0,46	7	1,22
Too specific	90	3.90	0	0	2464	11,01	43	7,48
No relation	210	9.10	0	0	9074	40,56	197	34,26
Empty tags	19	0.82	0	0	3179	14,21	62	10,78
Meaningless tags	10	0.43	0	0	2505	11,20	158	27,48
Article was deleted	397	17.21	0	0	1654	7,39	2	0,35
Sum	**2306**	**100**	1	**100**	**22371**	**100**	**575**	**100**
Summarized abountess categories								
Related	1541	66,79	1	100	3393	15,17	106	18,43
Not related	340	14,73	0	0	11640	52,03	247	42,96
Empty/meaningless tags	426	18,46	0	0	7338	32,80	222	38,61
Sum	**2307**	**100**	**1**	**100**	**22371**	**100**	**575**	**100**

Table 16 Aboutness in internal and external tags, in numbers and in percentage, before and after feature change. The summarized aboutness categories are based on rows with similar shade of grey. Tags from the day of the fature change are excluded.

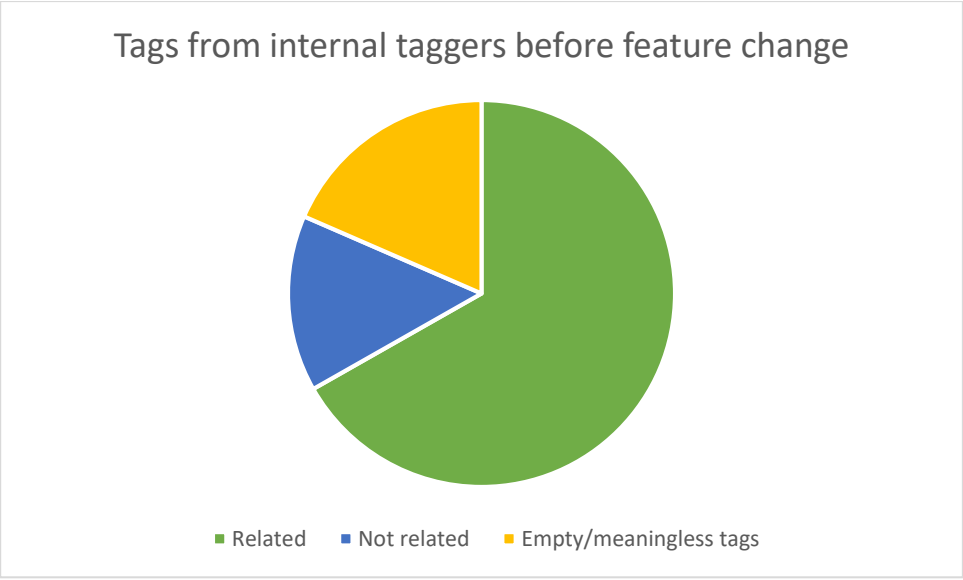


Figure 23 Share of tags in summarized aboutness categories, tags from internal taggers before feature change

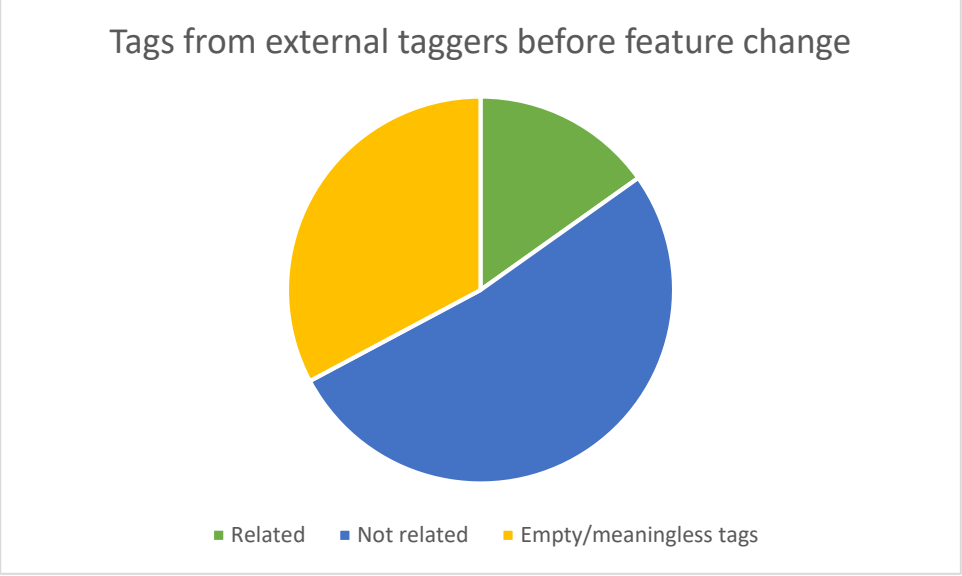


Figure 24 Share of tags in summarized aboutness categories, tags from external taggers before feature change

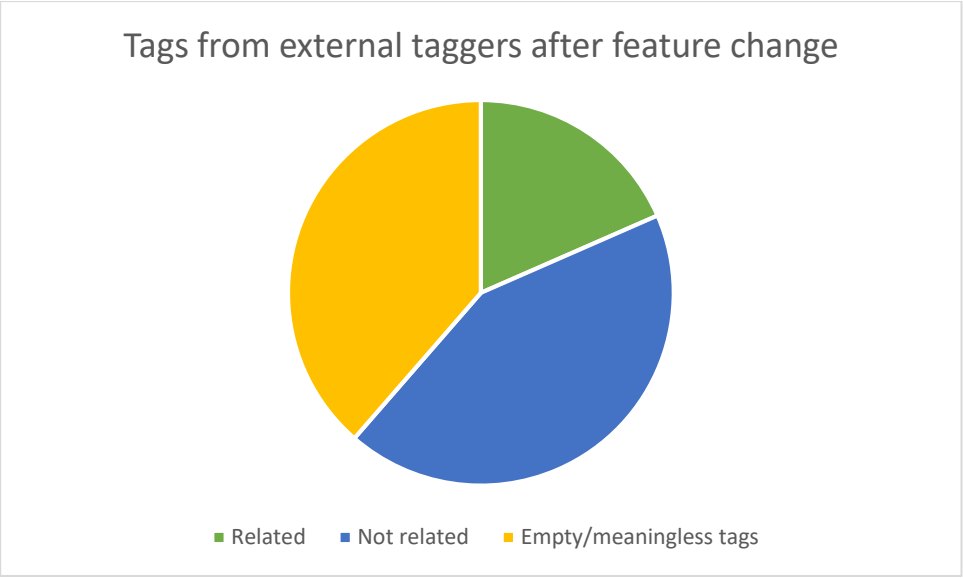


Figure 25 Share of tags in summarized aboutness categories, tags from external taggers after feature change

The aboutness categories can also be grouped to show specificity in tags. The categories *Too specific* and *Topic mentioned* are both used for tags that are within the theme of the article, but more specific than the topic of the whole article. If we group these two categories from Table 16 and summarize the share of tags, 45.36% of internal tags applied before the feature change were more specific than the topic of the whole article. For exernal taggers, the sums are 20.07% before the feature change and 17.74% after the feature change.

Likewise, tags that are about the same topic but more general than the topic of the article, can be grouped and the number of tags summed up. The aboutness categories *General topic* and *Too general* are both used for these tags. When grouping and summarizing them here, 4.98% of tags from internal taggers are more general than the article. For external taggers, 1.50% of tags before the feature change and 1.57% of tags after the feature change describe a topic that is more general than the topic of the article.

Note that the shared listed here, are based on the aboutness categories in Table 16, but differ from the summarized aboutness categories.

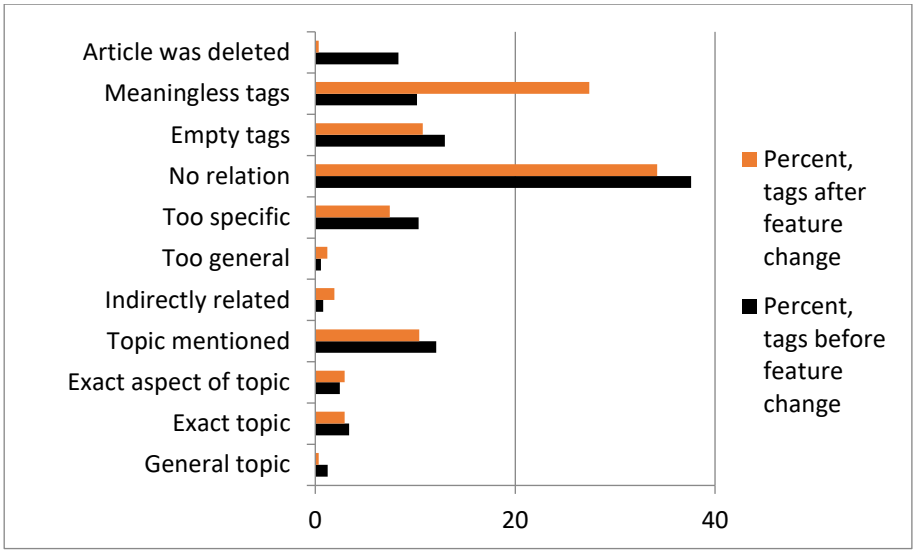


Figure 26 Aboutness categories before and after feature change, numbers for external tags, in percent

Figure 26 shows aboutness categorization in percentage for tags applied by external taggers. Numbers can be found in Table 16. They show other changes than the decrease in total number of tags before and after the change. The picture change when looking at each aboutness category, compared to the summarized categories. Related tags and tags that represent aspects of the article topic, increased slightly in percentage after the change. On the other hand, there is a decrease in the number of tags where the tag refer to the exact topic (Exact topic in Figure 26 and Table 16) of the article, the tag refer to something mentioned in the article (Topic mentioned in Figure 26 and Table 16), or the tag refer to the topic in general but still in a fairly desctiptive way (General topic in Figure 26 and Table 16). All together, the number of tags that refer to the topical content of the articles decreased slightly after the feature change.

The largest change is in the number of meaningless tags, which more than doubled from 10% to 27% of the tags. At the same time, the number of empty tags decreased, and the number of not related but otherwise meaningful tags ("No relation" in Table 16) also decreased.

Only one tag was applied by an internal user after the feature change, it is not included in Table 16. It is categorized as "Topic mentioned" and thus relate to the topic of the article. But only one tag does not show any tendency and it will not be analysed further.

The internal taggers applied better tags than external taggers, if good tags are those that refere to the aboutness of the articles and8 Table 16). Still, 14.73% of the internal

tags do not refer to the content of the article and are not useful for refinding this content through search or browsing. External taggers apply many tags that do not relate to the aboutness of the articles, 51.80% of their applied tags, in addition to the empty and meaningless ones. But the tags that do relate to the subject are as many as 3,499, thus a definite contribution to Cancer.dk. The only challenge is that there is no way to select and view these tags only.

When looking at the details, internal taggers applied more tags that related to the topical content of the articles, with the indirectly related tags as an exception. They applied more tags that were too general compared to the topical content of the article, but in the other two categories for tags not related to the topic of articles, they applied fewer tags than external taggers.

7.2.12.4 Aboutness in lay and professional tags

Summarized aboutness categories / Lay professional vocabulary		Related	Not related	Empty/ meaningless tags
Lay	**Number**	998	2137	187
	%	30.04	64.33	5.63
Neutral	**Number**	3143	7153	1481
	%	26.69	60.74	12.58
Professional	**Number**	506	1210	272
	%	25.45	60.87	13.68
Hard to categorize	**Number**	380	1500	195
	%	18.31	72.29	9.4
No categorization needed	**Number**	13	227	5851
	%	0.21	3.73	96.06

Table 17 Aboutness in lay and professional vocabulary

Table 17 shows how the lay and professional tags relate to the aboutness of the articles. It shows that lay tags tend to relate to the aboutness of the articles more often than neutral and professional tags. And it shows an opposite pattern for tags not related to the aboutness of the articles. One reason for this, can be that users have a better

understanding of the lay words and thus use them better. The difference however, between the percentage of tags with lay, neutral and professional words, is small, approximately 5 percent points.

	Before feature change				After feature change		Sum
	Internal tags	%	External tags	%	External tags	%	
Lay	347	15.05	2906	13.01	61	10.61	3314
Neutral	1737	75.33	9864	44.17	154	26.78	11755
Professional	146	6.33	1809	8.10	33	5.74	1988
Hard to categoreize	48	2.08	1920	8.60	106	18.43	2074
Empty tags, meaningless tags, etc.	28	1.21	5835	26.13	221	38.43	6084

Table 18 Lay and professional terms before and after feature change. 37 tags applied on the day of the feature change and 1 internal tag applied after the feature change are excluded.

Table 18 shows lay and professional categories for tags applied before and after the feature change. External tags are separated from internal tags, but the one internal tag applied after the feature change is excluded. Tags applied on the day of the feature change are also excluded. The table mainly show an increase in the number of tags hard ot categocize and the empty and meaningless tags. The number of neutral tags applied by external taggers also changed, down from 44.17% to 26.78%. For the remaining categories, changes are smaller, and the external taggers applied after the feature change are relatively small.

7.2.12.5 Aboutness in tag facets

The different tag facets have different numbers of tags that relate to the aboutness of the articles. See Table 19. The colomns of the table shows numbers of tags and shares of tags categorized by summarized aboutness categories. The communication facet stands out, with 65.35% tags that relate to the aboutness of the articles. An interesting thing about these tags is that many of them are requests for communication with the Danish Cancer Society, or attemts to communicate via tags directly. A typical example is contact pages tagged with *adresseændring* (address modification). These tags are not regarded as related to the aboutness of the page. But there is no place on

Cancer.dk you can give a new address, so the contact pages are the closest place if you want to inform the Danish Cancer Society about your new address or to end your membership. Another typical example is a landing page to unsubscribe to newsletters from Cancer.dk. These pages have many inscances of the tag *afmeld* (unsubscribe), which are highly relevant to the content. On the other hand, is this tag really useful to anyone?

Facet	**Tags related to aboutness of article**	**%**	**Tags not related to aboutness of article**	**%**	**Empty and meaningless tags, tags applied to deleted pages**	**%**
Cancer types and cancer in general	669	23,87	1963	70,03	171	6,1
Thoughts, feelings, and psychological symptoms	151	33,63	267	59,47	31	6,9
Treatment	297	42,86	344	49,64	52	7,5
Communication	232	65,35	117	32,96	6	1,69
Organizations and organized activities	555	32,49	854	50	299	17,51
Doing something else	32	22,54	67	47,18	43	30,28
Symptoms and side effects	356	28,92	760	61,74	115	9,34
Examination	119	49,38	84	34,85	38	15,77
Family, friends and others	94	51,65	73	40,11	15	8,24
Body and biology	397	27,46	957	66,18	92	6,36
Patients and professionals	87	21,64	276	68,66	39	9,7
Drugs to treat cancer	55	20,6	151	56,55	61	22,85
Food	602	38,44	889	56,77	75	4,79
Work	28	26,42	73	68,87	5	4,72
Places, including places of treatment	107	6,2	1243	72,06	375	21,74

Facet	Tags related to aboutness of article	%	Tags not related to aboutness of article	%	Empty and meaningless tags, tags applied to deleted pages	%
Cancer outcome	74	18,05	292	71,22	44	10,73
Preventing cancer	162	47,65	142	41,76	36	10,59
Health in general	33	43,42	31	40,79	12	15,79
Causes of cancer	231	44,51	228	43,93	60	11,56
Politics	4	23,53	10	58,82	3	17,65
Life with cancer	108	44,26	114	46,72	22	9,02
Life in general	18	19,35	63	67,74	12	12,9
Alternative treatment	18	12,95	40	28,78	81	58,27
Voluntary work	41	35,96	52	45,61	21	18,42
Other tags	126	20	431	68,41	73	11,59
Numbers, codes, words with no meaning	7	0,25	74	2,68	2676	97,06
Statements, questions	91	23,21	270	68,88	31	7,91
Name, e-mail address	48	3,51	1232	90,06	88	6,43
Empty tags	0	0	0	0	3259	100
Time, season	29	30,85	64	68,09	1	1,06
Things, goods	58	7,07	674	82,2	88	10,73
URL	13	38,24	20	58,82	1	2,94
From, genre, document	117	34,62	191	56,51	30	8,88
Other persons	40	19,14	140	66,99	29	13,88
Statistics	41	49,4	40	48,19	2	2,41
SUM	5040		12226		7986	

Table 19 Aboutness in tag facets

Other facets that attracted good tags from an aboutness point of view are:

- Examination
- Family, friends and others
- Preventing cancer
- Statistics

Four facets stand out with tags that are mostly bad, from an aboutness point of view. These facets are tags that describe *places*, tags about *alternative treatment*, *names and e-mail addresses*, and *tags about things*.

The *place* tags appear more often on pages where one may start to look for places. 95% of the tags that describe places, both geopolitical places and more general description of hospital wards and so on, were applied by external taggers. And most of them, 99% do not describe the content of the article. But these tags are still more frequent on geographically related articles like a list of wards or brances with information about their geographically whereabouts. Aproximately 1/3 of the tags in this facet do not describe the content of the article at all. Since they are often applied to geographically related articles, they look like searches or questions: “Does the Danish Cancer Society have a second-hand store in Copenhagen?” Aproximately 2/3 of the tags describe an aspect or a fragment of the content of an article that is related to the artiticle it is applied to. Usually, they are applied to articles that are local start pages for a topic, like an article with a brief introduction to igen, a chain of secondhand stores. This page links to a map of all brances in Denmark. But the cities mentioned on the map are applied to a more general page that only links to the map, og links to an article that links to the map. Thus, the topic is present at this section of Cancer.dk, it is too specific for the article it is applied to.

This pattern is also visible, but not as clear, on the external tags applied to food pages, where ingredients or dishes are applied as tags to overview pages that links to recipes. The food facet is not totally dominated by external taggers. 11.7 % of the tags are internal, but most of them, 65% of all internal food tags, describe a topic or sub-topic of the article.

The facet for tags about things also shows a similar pattern. Typically, these tags are applied to general web shop pages. Sometimes the tag describes things available in the Cancer.dk web shop, sometimes not. Thus, the things may be relevant in a Cancer.dk setting, like the tag *armbånd* (bracelet) applied to pages on sun protection. To see the relevance, one has to know that the web shop has bracelets made ov UV-sensitive pearls. Thus, they can be used as a tool to protect oneself (and especially kids) from the sun. But most of these pages do not mention these bracelets at all, thus the tags are not related to the aboutness of the actual page.

Tags with e-mail addresses are rarely related to the aboutness of the page they are applied to. No Cancer.dk pages are about individual e-mail addresses. These tags seem to be a result of misunderstandings and attepts to use the tagging feature for direct communication. E-mail tags are frequently applied to pages with forms, so there is reason to believe that the tagging field was seen as a part of the form.

Tags on *alternative treatment* need a different explanation. The nature of this topic do not make it easy to see a pattern similar to *places* and *things*, or even *e-mails*. But it is possible to see them as attempts to search for topics within alternative treatment. Cancer.dk has information about alternative treatment. But if users find it insufficient or are unable to find it, tags can be seen as attempts to find information.

Because of the low number of tags after the feature change, and the high number of facets, I have not separated the tags based on whether they were applied before or after the feature change.

7.2.12.6 Aboutness in URL sections

Cancer.dk is organized in sections. In the transaction log, these sections can be extracted from the URL with which articles are represented. The data in this chapter is taken from TagAppliedLog. This means that URL sections whith articles or pages that nobody ever applied a tag to, are not included here, only URL sections with one or more tag applied to one or more article or site in the section.

Because of the low number of tags after the feature change, and the high number of URL sections, I have not separated the tags based on whether they were applied before or after the feature change.

Among the Caner.dk sections with the most tags applied to them, *ditliv* (your life) and *omsorg* (care) have the highest share of tags that refer to the content of the articles. *Stoet+os* (support us), *pinkcup* [name of a breast cancer fund raising activity] and *Forskning* (research) have the highest share of meaningless tags. The remaining 13 sections are dominated by tags that do not refer to the content of the articles.

The sections at Cancer.dk attracts different tags when it comes to aboutness. If high quality tags are tags that relate to the content of the articles, some sections attract more of them than others do. A closer look at this can be taken when the aboutness of tags at sections are also distributed on internal and external taggers, see Figure 27 and Figure 28.

The internal taggers, who applied more tags related to the aboutness of the articles than the external taggers, also vary their tags from section to section when it comes to aboutness.

From the figures compared it is visible that internal taggers did not apply tags to articles in as many URL sections as external taggers did.

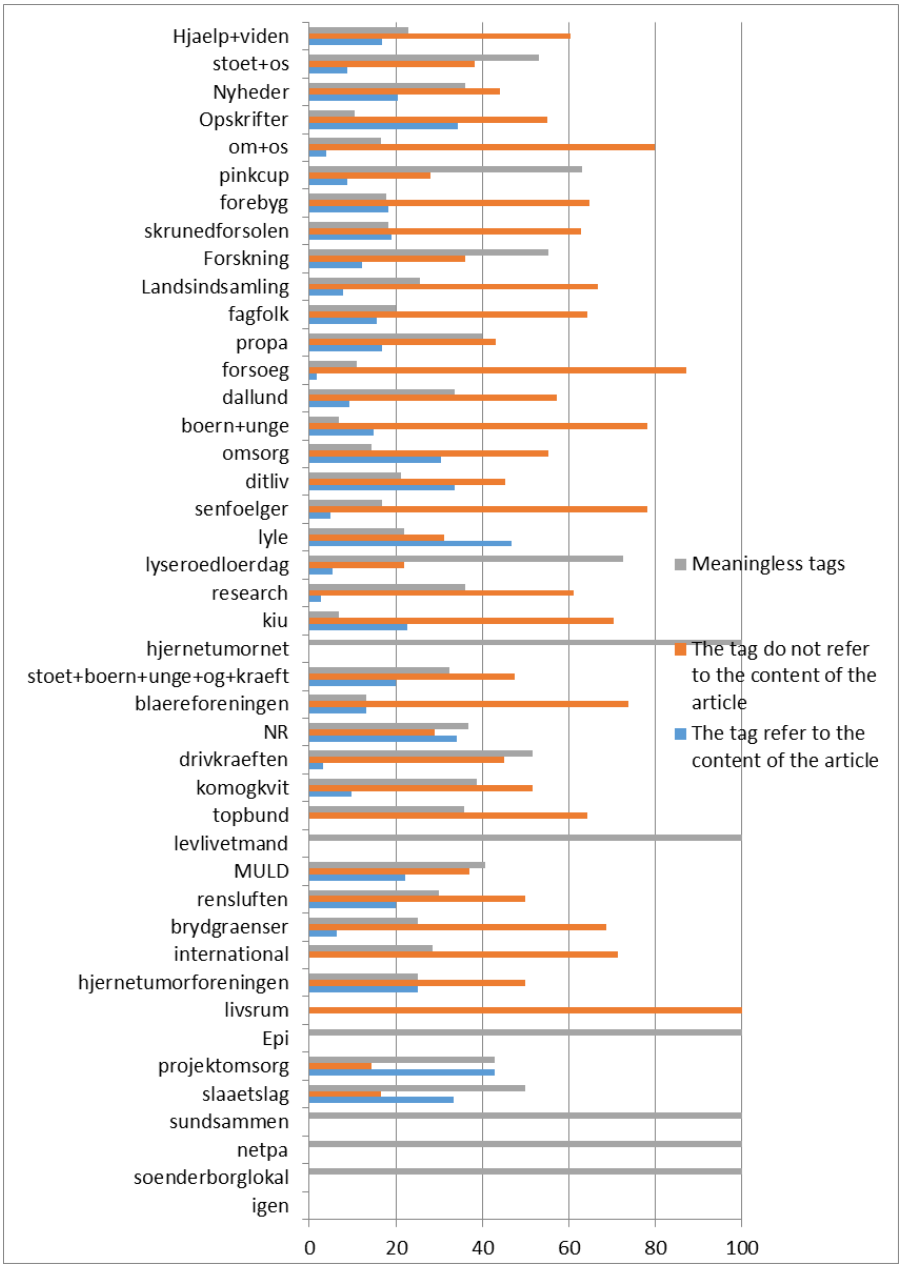


Figure 27 Tags applied by external users, distributed on Cancer.dk sections. Share of total number of internal tags in the section.

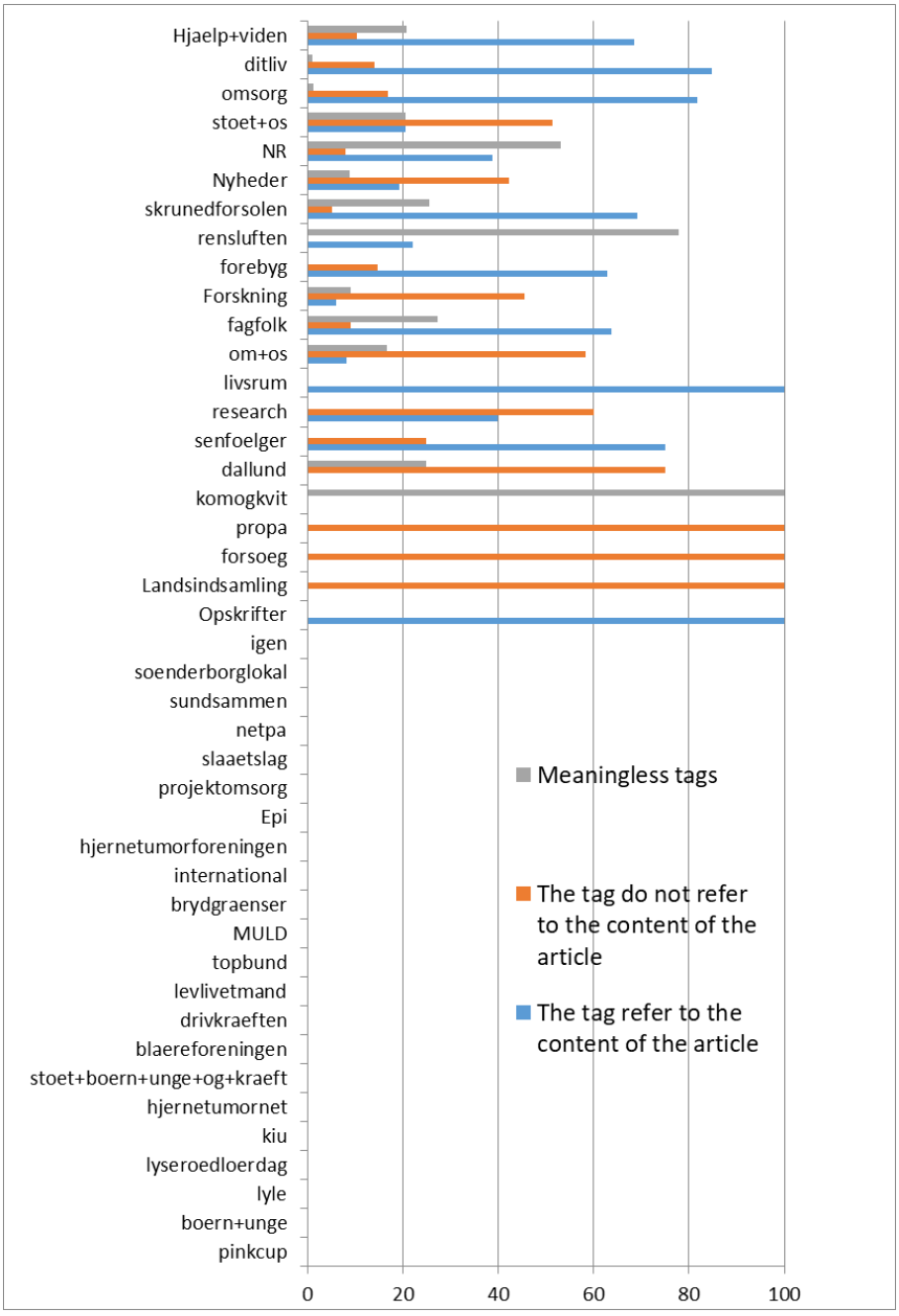


Figure 28 Tags applied by external users, distributed on Cancer.dk sections. Share of total number of internal tags in the section

7.2.13 TAG EDITING

The tagging feature included an administration tool to edit or delete tags. This was used mainly to delete tags. Usually, tags were deleted because they did not correspond to the content of the article. All use of this tool was logged into TagAdministrationLog. It shows the tag, whether it was deleted, excluded or edited, at what time. If the editor edited the tag, the transaction log includes the new tag. An edited or excluded tag could be treated in the administration tool again. A deleted tag was gone.

If an editor deleted an article with tags applied to it, these deleted tags would not be in the log. Thus, the log only contains cases where a tag was treated independently, and the editor wanted to edit or delet the tag itself. Tags that were deleted with articles were most likley deleted for other reasons than the content or other porperties of the tag itself. On the other hand, some tags that were deleted together with the article they were applied to, would have been deleted or edited if the article were not to be deleted. Because of this, some tags are 'missing' in the TagAdministrationLog.

The first months, the tag editing tool was mainly used by the chief editor Øyan. Then the responsibility was handed over to editor E3 (see chapter 9). She was a student who worked part-time for the Danish Cancer Society when she took over the responsibility. Øyan gave instructions based on his initial experience with editing tags. The other editors E1 and E2 had access to use it the editing tool, but only editor E2 reported that she had edited and/or deleted tags, and only rarely.

Tags that did not relate to the aboutness of the article were most challenging to the editors. From the beginning, these tags were seen as promblematic and irrelevant to the tag collection. First, the editors were hesitant about deleting them also because they knew we were studying the tags and I had stated that I did not want them to delete too many tags. In addition to this, the tag-editing tool that not not give enough information to locate tags to be edited in between the relevant ones. Only tag and URL were visible in the editing tool, not the document text itself. But in cases where it was obvious that the tag was far away from the article content, it was deleted if it was discovered (E-mail from Tor Øyan fr 09.12.2011 13:53). And the editors eventually established experience and routine when editing tags.

8,574 tags were treated in the administration tool. Most of them, 8,384, or 97.8% were deleted. 189, or 2.2%, were edited. In Table 20, the *Number of transactions* refer to the total number of treated tags, per month. Number of deleted tags show how many of the treated tags were deleted each month. *Distinct tags* show the number of distinct tags treated each month.

The option to exclude tags, and thus ban this word from being saved as a tag, was used only once. The excluded tag was *lort* (crap). Tags that were applied later, like *fuck lort* and *fucking lort* [crappier] was not excluded, only deleted. Thus, this seems like this option was not really used by the editors. When looking at distinct tags, 7,990 distinct tags were treated. 7,814 were deleted, 186 were edited and one excluded.

When compareing the number of tags treated in the administration tool with the total numbers of tags, 34% of the tags were treated in the administration tool. 33% of all applied tags were deleted.

Month	Applied tags	Number of transactions	Number of deleted tags	Distinct tags
November (from Nov. 8)	992	50	48	43
December 2011	2502	131	128	105
Jaunary 2012	2740	301	297	289
February	2319	2410	2408	2396
March	3287	1234	1230	1220
April	2719	203	168	198
May	3354	1369	1345	1343
June	2412	181	181	172
July	1655	210	205	203
August	2025	671	604	665
September	721	1249	1212	1237
October	38	276	269	270
November	134	35	35	32
December	200	220	220	212
January 2013	94	22	22	21
February 2013	59	12	12	12
Average		535,88		526,13
All year		8574	8384	

Table 20 Transactions in the tag editing tool, per month.

Figure 29 visualizes the numbers of deleted tags each month. The numbers vary and it is not clear why. But for the months after the feature change, there were less new tags to delete or otherwise treat. In February, September, October and December 2012 more tags are deleted than applied, thus older tags were treated as well as new ones.

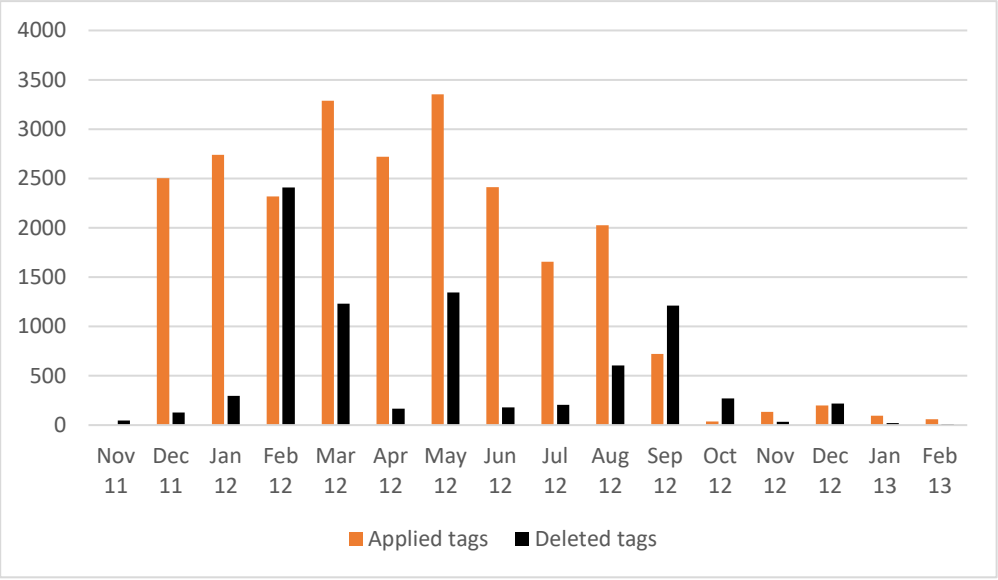


Figure 29 Applied tags and deleted tags per month

7.2.13.1 Categorization of edited and deleted tags

Unfortunately, the TagAdministrationLog did not include the id or the URL of the tag that was deleted or edited. Thus, it is complicated to say which occurrence of a tag was deleted, in cases where a similar tag was applied to many URLs, or many times to the same URL. However, a comparison of the list of tags and the list of deleted and edited ones gives numbers that indicate how often a tag formulation was deleted. When categorixed, the numbers in each category indicate how editors treated tags in various categories.

98% of all distinct tags from external users were deleted once or more. 70% of all distinct tags from internal users were deleted once or more. Thus, tags from external users were more likley to be deleted than tags from internal users. This corresponds to my expectations that users from internal ip-addresses were able to apply better tags than external users. In addition to the central editorial board of Cancer.dk, approximately 20 other decentralized editors applied tags. And all Cancer.dk employees had access to tag. But the editors who deleted and edited tags, were two of the internal taggers, and this may have influenced their judgement when doing this. Still, there is a distance between 70% and 98%.

		Deleted or edited tags	Total number of tags	%
Lay	1	2838	3284	86.42
Neutral	2	8738	11684	74.79
Professional	3	1374	1917	71.67
Hard to categorize	4	1707	1923	88.77
No categorization needed	5	6183	6272	98.58

Table 21 Distinct tags that were deleted, distributed on lay/professional categories

Professional and neutral vocabulary is more likley not to be deleted than lay vocabulary. 75% of neutral distinct tags and 72% of professional distinct tags were deleted once or more, compared to 86% of lay distinct tags. See Table 21 where tags that were deleted or edited are distributed on the lay/professional categories. The numbers do not explain whether editors were more liberal to lay vocabulary, or maybe tags with lay vocabulary were better. If so, it corresponds to a view that lay users perform better with lay vocabulary.

7.2.14 SEARCHING

The search feature's default choice was to search all words in the articles and structure of the site. Users could instead choose to include tags in their search. Users have included tags in their search in 1.22% of their searches. (Searches with no information about this are excluded.) See Table 22 for numbers related to search at Cancer.dk.

When the user had conducted one search, the result page included a search form with an additional choice. Users could choose to reformulate the search string, and/or to filter the search by category. Most of the information at Cancer.dk is categorized into these categories, and they correspond to the main browsing sections and some of the Cacer.dk sections based on the start of the URL of each page (see chapter 7.2.8 for results on tags connected to these categories.). 14.3% of the searches in the transaction log includes a filter on one of these categories.

Variable	Number
Number of user sessions[13]	4,816,300
Number of distinct URLs viewed by users[14]	622,565
Number of sessions with search	375,718
- Search for words in articles[15]	366,332
- Search both words in articles, and tags [16]	4,524
- Search type unknown[17]	4,862
Distinct search strings	69,490
Number of search filtered to a category at Cancer.dk[18]	53,689

Table 22 Key numbers on search at Cancer.dk

7.2.14.1 Tags and search terms

All together, 375,718 searches were submittet to Cancer.dk in the logging period. 7,761 of them, 2%, were empty. There are 69,899 distinct search strings. On average, a distinct search string was used 5.4 times.

A simple comparison between distinct search terms and distinct tags gives an indication on how these two kinds of user input overlap[19]. Of 10,141 distinct tags,

[13] select distinct sessionid from PageViewLog

[14] select distinct PageUrl from PageViewLog

[15] Search type="text"

[16] Search type="both"

[17] During some periods, especially in the spring 2012, there were problems with the search feature at Cancer.dk

[18] select * from SearchLog where Category not like ''

[19] Søk, Tagger med kategorier AUGUST 2013-2.xlsx

7,906 or 78% appear as search words. 22% of the distinct tags were never searched for. Thus, 78% of the tags can give a match when people serached. Compared to the 69,899 distinct search strings, 11% of them match a tag. This means that tags do contribute to Cancer.dk in terms of giving acces to articles. But these numbers do not give any information about the quality of tags and result lists.

7.2.14.2 Tags as access points to Cancer.dk

Cancer.dk used Google Analystics for statistics on how the site is used. 13,109 keywords resulted in search results where the user decided to visit Cancer.dk via a link in the result list five or more times. These were compared to 5,786 of the distinct tags, extracted from a list of totally 14,080 applied from the first five months of the tagging period. 843 keywords, 6,43%, match a tag exactly.

7.2.14.3 Tags as the only gateway to Cancer.dk

In October 2012, the tags so far were compared with all other words at Cancer.dk[20]. The comparison was conducted by students at Portland State University, supervised by the FIRE team members at this university. 173 search terms had been used to access pages in Cancer.dk where the term was a tag word on that page but no terms in the search appeared anywhere else on the page. The number is low. Since 78% of all tags were searched for, the low number indicate that many users apply tags with words that are already on Caner.dk and even in the article.

77, or 44%, of the 173 tags that are used to retireve articles, referred to the content of the articles they were applied to. 92, or 53%, did not. Four cannot be decided on, because the article was not at Cancer.dk when the tagging feature was launched and was deleted again at the point of aboutness judgement. Thus, the tag that really add something to Cancer.dk are misleading slightly more often than they are helpful. It is hard to say whether this is as expected or not. It has been found that tags are more helpful in search when they are not the only data in the index (Morrison, 2008; Pera et al., 2009). Thus, the expectations to tags alone was not high. On the other hand, these 173 tags could be better, especially since both user test persons and editors mainly agree that tags should be related to the content of the articles.

42 of the tags consisted of lay words or expressions. 87 were neutral, and 39 belonged to a professional vocabulary. See the Table 23 for details on tag properties on lay/professional vocabulary and their relatedness to the content of the article.

It seems like the tags with a professional vocabulary are better than tags with a lay or neutral vocabulary. There are more professional tags (22) that refer to the content of the article, than do not. Neurtal and lay tags show an opposite pattern. But there are only 173 tags, so the difference could be just a coincidence.

[20] Contribution, Tagger med kategorier AUGUST 2013-2.xlsx

7.2.15 BROWSING

There were 11,704,070 pageviews in the logging period. This is the number of times somebody viewed a certain URL. 250,592 of the viewed articles, 2.1%, were viewed because the user clicked on a tag. 97.9% of the articles were viewed with no use of tags.

Like many other information websites, Cancer.dk is built with a structure that facilitate browsing through different paths. The search feature can also be used to find information. The tags come as a third alternative. They can be both browsed and searched for, but still the browsing structure dominate the site. This is naturally, as research on tags so far conclude that tags are better as a supplement than as a main structure or a main index (Morrison, 2008; Pera et al., 2009). Based on this, it was expected that a minority of the pageviews would be a result of a click on a tag. The following paragraphs analyse what tags are used more than others.

Because the log for pageviews so not separate between tags applied before or after the feature change, or whether thye were applied by internal or external taggers, this analyse is not separated into those categories. Also, when users interacted with Cancer.dk, they were not informed on who applied a tag and when it was applied.

7.2.15.1 Use of lay and professional tags

From Table 23 one can see that lay vocabulary were clicked on relatively more often

Category		Number of tags	% share of tags	Number of clicks	% share clicks	Never clicked	% share of tags never clicked
1	Lay	3284	13,09	57307	26,1	1745	0,28
2	Neutral	11684	46,59	118329	53,8	1013	1,19
3	Professional	1917	7,64	21384	9,72	1115	0,21
4	Hard to categorize	1923	7,67	10230	4,65	532	0,36
5	No categorization needed	6272	25,01	12752	5,8	203	1,96
Sum		**25080**	**100**	**220002**	**100**	**877**	**4**

Table 23 Number of tags that were clicked on, categorized on lay/professional vocabulary. All tags are included.

than other tags. Lay vocabulary was used in 13% of the tags, but it constitutes 26% of the tags that were clicked on to view a page. Neutral vocabulary was used in 47% of the tags, but they originate 54% of the tags that were clicked on to view a page. The numbers for professional vocabulary show a similar pattern, but with smaller interval between the share of tags and the share of clicks.

The use of tags in category 4 and 5 are closer analysanalysed in chapter 7.2.15.3.

7.2.15.2 Use of internal and external tags

Whether the tag was applied by an intern or extern user, does not influence the use of tags. See Table 24. The share of tags that were clicked on to enter pages is 87% for external tags and 13% for internal. The shares of tags are approximately the same, 89% of the tags were applied by external users and 11% by internal.

	Number of tags	%, share of tags	Number of clicks	%, share of clicks
External	19667	89,4	884	87,1
Internal	2330	10,6	131	12,9
Sum	21997	100	1015	100

Table 24 Use of tags from intern and extern users

7.2.15.3 Use of tags according to facets

When tags are faceted, the use of tags are distributet evenly on most facets. See Table 25. Users show an interest in all parts of Cancer.dk, this is visible also in the use of tags.

The tags that include numbers, codes and words with no meanings were clicked on far more often than expected. The 2,394 tags have attracted 392 clicks, thus averagely every sixth of these tags were clicked on. It seems like these meaningless tags trigger some curiosity. That is interesting, because it indicates that when users find these tags, clicking on them is a way to find out what it means or why it is there. However, I will guess that the users did not obtain a full understanding from their click.

Facet	Number of tags	%, share of tags	Use of tags	%, share of use	%, distance
Cancer types and cancer in general	2020	11,56	44	6,25	-5,31
Thoughts, feelingd, and psychological symptoms	184	1,05	6	0,85	-0,2
Treatment	514	2,94	18	2,56	-0,39
Communication	261	1,49	11	1,56	0,07
Organizations and organized activities	719	4,12	13	1,85	-2,27
Doing something else	40	0,23	2	0,28	0,06
Symptoms and side effects	664	3,8	26	3,69	-0,11
Examination	80	0,46	2	0,28	-0,17
Family, friends and others	72	0,41	3	0,43	0,01
Body and biology	857	4,9	22	3,13	-1,78
Patients and professionals	250	1,43	3	0,43	-1
Drugs to treat cancer	158	0,9	3	0,43	-0,48
Food	1395	7,98	23	3,27	-4,72
Work	45	0,26	2	0,28	0,03
Places, including places of treatment	952	5,45	19	2,7	-2,75
Cancer outcome	215	1,23	2	0,28	-0,95
Preventing cancer	193	1,1	7	0,99	-0,11
Health in general	29	0,17	0	0	-0,17
Causes of cancer	362	2,07	7	0,99	-1,08
Politics	0	0	0	0	0
Life with cancer	60	0,34	0	0	-0,34
Life in general	29	0,17	0	0	-0,17
Alternative treatment	57	0,33	2	0,28	-0,04
Voluntary work	57	0,33	1	0,14	-0,18
Other tags	313	1,79	15	2,13	0,34
Numbers, codes, words with no meaning	2394	13,7	392	55,68	41,98
Statements, questions	239	1,37	20	2,84	1,47
Name, e-mail address	448	2,56	49	6,96	4,4
Empty tags	3260	18,66	2	0,28	-18,37
Time, season	58	0,33	1	0,14	-0,19
Things, goods	362	2,07	4	0,57	-1,5
URL	55	0,31	2	0,28	-0,03
From, genre, document	99	0,57	3	0,43	-0,14
	17472	100	704	100	

Table 25 Tag use distributed on facets

7.3 FINDINGS

Visibility of tagging feature: When the tagging feature changed in September 2012, the visibility was the only change. This resulted in a dramatic decrease in the number of tags. Of course there are variations in number of tags during the whole logging period, but not as dramatic as this. The visibility of the tagging feature influences the number of tags.

Aboutness: Analysis of the log files shows how difficult it is to apply tags. Especially the aboutness categorization reveals challenges with how tags relate to the topical conent of the article. It is, on the other hand, not clear from the transaction log wether the users themselves had a conscious attitude towards these difficulties, due to the nature of this material.

Some tags are good subject descriptions, some are not. Some express users' and patient' experiences and needs. Some of them requests answers. And quite many seem to be misunderstandings and individual thoughs. Mixed together, the tags as a whole are difficult to use and difficult to judge.

One possible explanation to the high number of tags that do not relate to the content of the articles can be that users misunderstood the tagging field for a search field. If so, it is an example of how it takes some effort to inform users how to behave on a website.

Facets: The tag content are influenced by the content of Cancer.dk. Many tags repeat article words.

The facets that fit for blog tags, needed expansion before they were used for Cancer.dk tags. This already indicates a difference between the blog tags and the cancer.dk tags: the Cancer.dk tags cover more facets.

Words that concern feelings are less frequently applied as tags at Cancer.dk, compared to Blogomkraeft.dk. This is also the case with tags that are not about the article content: wishes, seach terms. These tags can be seen as communication initiated by the user. But the content of the communication originates from the content of Cancer.dk. This influence is stronger than the users' background or situation. See the comparison between Cancer.dk and Blogomkraeft.dk tags.

The tagging feature is not a place of inappropriate or unfriendly behaviour. Such tags are rather rare. This can also be a consequence of how Cancer.dk set the tone. A serios and respectful site attracts serious and respectful tags.

User groups: The taggers' connection to the system, whether the tagger is internal or external, influence both the willingness to apply tags, the number of tags, and the quality of tags. The internal taggers are from from inside the organization. They were contacted as a group, they probably have a lojality bond to the organization. It also turned out that they were easier to encourage to tag. However, a crowd of external taggers was more stable. Their number of tags varied less during the logging period.

This result may also be influenced by the number of taggers, there were several more external than internal taggers.

The internal taggers were relatively dominating when it comes to tags about thoughts, feelings and psychological symptoms. Thus, these users applied tags more similar to the Blogomkraeft.dk tags. It seems that in this sense, internal taggers are better at choosing tags independently of Cancer.dk than the external users are. This conflicts with the expectation that end users can add new viewpoints that those within the system can not.

Tags from internal taggers are more diverse, they cover diverse categories an describe the article content from different angels. Their tags are more evenly distributed on tag facets. This may partly be due to a tendency of internal taggers to apply more tags per sessions and per articles.

Tagging is communication with the system. In an extended narrow folksonomy, there is no bookmarking feature and thus the users can not use the tags solely for their own purposes. They can not use tags to communicate with themselves. The communication is targeted towards the system and the editors and the organization behind the system.

8 HOW PATIENTS APPLY TAGS TO CANCER INFORMATION

8.1 BACKGROUND AND PURPOSE

This chapter focuses on tagging behaviour, defined as how users apply and use tags; and describe describe the charasteristics of tags. The study was organized as a usability test exploring the users' understanding, interaction and opinion about the newly implemented tagging feature. How do users tag, what is their purpose, how do they use the applied tags? Do they have an opinion about what a good tag looks like? How do people relate to their own and to other users' tags? As inexperienced taggers they do not only have to apply tags, they also must use energy to understand the system itself, its function and features.

This part of the thesis adds to previous findings by investigating taggers' viewpoints on and motivations for tagging. The study explores regular users with a scarce experience with the web and hardly any experience with tags, and older than regular taggers. Compared to previous studies on tags (see chapter 0), this study has a focus on users who are older than the regular taggers, and investigate how tagging can function as a real, democratic tool for anyone.

8.2 METHOD

Two usability studies were conducted with eight cancer patients in December 2011, one at the the Danish Cancer Society's head office in Copenhagen, Denmark and another at a gym center for cancer patients in Aalborg, Denmark. I used a procedure like the one I used when testing the prototype. The data collection focused on qualitative data providing explanations, opinions and examples of tagging behaviour. The study did not seek a representative or complete picture of tagging behaviour. The purpose was explorative, to provide insight into taggers thoughts, considerations and preferences.

8.2.1 THE TAGGING FEATURE AT CANCER.DK WHEN TESTED

The tagging feature was launched November 8, 2011. It was firs available for employees of the Danish Cancer Society, to ensure that a number of tags would constitute examples and guide the general audience when the feature was opened to them. The first tags from the general audience were applied on November 30, 2011.

The feature was easy to find, and the activity was high from the beginning. When data collection started in the morning at December 8, there were already 1,813 tags in the system. Around 1,006 of them were applied from IP addresses belonging to the Danish Cancer Society or their vendors, the remaining 807 came from outside. Most articles retrieved by the participants in this study had no or only few tags applied.

8.2.2 PARTICIPANTS

Recruitment: All participants represented the target group defined by Cancer.dk. The Danish Cancer Society helped with the recruitment. For practical reasons, all participants were cancer patients. In Copenhagen, I invited a user panel established by the Danish Cancer Society for user studies. Two cancer patients volunteered to participate. In Aalborg, members of a gym class for cancer patients connected to the Danish Cancer Society were invited. Six patients volunteered to participate. Only volunteers with some Internet experience were invited, so that they were likely to be able to use Cancer.dk and to have an opinion about tagging.

All ages were invited. Young users may be the most frequent taggers (Rainie, 2007), but older users constitute a larger part of the target group of Cancer.dk, and may also find tagging useful. More than 50% of all cancer cases in Denmark strike people older than 65 (Engholm et al., 2011).

In this chapter, numbers 1-8 refer to test subjects, in order to differentiate between them and annonymize them.

Characteristics: The participants' age were from 39 to 73, the average age was 61. There were seven women and one man. Five of them applied tags for the first time. As the study was conducted right after the launch of the tagging feature, tagging at Cancer.dk was a new experience to all of them. Seven of the participants had used Cancer.dk before, and they all had Internet experience. Two were diagnosed with cancer in 2011, but none of them regarded themselves as newly diagnosed. See Table 26 for further details.

Participant	Gender	Age	Internet experience	Knowledge of tagging before participation	Use of tags before participation	Applyed tags before participation	Time of diagnosis
1	F	39	Seven years or more	No	No	No	2011
2	F	60	Seven years or more	Yes	Yes, many times	Yes, many times	
3	F	61	Seven years or more	No	No	No	2006
4	F	65	Four to six years	No	No	No	2007
5	F	73	Seven years or more	No	No	No	
6	F	62	Seven years or more	Yes	Yes, a few times	No	2011
7	M	60	Seven years or more	Yes	Yes, many times	Yes, many times	2007
8	F	71	Less than a year	No	No	No	2005

Table 26 Participants

8.2.3 TAGGING SESSIONS

8.2.3.1 Setting

In Copenhagen, the two participants (numbers 1-2) arrived at the Danish Cancer Society's head office and met with one of the researchers. The two participants did not meet each other. In Aalborg the study took place after the weekly gym class. After a short coffee break, the six (numbers 3-5) participants gathered in a meeting room to participate in the test. They were able to talk to each other's during their questionnaire and tasks sessions. Both researchers and a research assistant were present.

8.2.3.2 Procedure

First, the participants filled out a questionnaire about their demographic data, general Internet experience, and experience with Cancer.dk. Then they solved tasks using the site. It was a goal that this part would lead the participants into a natural situation of searching information. The starting point was ideal; Cancer.dk is a prominent source of cancer information in Danish. The participants could browse or search and read along their path to an article that would constitute an answer to the task, as they would if they were looking for an answer in an information need, they had themselves. I did not ask for complete results, only one article that included an answer. This was done to leave the tasks open and flexible, so that participants could let their own interests influence part of their behaviour. During the following interviews, participants showed an interest in the content they had retrieved during the interviews. Thus, they were able to engage in the tasks.

I conducted the individual interviews. When a participant was ready, she/he went to another room with me for the interview. As some of the participants in Aalborg finished the tasks at the same time, some of them had to wait a little for their interview. They sat talking with each other and seemed to have a good time while waiting.

8.2.3.3 Tasks

All tasks asked the participants to find information at Cancer.dk, and to apply one or more tags to an article they judged relevant for the task. See Table 27. The requested information was easy to find at Cancer.dk. The tasks were about mainstream cancer-related topics, to avoid participants from getting frustrated or stressed by not finding the information.

At Cancer.dk with its extended narrow folksonomy it is not necessary to apply a tag already applied. To apply tags to an article already tagged may differ from applying tags to an article not already tagged. Therefore, the participants were given different tasks, to ensure that they all had a chance to be the first person to apply tags to at least one article. The fact that several articles could constitute an answer and thus have tags applied to them, also helped achieve this goal. There were three task sets, each with a total of four tasks. See Table 27.

Tasks			
	Set 1: Tasks for participant 1, 4 and 7	**Set 2: Tasks for participant 2, 5 and 8**	**Set 3: Tasks for participant 3 and 6**
1	Find the list of news items (or news articles) at cancer.dk. Click on a news article that you find interesting. Apply one or more tags to it.		
2	Find an article about how the Danish Cancer Society spends the money they get from their sponsors and donors. Apply one or more tags to it.	Find information about soy and cancer. Apply one or more tags to it.	Find an article about how a good work environment prevents cancer. Apply one or more tags to it.
3	Find an article about vaccination against cancer. Apply one or more tags to it.	Find an article about psychological reactions to a cancer diagnosis. Apply one or more tags to it.	Cancer.dk also has food recipes. Find a recipe for something that you think looks tempting. Apply one or more tags to it.
4	Re-find one of the articles that you previously tagged, using the search feature at cancer.dk. If you want to, apply one or more tags to it.		

Table 27 Tasks

Task 1 was common for all task sets. It was quite open and asked for a news article. In the preliminary study all five participants reported that they were interested in news from Cancer.dk, see chapter 5.3. The participants were expected to click on one of the five news items on the start page, or follow a link to older news items. One of the news items on the start page on the first day of data collection were still there on the second day of data collection. This gives a total of nine news items visible on the start page during the period of data collection. The news articles at Cancer.dk are not included in the subject-based structure of the site. Thus, all participants had a chance to apply tags to articles not already "indexed" by the Cancer.dk editors.

Task 2 and 3 differed in the three task sets. They all asked for specific information, like: "Find an article about vaccination against cancer". This increased the chances that the data material would cover many parts of Cancer.dk and hence give insight in tagging behaviour when faced to the variety of content at Cancer.dk. One or a few pages would constitute an answer to task 2 and 3 in the sets. They asked for information that could be interesting independent of what cancer type and stage the participants had, so that all three sets would lead to an interesting session for all the participants. All together the tasks led the participants to different parts of Cancer.dk and to articles including different genres and topics.

One of the tasks asked for psychological reactions to a cancer diagnosis. Previous analysis on this topic at Cancer.dk has shown differences in the level of specificity, compared to tags from a cancer blog, see chapter 5.2. Thus, it was a topic that was interesting to include in the tasks. Two tasks asked for soy and cancer, and work environment and cancer, information that could be interesting for any cancer patient. The one cited above asked about vaccination against cancer. This led the participants to articles about a vaccine against cervical cancer, a topic that had appeared in the news several times the year before data collection. Another task asked about the Danish Cancer Society itself and how they spend the donations they receive. This topic differs from the others, and was meant to be of interest because the Danish Cancer Society is almost entirely run by donations. The last task asked for a food recipe. Recipes are general popular at Cancer.dk, and now and then a certain recipe are among the most popular URLs on the site.

Task 4 was common for all task sets: "Re-find one of the articles that you previously tagged, using the search feature at Cancer.dk. If you want to, apply one or more tags to it." This was to find out how they would re-find an article with a known tag, and give them a little experience before the interview where it was a goal to find out how they reflected on the use of tags. It turned out to be a good starting point for reflections during the following interviews.

Questions about specific cancer types or treatment, and cancer outcome (survival, death etc.), were not included. This was not considered to be necessary, as it was not known on beforehand whether any of the participants were newly diagnosed or vulnerable in other ways. It was not considered important to this study to expose the participants to tough feelings; hence it was avoided. In retrospect, it is not known

whether these considerations were important. But participants were able to participate in the study with the dignity one may want to meet strangers with, including researchers.

8.2.3.4 Interviews

Interviews allowed me to explore the participants' considerations, ideas, explanations and purposes for tagging. The purpose of the interviews was to obtain insight into users of Cancer.dk opinion about the tagging feature, and their reflections on tags. The participants were encouraged to describe their experience on task solving and their tagging, and if and how tagging was useful to them. The interviews are therefore of high importance for the analysis, for the understanding of experiences with, motivations for, and challenges in tag use and tagging behaviour. Tags have enthusiastically been described as "power to the people" (Quintarelli, 2005). Thus, it is interesting to hear how "the people" describe this "power" themselves.

Main topics from Interview guide
1. Tell me a little about how easy it was to complete the tasks? 2. How easy was it to use the tagging features? **3. How easy was it to choose the words you used when applying tags?** **4. Did you use any sources when formulating your tags? Was this important / useful?** **5. In the last task you were asked to use the search field to re-find one of the articles you tagged. How did you do that? Which search terms did you use?** **6. What would you describe as good properties for a tag, based on your experience with tagging?** 7. Do you think tags are useful? 8. Do you think you will use tags and/or apply tags at cancer.dk later? Why/Why not? 9. (If yes for question 8:) What do you think that you will use tagging/tags for?

Table 28 Interview guide key questions

The interviews were semi-structured. An interview guide helped to make sure that each participant reflected on a few topics. See Table 28 and full interview guide in Appendix I – tagger study. They were asked about how easy it was to formulate tags, as this may influence the number of tags and their quality. They described good and bad tags, as this might reveal the participants attitudes towards tagging in general, and quality of tags in particular. I also asked them how they solved the last task where they should re-find a previous found article (task number 4, Table 27), to acquire reflections on navigation and how the participants see the tags as part of a larger

system. Each interview also had its own dynamics, and the participants were encouraged to give their opinions and reflections on any topic concerning tags and cancer information. The interviews provided interesting viewpoints on what properties tags should have, how they would use or not use tags, how others could use tags. Not all topics in the interview guide were covered by all participants. The interviews lasted for between seven and 30 minutes. In Aalborg the gym setting limited the time available for the tagging sessions, but mainly the interview length varied according to talkativeness of the participant.

8.2.3.5 Language

All participants spoke Danish. The interviewer spoke Norwegian. In general Danes and Norwegians understand each other when speaking slow and clear. No language problems were observed during the interviews. Questionnaries and all tasks were in Danish. In this thesis, all quotations are translated into English.

8.2.4 DATA COLLECTION METHODS

8.2.4.1 Transaction logging

Three methods were used to collect data. First, the transaction log includes all activity on Cancer.dk and thus capture the concrete behaviour. The participants' tagging activity was identified in the log using time, session and IP address from the log, supported by the interviews. This leaves a material with 98 applied tags, including one empty tag. This material may not include all the participants' activity, but it is quite certain that it includes all sessions that included one or more applied tags. Second, interviews were sound recorded and later transcribed. The interviews give explanations to the behaviour, from those who have this behaviour. Third, participants were observed during the whole sessions to acquire extra information about behaviour and how to explain this behaviour. The observation adds to both transaction logs and interviews.

8.2.5 DATA ANALYSIS METHODS

8.2.5.1 Coding

Each distinct combinations of participant-tag-URL were categorized based on the tag and article content supplemented with information from interviews. The categorization is described thoroughly in chapter 7.2.

Variable	Definition
Specificity	To what degree a tag covers the topic of the whole article, or covers a broader or narrower topic compared to the topical content of the article.
Understanding the tagging feature	To what degree the participant understands how tags are part of a structure and description of content at Cancer.dk. Understanding includes, but is more than to be able to apply single tags.
Purpose when applying tags	How do the participants explain why they apply a certain tag to a certain article?
Purpose when using tags	How do the participants explain why they click on, or search for a certain tag? And how do they expect others to click on, or search for a certain tag?
Vocabulary	How do the participants describe their own words compared to words used by health professionals and words used at Cancer.dk?
Sources for tags	How do the participants describe their chose of words, and do they reflect on how thei chose words for tags.

Table 29 Variables with definitions used for coding an analyse of interviews

Transcribed interviews were read through and examined, and so were the questionnaires. As a starting point, the material was coded and categorized to identify viewpoints and characteristics on the tagging behaviour (Kvale & Brinkmann, 2009). The coding was inductive, the final codes were adjusted throughout the coding process. Table 29 shows key variables with short descriptions. The coding was mainly data driven, but influenced by the pre-pepared questions the participants were asked.

8.3 RESULTS

8.3.1 APPLIED TAGS

Participants	Tagged URL's	Number of tags	Distinct tags
Participant 1	3	17	17
Participant 2	4	27	24
Participant 3	10	16	15
Participant 4	3	5	2
Participant 5	5	6	6
Participant 6	3	5	5
Participant 7	3	8	8
Participant 8	6	14	12
Totals:	**28 distinct URL's**	**98**	**89**

Table 30 Number of tagged URL's and number of tags applied to them

The participants applied between 5 and 27 tags each, on average 12.25 tags. Table 30 shows how many tags were applied to how many URL's by the participants. There are 89 distinct tags, 98 tags in total. When looking at distinct combinations of user, tag and URL, there are 92 such combinations. Table 31 shows an aboutness categorization of these tags. For comparison, see also Table 16, with aboutness categorization of internal and external tags.

59 tags, 64.3% (categories 1-3 in Table 31), describe the topic of the article and additionally 16 tags, 17.3% (categories 4-5 in Table 31), are related to the topic of the article. 16 tags, 17.3% (category 6 in Table 31), are supplemental or not related to the

articles they are applied to. They represent a challenge with tags, in spite of the fact that they were applied for rational reasons. They represent uncommon ways to understand the articles, or an attempt to communicate with the Cancer.dk editors through the tagging feature. The challenge is connected to any attempt to use tags as topical metadata. The tags that are not related to the topic of the article can then be seen as misleading. No topical metadata are perfect, but 17.3% misleading tags are a lot. The problem increases as there seems to be no way to clear up without challenging one of the properties that has been said to be an advantage about tags: the users' viewpoint.

Tag category	**Number of distinct participant-tag-URL combinations**	
Related tags	71	77.3%
Not related tags	20	21.7%
Empty or meaningless tags	1	1%
Total	92	100%

Table 31 Categorization of tags applied by participants. See also Table 16 for aboutness categorization of tags applied by all users.

None of the participants applied time and task related tags. These tags are also rare at Cancer.dk in general, unlike in Delicious with as many as 16% time and task related tags.

There are few synonyms within the collection of tags from each tagger (Kipp & Campbell, 2006). But in many cases, they applied tags that are synonyms to or grammatical variations of words in the article. Thus, the participants do have an understanding on the importance of synonyms in a system like this. During the interviews, they were also able to talk about this before being asked.

43% of the participants' tags are more specific or narrow than the topic of the article. This corresponds to previous findings where blog tags also were found to be more specific than Cancer.dk site structure and thus the topics of the articles at Cancer.dk (see chapter 5.2). From the transaction log and the tag study, I found that 3.90% of the internal tags and 10.93% of the external tags were too specific, so the participants or the study setting they were in differ from general taggers of Cancer.dk.

8.3.2 UNDERSTANDING THE TAGGING FEATURE

Six of the participants had been using Internet for more than seven years. The other two had a shorter experience. However, their practical skills varied, also among the experienced Internet users. One was teaching basic computer literacy; others only knew the basics. However, the participants with the least computer experience also had least understanding of the tagging feature. The ones with computer experience were the ones who were best able to reflect on different use of tags. Three participants understood the feature, three did not fully understand it, and the remaining two were somewhere in the middle. But they were all able to apply tags. In addition, they all understood that tags would be connected to articles and that everybody can read them, including staff from the Danish Cancer Society.

8.3.3 PURPOSES WHEN APPLYING TAGS

In the interviews for this study, the participants told about their purposes of applying tags and choosing which words to use as tags. These purposes are connected to the tag and its intended future use. I observed that users also might use tags in ways that they were not necessarily purposed for. In the groups below, you find four purposes of applying tags. The groups are based on intentions or purpose and not the resulting tags, thus tags may serve purposes they were not intended for.

The categories represent new tag functions or purposes, for instance compared to what was revealed in Delicious (Golder & Huberman, 2006).

8.3.3.1 Tags as topical description

Topical tags describe the content of the article they were applied to. When the study and tags were introduced to the participants, they were explained that tags can be subject descriptions or article descriptions from a users' perspective. Users may apply tags to help themselves and other users finding information at Cancer.dk. The tasks had a focus on finding information about different subjects. All this could lead to a subject focus when tagging. The participants seemed to put emphasis on this focus when tagging and talking about their tagging. Six of eight participants had a main focus on tags as subject descriptions, and most of their tags correspond with this view. Among these are the participants who had the best computer skills and best understanding of the tagging feature. Examples of their subject description tags are *soja* (soy) and *antihormon* (anti-hormone), both applied to articles about these topics.

Some of the taggers stated that they wanted tags to be exclusively topic descriptive. One stated that if tags do not relate to the subject of the article, it would decrease the credibility of Cancer.dk: "No, it should be in the article, or else […] the article will lose its credibility" (2). Thus, tags should be understandable for other users and the topical relationship between tag and article should be clear and understandable for others. Another participant was not as strict: "*I really can see the feature as an*

amateur's entrance to it, where you as user kind of expect the expert, the one who made the page, and the ones who know something about this, to have written the other parts. And then you can, as a user, add something very simple to this." (6) Here, the participant sees the tag quality from the quality and credibility of the rest of Cancer.dk.

Participants who did not apply subject descriptive tags, all agreed that such tags could be useful. The participants often connected a focus on topical desctiption to a focus on finding information. A participant said about finding words for tags: "*Then one should find the right ones,* […] *think about what people search for*" (1). Another reason to apply topically descriptive tags may be to explain something or give alternative wording. Not all the participants were able to define a good tag, but the ones who did stated that a good tag is a tag that describes the topic of the article, or at least more generally has a clear relation to the content of the article. However, to the participants, the topical tags do not have to describe the general topic of the article, it is sufficient that it describes a subsection or an aspect of its topic. Thus, topical tags do not equal subject headings. The requirements are not as strict.

8.3.3.2 Explicative tags

Explicative tags seek to explain an aspect of the content of an article, or a word in it.

Tags reflect the tagger's perspective and thus include vocabulary that users can understand (Peters, 2009). Two of the participants took this a step further and discussed to apply tags solely to explain difficult wording or other aspects of article content to others. This was something they wanted to do. To one of them, a nurse (3), "others" were laypersons. In her experience, some cancer patients did not quite understand their own illness or treatment. She saw a need for explainations and that tags are a way to explain.

The other one (6) observed that some information needed explanation and mentioned this when asked whether others could use her tags. Her starting point was more specific, as she said that tags could be applied as "*...little alerts, maybe to somebody you knew were looking for something like this. Or more general: Oh, an explanation is missing [] there is a lot of good stuff here*" (6).

Both these taggers applied topical tags that are clear and understandable. Examples are:

- *alternativ morgenmad* (alternative breakfast)
- *Fare ved rygning* (dangers in smoking)
- *Forskning* (research)
- *Sommertomater* (summer tomatoes, tag applied to a gazpacho recipe)

It is not easy to see the explainatory motive in the tags, the tags look like topical tags. However, the willingness to apply tags that are synonyms to article words can be seen as an indication on explicative tagging. Despite the similarities between topical tags

and explicative tags, the motivation is different. Topical tags describe what is already in the article, explicative tags seek to add an explanation and thus add content to the site. None of the participants described practical challenges like how to know what part of an article a cartain tag explains. If the article concerns similar or opposite concepts, this would be a problem.

8.3.3.3 Opinion tags

Opinion tags express an opinion on the content of articles. One participant (4) used only two tags, to value the article content. The tags were:

- *fremragende* (excellent)
- *vigtigt* (important)

This participant used Cancer.dk for the first time, and she really liked it. A third opinion tag was applied by another participant (8). This tag also seems to value article content, as it was applied to an article on a decrease in the number of new cancer patients in Denmark. The tag was:

- *god nyhed* (good news)

This type of tags is well known from other systems like Delicious (Golder & Huberman, 2006).

8.3.3.4 Tags that express wishes for more information

Tags applied to express wishes do not necessarily relate to the content of the article. Instead, the tag is applied to express a wish for information. It is a message to the editors saying: "I want Cancer.dk to have more information about this." One participant (3) stated that she had applied tags to express her wish for additional information. She explained that it was impossible to find information about her specific cancer illness, and she wanted Cancer.dk to include it. To express this, she applied tags. This purpose of tags is opposite to subject description. The article she chose to apply these tags to was one that she probably found when solving the task: "Find an interesting news article..." The article is titled "Overraskende stor stigning i kræftoverlevelsen" (Surprisingly large increase in cancer survival) and mentioned different kinds of cancer. The applied tags describe the cancer type the participant had. She wanted Cancer.dk and the web in general to have more information about this cancer type. From the interview, it is clear that she did not only want information about survival concerning her cancer type, but information in general. Thus, it seems like she chose this specific article simply because it was there. The participant told that she previously had told the Danish Cancer Society about her information need, but that nothing had happened on the site. She saw this as another opportunity communicate her wish.

Another participant (6) formulated the contrast between topical tags and tags that express a wish and said: "*To me it is negative to write about something it [the article] is not about, this is one of my opinions*" (6). She did not want these kinds of tags in the system.

8.3.4 PURPOSES WHEN USING TAGS

The four purposes of applying tags described above say something about intended functions when applying tags. But the participants also had some thoughts about using tags, and they reported on their own use of tags. This usage can be categorized in **tag usage purposes**:

1. Exploring another taggers' use of Cancer.dk
2. Searching, within Cancer.dk or from search engines
3. Re-finding articles
4. Finding explanations

The participants reported that they had used tags according to the two first categories when solving the tasks. The other categories were discussed during the interviews.

Using tags to **explore** other taggers' use of Cancer.dk came as an interesting extra feature for two of the participants. They had not expected this to be interesting but ended up spending time looking at tags. They found it interesting to see others' use and interpretation of articles. One of the two participants had found an article tagged with the names of some Danish cities. This made her wonder, and she had no idea why this particular article had these tags. She did not mention the possibility of somebody making an error and seemed to take it for granted that the tags were there for a logical reason.

The participants used tags to **search** for information, through seeking and browsing, when solving the tasks. They were told that their tags were not indexed and searchable immediately, but they searched other tags in the system and discussed how tags could improve search results. This function corresponds to the understanding that tags should or can describe the subject of the articles, and is the most widely tag using function in the material.

To some participants, the search function within Cancer.dk was not a prominent part of their mental model of cancer.dk as a system. They did not reflect deeply on how tags influence internal search at Cancer.dk. Others took a broader view when thinking about searching tags: One participant was skeptical to the lay words in tags, but then started thinking about how they could lead people to Cancer.dk: "*If there are tags at Cancer.dk, then they will also be out at Google, or what? That could be a good thing* (Hvis der kommer nøgleord ind på cancer.dk, så vil de komme ut på Google, gør de ikke det)" (7). Tags was seen as useful **access terms** when people use Google to find cancer information. Lay words could help them find cancer.dk easier.

To use tags to **re-find articles** may be seen as a sub-function of searching. But the participants discussed this partly as something separate. They saw this as a way to find known items in an efficient way. "*If I had known that I applied this a few days ago, and I rapidly wanted to re-find it, I would clearly use it* [the tag]" (6). Like searching, this use of tags was connected to topical descriptive tags. None of the participants suggested using other kinds of tags for this purpose. Tags formulated for self-reference like the ones found in Delicious (Golder & Huberman, 2006), were not in the participants' mind.

To use tags to find explanations is a way of using tags only visible in the participants' intention to help others. These "others" may be laypersons, newly diagnosed cancer patients or friends, or family of cancer patients. One said: "*So I have been thinking, also when I sat and applied those tags, how one feels, especially when newly diagnosed* […] *It is a chaos*" (2) These "others" may use the explicative tags and thus find explanations.

The participants regarded themselves as people who had learned about cancer and about finding information about cancer through their illness and treatment. None of them reported on finding explanation or learning something new from tags only, or planning to do so in the future. This was not necessarily expected, as the tagging feature was rather new at the time of data collection. The tasks may also have directed their focus, but it is possible that tags have nothing to offer in this case. After all, tags are mostly single words.

However, the participants reported that they learned from solving the tasks and reading articles on Cancer.dk in general. To find out if tags really can help people find explanations or learn something, it is most useful as a starting point to study people with less experience with cancer than the participants. It is also possible that helping is more important to the users than getting help.

8.3.5 VOCABULARY AND SOURCES FOR TAGS

The vocabulary problem (see chapter 4.4) is visible in the interviews. Participants talked about tagging to explain things to other users. They also touched upon it when talking about tags in general. One said: "*One should not use too fancy words, but such words for ordinary people who have nothing to do with the field who do not think of these technical terms. It [fancy words] is of no use. This is not how you write tags*" (2). This view was explicitly shared among five of the participants. As often intended with tags, all participants found it natural that tags were words or expressions from daily language. Tags were seen as the consumer or lay vocabulary, as opposite to the vocabulary used by health professionals.

The participants do not agree about the need for consumer vocabulary to be available as lead-in terms at Cancer.dk. This seems to have two reasons. One is their need no learn a professional vocabulary. They claimed to have done this to a certain degree, and two of them said that they preferred to use professional vocabulary when finding

information about cancer. This is something other researchers also has observed (Glenton et al., 2006; Smith et al., 2002). The participants are on the other hand aware of that others may not have this knowledge of professional vocabulary. One participant (7) seemed to worry that tags may bring slang words into the system, and found this unnecessary. From the list of tags at Cancer.dk, there was no need to worry, but this was clearly something that influenced his view on tagging.

As mentioned above, differences in vocabularies between patients and professionals' origins in different mental models concerning the domain of health and medicine. None of the participants did go that deep into this subject. They did not reflect on how differences in mental model affect their relationship to Cancer.dk and the health services.

Participants did not have a strong opinion about the source of their tags. Their view on this seems to be dependent on their understanding of the whole system. If they did not know that title words were searchable, they were more likely to want them as tags, to make them available for search.

8.4 MAIN FINDINGS FROM TAGGER STUDY

This chapter reports a study on the tagging feature at Cancer.dk. Eight participants were interviewed after completing four tasks each. The tasks asked them to find information at Cancer.dk and to apply tags to articles when the information was found. The participants liked the tagging feature and were all able to use it, even though some of them did not understand the complete functionality of the tagging system.

The participants retrieved articles and applied tags because they were asked to. Thus, their behaviour was not necessarily fully natural. But their reflections on what purposes tag have and can have for themselves and others are funded on their own experiences.

Our data indicate a connection between computer skills and understanding of the tagging feature on one side, and a focus on tags as topical descriptors. But other tag purposes are also found: tags to explain, tags to value articles and tags to express wishes for additional information in articles. Topical description is the dominating purpose when applying tags at Cancer.dk. All participants agree that a good tag has an understandable relation to the subject of the document it is applied to. With a focus on the use of tags, slightly different tag usage purposes appear: Exploring other taggers' use of Cancer.dk, searching, re-finding articles and finding explanations. Other findings reveal that taggers re-use words from the documents they tag or other information available at the moment of tagging (Bar-Ilan et al., 2010), but the participants do not reflect much about their sources for tag formulation.

All the participants agree that subject descriptive tags are good, but they do not agree on whether other types of tags add value to Cancer.dk. Especially tags that express wishes for information are controversial. One participant applied such tags. Other

participants had a strict opinion that such tags were useless and could even disturb the credibility of Cancer.dk. This is correct. The wishes from an individual tagger do not give any help in the dialog between users and Cancer.dk, and there is no reason to believe that the editors are able to identify the wish act upon it. These tags are misleading and not wanted in the system.

The focus on helping others may have been strengthened by the tagging feature as an extended narrow folksonomy. The feature does not support personal information management through tags. The helping focus is also visible in the applied tags. Many of them are synonyms or reformulations of article words.

9 EDITORS' VIEW ON TAGS AND TAGGING

9.1 BACKGROUND AND PURPOSE

In December 2012, I interviewed the editors of Cancer.dk about the tagging feature and the tags. The feature had been present on Cancer.dk for abuot a year, so the editors now had experience with it. The tagging feature connects closely to Cancer.dk and is maintained by its editors. They write articles for the website in cooperation with medical and social experts, and structure the site. They edit and delete tags when needed. All together, these persons are responsible for the system of which tags are a part.

The end users' view on tags and the tagging feature is the most important when one wants to find out how tags can be used on a site like Cancer.dk. However, the editors are the ones who maintain the site. They sum up the responses and the behaviour they receicve and observe, and change the site or individual articles if they find it necessary. They use experts and other recources available in the Danish Cancer Society when needed, and they decide what to do when they discover conflicting needs among the users. Their decisions frame the tagging feature and give it its shape. Thus, they could give information about their thoughts and actions on tags.

I wanted to interview them to obtain their view on the tagging feature and the tags. I wanted to know their opinion about the tagging feature, good and bad tags, and any damage they believe the tags may do, if any. The editors did not access and analyse the transaction log or similar data, but they have a close relationship with Cancer.dk and its users. I wanted them to share this information.

9.2 METHOD

In addition to the editor in chief, there are two full time employees and one part time employee in the editorial board. I interviewed all three of them. To help them, there are decentralized Cancer.dk editors in all departments of the Danish Cancer Society. They are responsible for content on parts of the site, but they also have other responsebilities. I did not contact any of them. It could have been interesting to have their view as well, but in this project is was a priority to talk to users and the editorial board. The users because they are the target group of the tagging feature, and the editorial board because they are in the crossroads between the users and Cancer.dk' information policy.

Again, I chose interviews to let the editors give their own opinion. An interview guide contained questions to ask. See Appendix 2. It includes a set of questions that I asked all three editors. In addition, the interview guide has suggestions for follow up-questions and alternatives. I wanted the editors to comment on some of the results from the user study: Tagging purposes, use of tags and sources for tag formulation. Thus, this information is listed in the interview guide for follow-up questions. I did not give the editors the information immediately in the main questions, because I wanted them to give their own view first.

The interviews were conducted in 2012, December 13-19, by telephone from Oslo. The editors used phones at their desks at the Dansih Cancer Society in Copenhagen.

Methodically, these interviews were carried out in the same way as the interviews with the user test persons. But we did not meet in person, only by phone. This of course give limited possibility to observe non-spoken signals. On the other hand, these interviews were longer, and they were a follow-up on a long project that they had been a part of for over a year. The theme of the interviews was something the editors had a professional relationship to. Thus, I expected them to be able to describe their opinions with no particular restistance and that talking to them on the phone would give a sufficient imige of their opinions. This view was confirmed by the communication I had with both the editor in chief and the editors themselves before the interviews: They seemed to be polite, but also honest and open about their view on tags and tagging. During the interviewes, the editors also included criticism of both the tags and the tagging feature. See chapter 9.3 for results. I see this as a second confirmation that this time, telephone interviews gave enough information about the editors' opinions.

The tagging feature was funded partly by the Danish Cancer Society and partly through project funding from the research team. This may lead the editorial board to be more polite and positive than they would if they talked to researchers who were not involved in the tagging feature and who were not specifically involved with tagging at all. On the other hand, in this sense, the editors are also responsible for the site and its content. They did not take part in the tagging project to produce research results, but saw it as an opportunity to test tags on the site. When the logging period was over, they decided to take the tagging feature off the site. This was known to all of us before the inerviews. All together, there is reason to believe that the most polite solution for the editors was to give a balanced view on tags and tagging.

Again, the interviewer spoke Norwegian and the editors Danish. I discussed with editor E3 wether English was better, she was not very much used to Norwegian language. But we ended up with Danish and a Danish-influenced Norwegian instead. No language problems were observed during this or the other two interviews, or afterwords when the interviews were transcribed and examined. In particular, I looked for misunderstandings because of different pronunciation or similar words that have different meaning. Because Danish numbers can be complicated to a Norwegian, I made sure during the interviews that I had heard what they had meant to say.

9.2.1 THE EDITORS

The three interviewed editors, all of them work together and know each other, together with the editor in chief, are the complete central editorial board behind Cancer.dk. They are the ones that makes decisions and do what it takes to carry them out. See an overview in Table 32.

	Editor E1	**Editor E2**	**Editor E3**
Relationship to the Danish Cancer Society	Full-time, experienced editor	Full-time editor, started in the job after the tagging project started	Part time editor, worked for less than a year
Gender	Female	Female	Female
Interview duration	34 minutes	41 minutes	47 minutes

Table 32 Interviewed editors

Only one of the editors (E1) had been in this position since the workshop in 2008. I met her there. The other two became a part of the editorial board later, and did not have the same inside knowledge about the tagging project and its history. Editor E3 worked part time for the Danish Cancer society, and is the one person who edit most tags. All together, the three editors had a different background different familiarity to the tagging project. This may incluence their viewpoints and feelings about the tagging feature, and thus the interviews.

The editors agreed on many things when it comes to tagging. The differences between them were mostly differences in what emphasis they put on diverse aspects on tags and the tagging feature. An example is to what degree they believe tags that they regard as bad will disturb users.

9.2.2 DATA COLLECTION AND ANALYSIS

I transcribed the interviews and read them thorough. As a starting point, the material was coded and categorized to identify viewpoints and characteristics on the tagging behaviour (Kvale & Brinkmann, 2009). The coding was mainly data driven, but

influenced by the pre-pepared questions the participants were asked. Thus, I coded and looked for their opinions about:

- Good and bad tags
- The tagging feature
- To apply tags
- Who should apply tags?
- Usage of tags
- Tag editing

I also read the transcribed interviews to find information related to findings in review, blog tag analysis, tag analysis, and tagger study. Thus, I coded and looked for:

- Tag facets
- Purposes when users apply tags

Some of the items above are closely related: Who should apply tags is related to what porposes these users have when tagging. And tag facets are related to what usage of tags is possible.

9.3 RESULTS AND DISCUSSION

9.3.1 USAGE AND OPINIONS OF THE TAGGING FEATURE

The editors' practical use of Cancer.dk differ. Editor E1 said: "I never use the search field because I know where things are." (E1). Editor E2, on the other hand used the search field a lot when she is looked for something, because she knew the site so well that she always knew what to search for. They have opposite behaviour, but similar explanations, and both make sense.

Editors E1 and E2 have not used the tags themselves. Again, they said this was because they knew the site so well and thus did not need tags. This is a sufficcient explanation. But it is worth asking whether their general experience with tags at Cancer.dk may also be an explanation. The editors had not tried to identify explicative tags or tags that expressed wishes for more information. These tags could inspire them to do changes in the article according to the taggers' suggestions. When asked, this was a new idea to them.

When editors applied tags, especially in the beginning of the tagging period, some of them applied tags that were already in the text, and some did not. This seems to be the same as for general taggers; at least some participants in the tagger study applied tags they had already found in the article text, some did not and saw this as useless.

Their opinions about the tagging feature changed during the project. At project start editor E1 were open and positive, and hoped that the tags would improve the search feature and user's general ability to find information. With a years' experience, she felt that neither the feature nor the tags had fulfilled the expectations. She saw several challenges: In the beginning, the tagging feature slowed down the site. In addition, the tags seemed to describe what the users could not find, not the content of the article. Editor E2 were more skeptical from the beginning, based on her own experience that she rarly applied tags or used tags when she visited other pages. When the feature had been live for a year, she concluded that it did not fulfil its purpose. Editor E3 was the newest of the editors, so she could not report on her expectations before the tagging feature was launced. She liked the idea that users can index content. But from the tags she concluded that the users did not understand the tagging feature.

All three editors liked the change in the tagging feature that happened in September 2012, approximately three months before the interviewes. See chapter 6 for more about the feature and how it changed. They mainly liked it because the feature was less visible and thus attracted less tags. At the time of the interviews in December 2012, their opinions also seem influenced by another change that happened in October. The site was moved to a new server and now loaded much faster. Thus, the number of tags was at a level they could control, and their feeling that the tagging feature delayed the site, was gone. The feature itself was not changed in September, only its visability. Editor E2 told that she did not believe that is was easier to understand, but there was a hope that those users who found it were more likley to understand it and apply good tags.

Editor E1 was quite happy with the feature for applying tags after the feature change. "Because it is such a discrete feature when it is sort of collapsed. And designwise it harmonies with the sorrondings. So, it is- yes. It is kind of huddled together so that I can ignore it if I want to, and easily find it if I need it. So, I think that works, really." (E1). But when she describedthe feature for applying tags itself, she is more critical: "And then *apply*, *apply tag(s)*. Well, you do not get much help. When one is standing at a specific article and click on *apply* […] There one must understand what tags are. And then people maybe write, what they are looking for." (E1). (The Dansih word she used was *tilføj*, the word on the button to apply tags.) It seems like editor E1 agreed with E2, those users who found the feature after the feature change, were more likley to understand how to use it.

The experience from the user tests reveals that only a little explanation was needed. Then even inexperienced users were able to apply tags. This information was available on the tagging feature (see Figure 3), but there is of course no guarantee that users read it. The editors however observed clear difficulties with the tags, that may

originate from the feature design itself. The feeling they had, that the feature gave too little information, first seems to originate from the tags that it resulted in, not from their first meeting with the tagging feature.

Editor E1 tried to, but was unable to describe a perfect tagging feature. But she was clear on what she wanted such a feature to do: This goal that a tagging feature should attract tags that are not already in the article, that gives increased value for search, indicate a view on tags as information description. Thus, tags should not be communication, as many tags seams to be.

Tags were not the main interest or focus of the editors. Cancer.dk and the information they provide were. When they realized that many tags did not meet their expectations, they lost interest. Tags were a side project, and none of them seemed to have the full overview on tags and tagging on Cancer.dk.

9.3.2 GOOD OR IDEAL TAGS

The editors' definitions of good tags correspond with how they saw tags could be used: for searching and browsing. Secondary, good tags add value and provide information about an article's content after you found it, no matter how you found it.

I asked the editors to give descriptions and examples on good or ideal tags. Synonyms that supplement the words in articles are prominent in the answers. "There may be three words for it, but we had to choose one." (E1). The other two words are then ideal tags for that page.

The editors were already aware of the importance of synonyms when they wrote texts for Cancer.dk. Synonyms give alternative access points to articles, and can also inform users on diverse words used to describe similar phenomena. With cancer information, users may not be aware of diverse wording. For instance, they need to learn that two words describe the same cancer type or examination. When the editors described a cancer type, they often made sure that both the Danish and the Latin name for this particular type is included in the text. When drugs were listed, they often used both brand name and generic name and thus included different words that described the same drug. For more general words, it is not as easy to provide alternative words in a text, so many times there was room for the taggers to apply synonyms. Thus, the tags served some of the same functions as lead-in terms or index terms would serve in a controlled vocabulary. Editors seemed to see tags as the indexing system the site did not have, and evaluate them as such.

Some synonyms may be impossible to use in the text. If the editors write *coloncancer* and *colon cancer* in the same text, one of them will look like a misspelling. Therefore, if one is in the text, and the other one applied as a tag, both variations are covered. Mix of Latin and Danish is the same. *Breast cancer* or *brystkræft* may be used in the articles, but *breastkræft* or *brystcancer* are mixed compounds and not as good. However, some users use them and search for them. If these kinds of compounds are

applied as tags, they may serve as lead in terms and the text will not be disturbed by these compound words that are not correct because of the language mix.

A good tag gives a resyme of the text, or the essence of the text, stated editor E2. It should be given from the user's viewpoint, as opposite to herself and all those who have a professional relationship with cancer and cancer information. But the main point was the content description, so that tags could be used for search and retrieval. Her explanation was broader than just saying that tags are synonyms to aricle text words. But her examples confirmed that she included the synonym tags in her resyme-description of good tags.

9.3.3 BAD, UNWANTED AND DISTURBING TAGS

Editor E2 had two ways to define a bad tag. First, she said: "A bad tag is something that confuses more that it benefits." (E2). Then she added: "A bad tag is everything else but these […] words that describe the text." (E2). The first one is natural and important, but not very concrete. It leaves the question: Confusion and benefit for whom? Users are diverse. The second definition is just the opposite of a good tag to her, a tag that describes the content of the article. This means that good and clear tag that expresses an opinion about the article, or a wish for information, is still a bad tag. This is in correspondence with what she said about tag purposes (see chapter 9.3.4).

The editors expressed different opinions about how damaging bad tags are. Editor E1 saw them as noise. This could be a challenge in searching and browsing, but not necessarily. Editors E2 and E3 connected bad tags with confusion and even loss of credibility for both Cancer.dk and the Danish Cancer Society. I also ovserved this diversity among the user test persons. Some found that bad tags gave bad credibility; others did not see bad tags as a threat at all. This revealed diverse views on tags as a part of a home site.

Editor E1 saw tags that did not describe the content of the article as noise. Examples were misspellings that misguide users and thus lead them to poorer search results because the misspelling may not be present in the articles that are most useful to the user. In general, these tags are a challenge, but not a cathastrophe.

Editor E3 wanted good tags to be clear and short, words that others could easily decode in order to understand its relation to the article. She also ended up excluding opinion tags and wish tags as good tags. She found these types of tags interesting, but they should be a part of a separate feature, dedicated to opinions or content wishes.

Editor E1 gave two examples of tags that users frequently applied to the page with contact information (http://www.cancer.dk/om+os/kontakt/), which she regarded as bad or at least unfortunate. The first example is *lotteri* (lottery), the other one is *adresseendring* (change of address). The first tag does clearly not describe the content of the contact information page, and it would be better to apply lottery-related tags to the lottery-related pages. But the other tag was not as clear as an example. Among

many addresses, there was an e-mail address to use if you wanted to give the Danish Cancer Society your new address. Therefore, you were looking for this page! The tag did not describe the complete content of the page, but how to use one of the e-mail addresses. The tag therefore was useful, especially since there were no other pages dedicated to change of address. I do suspect that this editor may have forgotten this part of the page. However, it may indicate that editors prefer topical tags to tags that describe things to do or tags that serve other purposes. They confirmed this when directly asked.

When asked whether there were tags that were both good and bad, the tag *terminal* (terminal) came up as an example. This word is not uncommon in Danish, but taken from Latin and thus a foreign term. Ediotor E1 said: "If I do not know what it means, and I click on it. And I land on a page about my illness. And I actually did not want to read, because I was not ready […]" E1. Then she concludes that this is a general danger with using Cancer.dk. Even though it is a goal to protect the users, they may stuble on information they did not want at that point. Tags like *omgivelsenes reaksjoner* (reactions from your surroundings) and *arvelighed* (inheritability) are examples of the same kind. This is related to a general experience that the editors expressed, that to use Cancer.dk as a patient, you have to acknowledge that you are a patient. This may be a hard and a slow process, and stumbling on unwanted information may be a part of it. Editor E2 mentioned tags that are interesting to some users, and not to others. Her example was tags applied to an article about brain tumors. If the article has tags with my type of brain tumor, it may be a good. If the tags name other types of brain tumors, they are irrelevant to my need.

For patients and their friends and family, there are risks using Cancer.dk. You may find information that you do not want, or more information than you can absorb in the moment. Editor E2 wanted tags to be a part of the balance, not to give users information they did not want, and not to overwhelm the users. She mentioned the tag cloud as a site that had the possibility of overloading users with links to a large amount of information. Editor E3 did not see this as a problem, but was more concerned about bad tags in the tag cloud, meaning tags that did not lead to articles with the same content as the tag.

9.3.4 PURPOSES WHEN APPLYING TAGS

During the interviewes, I mentioned the purposes for applying tags observed in the tagger study (see chapter 8.3.3), and asked for their opinions. The editors mainly saw the tagging feature as a way for users to supplement the work they did write and structuring articles. This was visible in their view on ideal tags as synonyms that supplement article terminology. It was also visible when they commented on tags that explained something to other users. Ideally, they believed articles should be in a way that made this tag purpose unnecessary.

According to editor E1, a good tag can supplement the content of an article to a certain degree. The suggested tags would cover what was missing in the article. An article about courses of treatment, with a link to an external page about course of treatment for a number of cancer types, had tags with the name of a few of these cancer types. She suggested to apply the remaining cancer types as tags as well, so that the tags could be lead in terms to both the article and its important links to the external page. The tags would not describe the content of the article, but the content of the linked article. Editor E1 based this suggestion on the history of the article. When the tagging period started, the article included many links, one for each cancer type with a formal guaranteed course of treatment. When the courses of treatment changed, it was decided to reduce the links to only one. This was easier to maintain. But since the original list of links included information about what cancer types was included in the list of formal guaranteed courses of treatment the article lost this information. As the list was not replaced in the article text, a second choice would be to apply it as tags.

Editor E3 observed many tags that related to small parts of the article. Examples were articles where a drug was mentioned, and then its name turned up as a tag. Articles about new research results attracted tags about what cancer types these results might relate to, wether they were already mentioned in the article or not. To sum up, editors accepted tags as topical descriptors, and included tags that were too specific or even described links, as long as they were topical.

Editor E2 was clear that she did not like opinion tags and tags that express wishes for more information. Such purposes should be in separate features, and Cancer.dk already had a form where you could contact the editors to express your wishes for more information: “So if someone express a lack of information about a specific [cancer] illness, one should not write it as a tag, but express it in some other way.” (E2). From a users’ point of view, this may not be as easy, and the tagging feature seemed to be en easier channel for some.

Editor E3 suspected that some articles established a culture for what tags users are expected to apply. Her example was tags like *Min mor er død* (My mother is dead) and *Jeg savner min far* (I miss my father). If a page has a few of those tags, other users could be encouraged to apply similar tags. This might absolutely be the case. However, it is also possible that the article text itself attracted such tags. The analysis of tags showed that the share of tags related to the aboutness of the articles varied in different sections of Cancer.dk.

When deleting tags, editor E3 said she primarily guessed that tags that did not describe the content of articles were misplaced searches. The thought of tags as wishes for information, was new to her. Wether this would have incluenced her future tag editing behaviour, is not clear.

Editors wanted tags to be a part of Cancer.dk and blend in as a smooth part of the site. The important priorities that goes for Cancer.dk, also goes for tags at Cancer.dk. Tags should not disturb users more than necessary, and they should be a part of an efficient tool that support searching and browsing and support the credibility of the site and

thus the Danish Cancer Society. This contradicts another view that the editors shared: That tags are av voice from the users and by nature do not always blend smoothly in on Cancar.dk

The editors had a clear picture of when they communicated directly with the users, and when they did not. They did not see tags as part of this communication. To them, tags were a part of the strufture of the site, an indexing from the users' viewpoint. They preferred to communicate with the users through dedicated forms on Cancer.dk, or e-mail. This distinction is not as clear for all users. Some users saw the tagging feature as a way to communicate, and thus used the feature accordingly.

9.3.5 WHO SHOULD APPLY TAGS?

From some of the editor's suggestions and opinions, it may seem like they want only editors to apply tags. This is not the case. They still mean that the tags have an important role as a supplement to the content they produce themselves. They are all very much aware that users do have a different perspective on cancer information. But the high number of tags that does not seem to be related to the content of the articles they are applied to, makes them question whether it is possible to find a clear voice from the users among the tags. "The tags should ideally catch what we do not know about the users", said editor E1. When asked directly about who shold apply tags, the editors' answers are mainly the users, but also editors. Thus, she does not want to change who have access to apply tags.

Editor E2 said that she believed an interest in- or an understanding for information technology is important when users apply good tags. Other users are less likley to understand and use the tagging feature. This corresponds to findings that taggers are early adapters of techonology (Rainie, 2007). The user study also revealed that those with the best computer skills also had the best understanding of the tagging feature. With a tagging feature that is open to anyone and visible at all pages, it is not easy to restrict the feature to those who understand the feature and are able to apply good tags.

9.4 MAIN FINDINGS FROM INTERVIEWS WITH EDITORS

The different purposes users have when they apply tags seem to be a challenge to the editors. They know that it is difficult to distiguish between these purposes when you look at the tags only. It is hard to say wether a certain tag is a description of the article, a supplement, a wish for more information, or what. Then it is hard to use the tags for searching and browsing, because you never know what lies behind a tag: an article about the topic that the tag describes, or an article that lack this topic?

Cancer.dk already had features for user feedback. The editors referred to this when they rejected tags that expressed wishes for information and attempts to communicate

through tags. They prefer this solution, because then it is clearer what the user message is.

From the interviews, it is my impression that the editors in a way did not want tags, but a controlled vocabulary. This would fulfill some of the purposes that both editors and users have when they apply tags. A subject language that includes synonyms and possibly also include relations between terms (hierarchial relationships etc.) would give lead-in terms that users and editors need. Such a solution is however contradicted by the editors' view that tags are mainly the users' voice into the system. A controlled vocabulary can never replace this, as the editors were also clear about.

The editors see tags that do not relate to the article content as a main challenge with tags on Cancer.dk. Tags with inappropriate or unfriendly content is not a problem to them. They agreed that these tags should be deleted as soon as possible. But since there were relatively few such tags, they did not bother to mention this as a challenge.

10 DISCUSSION

In this chapter, I will discuss the findings mainly in chapters 7-9 and give examples from the data material. When relevant, I will also comment on findings in the preliminary studies (chapter 5). I will mirror this to the theoretical framework in chapter 3.

The discussion will give answers to the research questions (see chapter 1.2).

10.1 WHAT CHARACTERIZES TAGS ON CANCER.DK?

(See Research question 1: What characterizes tags on Cancer.dk?)

I analysed tags from two websites when working with this this thesis, preliminary studies included. Blogomkraeft.dk had a narrow folksonomy, a blog where only the bloggers apply could tag, and everybody could read and use the tags. Cancer.dk had an extended narrow folksonomy. For both websites' tagging features, the possibility to use the feature for individual document- or information management was limited. When a user applied a tag, it became a part of a common pool of tags. There was no individual collection, of neither tags nor documents, to which you applied tags. There was also no possibility to correct a tag, the editors in the Danish Cancer Society had to do this. When editing tags, the editors did not communicate with the taggers of individual tags.

Tags may have diverse qualities; they can be good or bad. In this thesis, opinions about this are reported and discussed. The clear definitions of a good and a bad tag is individual and may vary among the users. As a starting point, I see a good tag as a tag that is meaningful and useful to many users. And a bad tag is meaningful and useful to few or no users. In between them are tags that are neither good nor bad, not too good, not too bad etc.

10.1.1 LAY AND PROFESSIONAL VOCABULARY

(See research question 1.1: Are tags characterized by lay or professional vocabulary?)

Tags can be categorized as lay and professonals. The relative number of lay, professional and neutral tags vary between internal and external taggers, and before and after the feature change.

To my knowledge the categorization of tags into lay an professional vocabulary is new. For a tagging system, the amount of tags with words from a professional

vocabulary seems large. After all, tags can be seen as the users' possibility to index and describe documents. On the other hand, users come with all kinds of expertise, including medically professional vocabulary.

Both internal and external taggers applied a high share of neutral terms. Both groups also applied more lay terms than professional ones. I assume that most of the external taggers were laypersons when it came to medicine in general and particularly cancer medicine. However, the external taggers applied the highest share of the professional terms. Results from the usability study can shed light on this pattern. Participants stated a willingness to learn and use the professional vocabulary, and two of them preferred this vocabulary because they regarded it as more precise and better for information searching and in communication with their therapists. This view is in line with research that reveals how professional vocabulary is better suited to match the controlled vocabulary used to retrieve information in bibliographic databases.

Others have a professional background themselves, one of the participants was a nurse. Another explanation can be found in the participants' wish to explain: If you add a professional word to an article where a corresponding lay word is used in the text, this can be seen as a way to transform the article to an explanation of the professional word. One can also see the differences between internal and external taggers as differences in their appearance (on appearance, see chapter 3.6.1.1). External users who apply tags play a role in a performance dominated by a field of knowledge that they need to learn or at least cope with. Thus, they use the professional words more often. Internal users who apply tags play a role as teachers in the universe of cancer. Thus, they need to use the precise words to let everybody understand.

I identified facets where tags were dominated by neutral vocabulary. Examples are *Life in general* and *Work.* These facets belong to an everyday sphere where neutral vocabulary is natural. It seems like the lay/professional duality is only visible on articles about topics within health and medicine. Thus, editors who worried about the consumer vocabulary problem only have to worry about articles dedicated to medicin and healt topics.

After the feature change, external taggers applied fewer neutral tags and more lay tags. I have no explanation for this change. But the number of tags after the feature change is low, 61 lay, 154 neutral and 33 professional tags, so the change may be accidental. In retrospect, it would have been useful to conduct more interviews a year later than my original user interviews, in order to get more explanations on the behaviour of taggers. Unfortunately, that was not possible

The tags on Cancer.dk seems to be based on a use warrant, as expected. When users learn a professional language and use it for tags, this have similarities with what could be found in a literary warrant found in literature about cancer and medicine. But when the users apply them as their own words, and there are no other rules for the selection of words, the tags are applied based on a use warrant. However, the difference

between internal and external taggers when it comes to lay and professional tags is not big.

Some tags can be dependent of each other, it is likely to believe that sometimes when the tags *breast* and *cancer* are applied to the same article, they belong together. But there is no other evidence or indication of tags applied to Cancer.dk based on a structural warrant.

Lay vocabulary were clicked on relatively more often than other tags. Lay vocabulary was used in 13% of the tags, but it constitutes 26% of the tags that were clicked on to view a page.

The user test persons did not agree on whether they preferred lay vocabulary or not, in their overall contact with Cancer.dk and the health service in general (see chapter 8.3.5). But they did expext tags to be dominated by lay vocabulary. The fact that lay tags were used more, indicates that these tags have something more useful than the professional and neutral tags.

10.1.2 THE TOPICAL CONTENT AND ABOUTNESS OF THE TAGS

(See research questions 1.2: What is the topical content of the tags and how does this relate to the aboutness of the documents?)

Tags have been compared to subject headings, because in many systems tags and subject headings have similarities in content and wording. From interviews with patients and editors of Cancer.dk, these similarities are also visible in their expectations to tags. Tags are to a certain degree expected to be subject headings from the users. The tags on Cancer.dk in general do not live up to these expectations. 19.96% of the tags describe the topical content of articles to which they were applied. Empty and meaningless tags were easy to delete and remove from the system. These 31.62% of tags were thus not very visible to the users. The remaining 48.42% of tags are meaningful words but not topical descriptions of the articles. This is still a large number, and these tags are hard to distinguish from the topical tags. The usability of tags as topical descriptions is thus restrained by the presence of tags that do not relate to the content of articles. (All numbers are taken from Table 15.)

Interviewed patients suggested other purposes for tags, and these purposes are visible in the tag log. They saw tags as a communication with the Danish Cancer Society and with the system itself. Other patients and the editors rejected these purposes. But there is also reason to question the taggers ability to apply topical tags. With Goffmans words, they are actors who do not know what play they were in.

The summarized aboutness categories in chapter 7.2.12.1 is not the only way to group aboutness categories and in short say: These are the tags that are related to the topical content of acticles. In chapter 7.2.12.1, and the numbers from Table 15 mentioned above, categories *General topic*, *Exact topic*, *Exact aspect of topic*, *Topic mentioned*

and *Indireclty related* were regarded as categories for tag related to the topic of the articles. If tags from categories *Too general* and *Too specific* are also included, 30.83% of tags relate to the aboutness of the article, and 37.55% of tags do not. Both editors and patients were quite liberal when it came to this, but I never asked them to judge or evaluate specific tags. And if they see it this way, this is not the same as finding as many as 30.83% of tags useful when searching or using tags in other ways.

10.1.2.1 Specificity

(See research questions 1.2: What is the topical content of the tags and how does this relate to the aboutness of the documents?)

It is known from subject indexing that subject languages often recommend using the most specific term possible (see chapter 3.1.4). From the aboutness categorization of tags, the categories *Exact topic* and *Exact aspect of topic* qualify as "the most specific term possible". Only 4.85% of tags belong to these categories, all tags considered. But the users were unable to point out this strict view on topical descriptions. The other topical tags do not follow these rerommandations, I have defined more specific and more general tags as topical as well. This is in accordance with the results from interviews with both users and editors. In other words, the recommendation to use the most specific term possible when making topical descriptions of documents is not intuitive to users and editors.

Cancer.dk differ from Delicious, as Munk & Mørk observed it (2007), when it comes to specificity. They reported that tags can be compared to basic cognitive categories, and thus often more general than the most specific trem possible, for the whole document. Cancer.dk tags are opposite. Only 4.98% of internal tags and 1.50%-1.57% of external tags are more general. On the opposite side, around 45.36% of internal tags and 17.74%-20.07% of external tags are more specific. (Numbers from chapter 7.2.12.1.) These tags describe only a part of the documents to which they are applied. Thus, Cancer.dk tended to attract more specific tags than Delicious did.

One reason for this, could be different level of knowledge between Delicious users and Cancer.dk users. The level of knowledge will influence the basic categories, or the user warrant is a specific. Cancer patients have competence on their own illness; thus, their user warrant is more specific on this field than it is for others. From the interviews, it is also clear that the participants applied tags based on the topic of parts of documents. This gives more specific tags compared to the topic of the whole article. Delicious, on the other hand, cover all themes, not only cancer. Thus, users can not be expected to be experts in all fields, and their user warrant may be more general. This is a speculation, and I do not know of studies where level of knowledge for Delicious users and Cancer.dk users are studied.

Another observation also influences the specificity. Cancer patients' willingness to help each other, may lead them to be specific. If they want something to be explained, they need to describe it precisely and clearly.

A third point is the fact that the tags on Cancer.dk form an extended narrow folksonomy. Users do not have a separate pool of tags, and thus it is less attractive to apply tags that are only of interest to oneself.

10.1.2.2 Aboutness categorization of facets

(See research questions 1.2: What is the topical content of the tags and how do tags relate to the aboutness of the documents?)

When Cancer.dk editors wanted a tagging feature in the first place, they wanted the users' view on their articles, implying that this view differed from their own view. They implied, not only that users may find information through other descriptions then the ones they could make, but also that the aboutness may differ depending on the user. A user's view on the aboutness of an article may differ from the editor's view on the abuotness of the same article. They may both be rigth in their desrciption, only they have different roles towards the article content.

10.1.2.2.1 Facets with a high share of tags that relate to the aboutness of articles

In chapter 7.2.12.5 I listed facets that attracted good tags from an aboutness point of view:

- Communication
- Examination
- Family, friends and others
- Preventing cancer
- Statistics

The *Communication* facet included tags that were requests for communication with the Danish Cancer Society, or attemts to communicate via tags directly. Thus, it is not clear how many of these tags were intended as article descriptions and how many were applied with other purposes and then happened to also describe the content of article.

The attemts to communicate that are visible in these tags, have similarities with what we can find in Twitter hashtags. The communicative purpose is the same. However on Twitter the tweet with tags is a message that reaches readers directly, this is what

Twitter do. On Cancer.dk the tag is not applied in a setting meant for communication in this way. These Cancer.dk tags are also formulated in diverse ways. One can see that the user do not follow a script or a culture for communicating. They are attempts to communicate, but du not succeed.

For the other facets it is hard to explain why they stand out in this way. But among the tags categorized as *Examination*, *Family, friends and other*, *Preventing cancer* and *Statistics* there are not only many that relate to the content of articles, but also a high share of tags that do not relate to the content of articles. Thus, the facets are characterized by a low share of empty and meaningless tags.

10.1.2.2.2 Facets with a low share of tags that relate to the aboutness of articles

External taggers applied 95% of the tags that describe places, both geopolitical places and general descriptions of hospital wards and so on. Most of these tags, 99%, did not describe the content of the article to which they were applied, they were too spesific. Still, these geographical but non-topical tags are more frequent on articles where geography is an aspect, like a list of wards or brances with information about their geographically whereabouts. Thus, these tags, for example *København* (Copenhagen) applied to an article that is a starting page for second-hand stores, looks like a search or question: “Does the Danish Cancer Society have a second-hand store in Copenhagen?” The starting page for second-hand did not mention Copenhagen, thus the tag was not related to the topic of the article. But the tag is probably not chosen by accident.

This pattern is also visible, but not as clear, in the facets Food and Things. For example, a starting page for recipies, has tags applied to it that express specific ingredients or dishes. However, a starting page for recipies do not include these details and thus the tags are non-topical when applied to it. The facet for Name and e-mail address have a low share of tags related to the aboutness of the articles, but here no such pattern is visible.

Taggers do not apply non-topical tags randomly. The overall theme of the article or the group of articles does influence the tags. But the purpose of these tags is not clear.

10.2 THE TAGGING FEATURE

10.2.1 THE TAGGING FEATURE AND ITS INFORMATION TO THE USERS

(See research question 2: What role did the tagging feature on Cancer.dk itself play?)

In the tagging feature on Cancer.dk (Figure 1), users got information about tags and tagging. This constituted the context in which users applied tags. The text over the tagging field said *Tilføy nøgleord* (apply tag(s)). Thus, it stated that it was possible to apply a tag, and that you could do it here. This implied some assumptions about the user: They had to understand what a *nøgleord* (tag) is and have some idea on the point of applying them.

For users who know tagging from systems using English language, the information had to make it clear to them that *nøgleord* meant tags. This was a dilemma when setting up a tagging feature in a non-English language, assuming that the English word was known and recognized to some users. It was hard, or maybe even impossible to find a Danish word that could both give the phenomena a Danish name, in order to include novise users, and at the same time give associations to the English word for those who were familiar with it.

One of the most used and most studied programs with tags is Delicious. Their interface does not give much more information about tags and tagging. Only a text, *Tags:* and the field to write these tags. A question mark next to the text, gives information when hovering the mouse over it: “Tags are an important part of organizing on Delicious. Using tags that are smart and simple benefits both you and the community.” This scentence is followed by practical advices on punctuation, spaces and so on. There is no further information or definition of the term *tag*.

There are however three major differences between the tagging feature on Cancer.dk and Delicious. The *first* difference concerns the feature and the documents available for tagging. On Delicios, the tagging field is a part of a system where metadata in the core of the user activity. They have found an URL and use Deliciuos to organize it into their profile. Other metadata fields, such as title and URL surround the tagging field. On Cancer.dk, the tagging field implies actions that is not in the core of the user activity in the same way, which is to find cancer information. The tagging feature is surrounded by information that is not interactive or social, and not connected to a user profile. In Goffmans words, the setting is different, or the stage decorated in different ways. This may influence taggers when applying tags, with possible different tags as a result.

The *second* difference concerns the users. Research has found that many taggers are early adopters of technology (Rainie, 2007), and there is reason to believe that Delicious have more of these users than Cancer.dk. Cancer.dk users have diverse

experience with technology. They use the site based on an interest in a topic, not an interest in technology. Cancer patients are also older than the general population (Engholm et al., 2011). On the other hand, Cancer.dk may not be a favoured source for older cancer patients. Still, Cancer.dk users seem to be less experienced with technology. The personal front of many did not enable them to apply tags that were useful. They did not know what play they were in.

A *third* difference concerns how users may form teams. I have assumed that most external users did not know each other and thus did not form teams or feel as part of teams. As cancer patients they may feel like a part of a group. But as taggers they probably did not. The tagging feature, the setting, simply did not invite them into teams, directly or indireclyt. I believe that teams can be formed through a tagging feature in two ways. Either, taggers must be enabled to create profiles and thus introduce themselves to other taggers. Or, the tagging feature is connected to a cause or a limited subject where users can be expected to share a common interest, such as a local history photo database. Cancer.dk did not provide any of these mechanisms. Cancer information could be a cause and a limited subject to form teams in this sense based on. If so, it just did not happen. Cancer patients and cancer information seekers are, after all, diverse individuals.

The possibility to form teams may have influenced the bloggers on Blogomkraeft.dk who applied tags to their own postings (see chapter 5.2). Here, the bloggers vrote about themselves and could read how the other bloggers presented themselves. Thus, they learnt from each other and could adjust their behaviour based on what they learned. In Goffmans words, they formed teams and adjusted their performance based on the information about the other team members.

10.2.2 THE TAGGING FEATURE AS A PLACE TO PERFORM TAGS

(See research question 2: What role did the tagging feature on Cancer.dk itself play?)

Taggers seem to need a context to provide good tags, for instance a tagging feature that let them see tags as metadata. Cancer.dk did not fully provide this context.

It is not easy to include a tagging feature on an information site like Cancer.dk. In this case, there is no lack of tags, but there is no ‘control’ of their content and usefulness. The success of a tagging feature depends on not only the tagging feature itself, but also the users and how the feature corresponds with their user behaviour and their interaction with the site.

The internal taggers applied more tags that are topical. If this type of tags is wanted, one idea could be to let only internal users apply tags. This would contradict the original idea with tags on Cancer.dk: A main motivation for the Danish Cancer Society to include a tagging feature on their site was to include the users’ voice. But this users’ voice was also intended to improve information retrieval on the site. For this purpose, topical tags are better suited, and internal users are a better source for

such tags. However, only 21.41% of internal tags qualify as subject headings, internal taggers are not an excellent source for tags that are similar to subject headings. But they are better than external users.

No matter how the quality of tags is defined, the tagging feature itself seem to influence how successful the tags are. In Goffman's words, the Cancer.dk tagging feature constitute a setting. If the performers do not know how to performe in this setting, the setting itself has to give them hints and help. This is a challenge. A tagging feature cannot take up too much place, not on the site and not in the users' mind. Thus, instructions and labels need to be short and easy to read and grasp.

The experience from the user study on the newly launced tagging feature did not predict the high number of non-topical tags. The four purposes of tags indicate a number of non-topical tags, but not as many as the present result. This may indicate that the short instructions that these users received as test persons, were enough to give their tags a profile that is not typical of all tags. It is unknown how the Cancer.dk tagging feature could incorporate them and thus influenced all tags. However, the differences between internal and external tags shows that some user groups need fewer instructions than others do.

The chief editor had a hyphotesis on tags that were not topical descriptions. Instead, some users misunderstood the tagging field as a search field. They believed that they submitted a search but applied a tag to the page. Then, if they realised this, the users did not have a possibility to delete the tag afterwords (E-mail from Tor Øyan fr 09.12.2011 13:53). This suggests that users were unaware that they were in the front region. I am unable to reject this hyphothesis. But if this was true for many users, it is natural to expect a change in the aboutness of tags after the feature change. The change redused the visibility of the tagging field, and thus should have redused this seemingly accidently use of it. Did this happen? The relative number of topical tags decreased slightly, from 19.99% before the change to 18.58% after the change. Admittedly, the number of non-topical tags decreased from 48.55% to 42.88%, and the number of meaningless and emty tags increased accordingly. (Numbers from Table 15). When looking at only external tags before and after the feature change, the picture is different. They applied 15.17% topical tags before the feature change, and 18.43% after, an increase. The number of non-topical tags decreased from 52.03% to 42.96%, and the number of empty and meaningless tags increased accordingly. Thus, external taggers improved their tags slightly after the feature change, in terms of topical tags, but in total numbers of tags the share of topical tags decreased because internal taggers only applied only one tag after the change. Altogether, I believe the experience from the feature change neither confirm nor weaken the chief editor's hyphotesis, due to the small changes.

It was a challenge for the editors to decide what to do with the non-topical tags. From the beginning, they were seen as problematic and irrelevant to the tag collection. According to the performance metaphor, deleting tags would be to take a show off-stage not allowing the audience to see it. The editors were reluctant, but this was the

solution they ended up with: If it was obvious that the tag was far away from the article content, it was deleted as soon as it was discovered (E-mail from Tor Øyan fr 09.12.2011 13:53).

With tagging seen as a performance, there is a mismatch between the audience's expectations and the performer's actual performance. The taggers, as performers, are not aware that they are expected to perform tags that are topical descriptions.

If topical descriptions, explicative tags and opinion tags are categorized as tags that describe the topical content of articles, only tags that express wishes for more information cause the non-topical tags in the transaction log. The interviewed users cannot be seen as representative for all users of Cancer.dk. But it is an indication that only one of them applied tags to express such wishes. All the 12,227 tags that do not describe the topical content of articles are not likely to be wishes for more information.

In addition to this, I observed that usernames and lottery numbers are in the log as tags. They were probably not meant as tags, but the users seemed to have difficulties finding the correct field to apply them to.

When looking at the facets, there may be groups of tags that have other purposes. E-mail addresses, usernames and other codes can be attempts to personalize the tagging feature. With an extended narrow folksonomy, you cannot use tags to build up a personal collection of documents. But if you apply an individually unique tag to all the documents you want in such a collection, such as your name or e-mail, this can be used to identify such a personal collection, nevertheless. In addion to your individual tag, you may further diverse this collection with other tags.

In Goffmans' terms (1959), this is a way to intentionally move the focus of the tagging feature from the front region or stage, to the back region or backstage. Everybody can see your tags, so in a way you are still performing. But the focus is on your own personal collection rather than the joint pool of tags.

10.2.1 NUMBER OF TAGS

(See research question 1: What characterizes tags on Cancer.dk? , and research question 2: What role did the tagging feature on ancer dk itself play?)

Although I have not studied tags form systems outside the Danish Cancer Society, I believe this study enables me to compare Cancer.dk tags to tags form other systems. Both the study of tags and the interviews with taggers revealed tag categories that differ from observations in other systems like Delicious (for instance Golder & Huberman, 2006). This comparison is also in line with the nature of this thesis as a case study. One cannot compare tags from various tagging features directly without also discussing the various tagging features themselves. Thus, here I will discuss

properties of tags found in this project and compare it to properties of tags found in other studies.

Before the feature was live, I discussed with the other researchers and the people in the Danish Cancer Society whether there would be enough tags to analyse during the logging period. We did not discuss the possibility of a too large number of tags, and I assumed it would still be possible to find a way to analyse the tags the way I wanted.

Systems like Deliciuos, Citeulike and Flickr attract large amounts of tags. On the other hand, smaller specialiced systems attract fewer tags (e.g. Broughton & Budzak, 2010; Peterson, 2009). These smaller systems often cover a limited topical domain. Sometimes the tagging feature is based in an information site, and not a social site (Kalbach, 2007; Peterson, 2009). This is the case for Cancer.dk.

Cancer.dk also had a tagging feature that did not fully support tags for self-organizing purposes. Users have diverse motivation for applying tags, and a limited feature may limit the number of users who are motivated to apply tags (Ames & Naaman, 2007). On the other hand, Cancer.dk was and is a system with many users. From the people behind the site, I found that many of them were eager to help each other's. Thus, if users saw tagging as a way to help others, there was a chance to attract many tags. When users applied tags to explain things to other users, this indicates that they in fact tried to help others (see chapter 8.3.3.2).

The review refer to experiences regarding participation in systems with tags (see chapter 4.3.3). Depending on the system, between 6% and 28% of the users of the systems had applied tags (Fox & Jones, 2009; Rainie, 2007). This corresponds to well-known patterns concerning participation and passive reading in social media in general. Most users are passive readers of social content. Only a minority participate and add content to the site (Uden-Kraan et al., 2008).

The tagging feature attracted many tags. During the logging period, Cancer.dk had on average 184,417 users a month, according to their Google Analytics statistics. I have not counted the number of taggers among these users, and thus do not know how many of the Cancer.dk users applied tags. But in 0.35% of all sessions, one or more tags were applied. As users will have to spend one or more sessions on Cancer.dk, it is likely to believe that the percentage of users who apply tags is 0.35% or higher. And 2.14% of pageviews involve use of a tag. Thus, a minority apply tags, and the other users observe them and sometimes click on them. In this perspective, the 25,253 tags seem to be in line with or slightly lower than other systems with tags.

When subject indexing, one usually restricts the number of tags per document. The actual number wil depend on the rules and the specific subject language, but typicaly 1-6 subject headings or one classification number. 8.93 tags per URL is higher, and this average number hides that some URLs had many more tags. The number is unknown due to problems when combining TagAppliedLog and TagAdministrationLog (see 7.1.3 for details).

But the relative high number of tags give potentially more ways to find articles, and more ways to describe articles when found. With broad folksonomies, it is meaningful to let high frequent tags be more visible in the system. It is also possible to let the most frequent tags represent the document, assuming that they are better descriptions. Cancer.dk had an extended narrow folksonomy that did not encourage users to apply tags already in the system. But if the Danish cancer society wanted, the high number of tags indicate that a broad folksonomy would give enough repeated tags to discriminate between high and low frequent tags. If the high frequent tags are better descriptions, tags on Cancer.dk could have been more useful in users' information interaction.

10.2.2 THE CHANGE OF THE TAGGING FEATURE

(See research question 2: What role did the tagging feature on Cancer.dk itself play?)

The changes in the tagging feature give two different situations when applying tags. When analysing the tags, changes that occur at the same time as the feature change, probably are caused by the change. Since the tagging feature was live for almost a year before the feature change, some taggers may have familiarized themselves with the tagging feature and behaved based on this. Thus, when the change happened, they may not have changed their behaviour until later. There is no way of knowing this. Changes that can be observed a while after the feature change, may be caused by it. But these changes may also be caused by something else.

Changes that I have observed in the tags can co-occur with this change, or even be caused by it. The changes are:

- Users applied 2,557 tags per month before feature change, 105 tags per month after. The change happened instantly after the feature change and was probably caused by the feature change. The feature was less visible, thus probably fewer users found it and used it.

- Users applied tags to more URLs before the feature change, compared to after the feature change. This change may be a consequence of the lower numbers of applied tags, which was caused by the feature change.

- The number of applied tags per session decreased a little after the feature change, most months, on average. December 2012 is an exception; this month had the highest number of tags per session compared to all other months in the logging period. Thus, it seems like the number of tags per session was not influenced by the feature change.
- Most months after the feature change have a lower number of tags applied per article compared to before the feature change. Again, December 2012 is

an exception, and it is unknown wether the feature change caused the decrease in tags per article or not.

- The use of tags varied each month, both before and after the feature change. Also, after the feature change, tags applied before the feature change was still available for usage. Thus, the use of tags is less likely to be influenced by the feature change.
- External taggers applied more tags related to the aboutness of the articles after the feature change compared to before, up from 15.17% to 18.43%. The share of related tags is still low, and the change is small. But a possible explanation is that the less visible tagging feature was used by a relative higher number of users who wanted to describe the topical content of tags.
- There is an increase in the share of empty and meaningless tags from November 2012. As the feature change was in September this year, there is no reason to believe that the increase was caused by the change itself.
- There is an increase in the number of deleted tags in August and September 2012. This may have caused changes in tagging behaviour, because it affected which tags were visible to users.

10.2.3 SETTING

(See research question 2: What role did the tagging feature on Cancer.dk itself play?)

It seems like Cancer.dk taggers have less knowledge on how to apply tags and what they can expect from tags. The differences were not an important issue when we planned the tagging feature. I knew this was in a "new" setting and were not aware of any research published on tags in such settings. When interviewed, the editors did not mention these differences. But they found that the tagging feature do not provide enough information to taggers. This was especially clear to them when they had seen the high amount of tags that did not describe the aboutness of the articles to which they were applied.

The collections available for tagging in various tagging features, give different settings for users to apply tags. Twitter, Delicious and Cancer.dk with its tagging feature represent three types of tagging features. Seen from one angel, Twitter is the hardest one to use, not because of the tags, but because you expose yourself when you publish a tweet. From another angel, both Twitter and Delicious are easy to use. Both systems have a rather narrow functionality. On Twitter, users publish their few words or a picture. If they want metadata, they use hashtags. On Delicious, users store their bookmarks. It is not hard for the users to figure out what play they are in, because Twitter and Delicious are simple settings where users learn the script quite easily. If they want to characterize the websites through other means than title and URL, they use tags.

On Cancer.dk, users came to interact with information. For many users, they only interact and read, in order to be informed. Other users want to check their lottery number or find a local group they can join. If so, the actions needed to achieve their goals are different. On Twitter and Delicious, users can also have many different goals, but the script is narrower. In other words: applying tags on Cancer.dk was more difficult than applying tags on Twitter or Delicious, not because of the tagging feature itself, but because it was part of a larger script. The website itself did not only inform users on how to apply tags, but also about preventing cancer and supporting an upcoming fundraising event or take part in a local group.

The setting gave so many clues on how to behave, that many users did not find out what part of the script they were to use. Compared to Twitter and Delicious, the Cancer.dk did provide the information needed to apply tags in a good way. But this information drowned in information about other things to do in the same setting, on the same website. The result was many good tags, and even more bad tags. Thus, the Cancer.dk tagging feature ended up almost as a theatre of the absurd. It was not a setting by itself but surrounded by a larger setting.

10.2.3.1 A need for a personal profile?

(See research question 2: What role did the tagging feature on Cancer.dk itself play?)

The interviewed taggers expressed a willingness to apply tags to help others. This was more dominating than their interest in a system where they could apply tags for themselves, through a personal profile where they could easily refind their own tags and not least, the articles they to which they were applied. Their minor interest in a profile may be a result of the fact that this was not a possibility in the Cancer.dk tagging feature.

In Goffmans words, the tagging feature on Cancer.dk gave users a limited possibility to show their personal front. It seems that this personal front could bring some context to the tags and thus give all actors a script, or hints on how to apply tags. It would link tags to taggers and thus show tags as scenes in a longer play, as opposed to independent plays.

This would correspond to a broad folksonomy where diffenent users can apply the same tag to the same article. Taggers apply tags to their own pool of tags and thus create a sub-collection of their own, tagged articles. If the main motivation for applying tags is to help others, this may seem as a contradiction to make them help themselves by applying tags for themselves first. This was also partly why the Danish Cancer Society did not want a feature where you had to log on to apply tags, because this was a consequence for such a feature. They did not expect users to want this log-on feature. Maybe they were right, maybe users would not have applied tags to a broad folksonomy on cancer.dk. However, it would have added context to the tags so that they could learn a little more on tagging through their own profile.

10.3 USERS' AND EDITORS' VIEW AND BEHAVIOUR TOWARDS TAGS

10.3.1 WHAT IS A GOOD TAG, AND WHY WAS IT APPLIED?

(See research questions 3: What are the users' and editors' view and behaviour towards tags?)

A problem with the tags on Cancer.dk is that so many of them were not useful as index terms, or subject index terms. (On indexing and subject indexing, see chapter 3.1.) They simply did not meet the requirements of controlled vocabularies. On the other hand, users applied tags as a part of their information interaction (Toms, 2002); this interaction did not include any requirements for tags.

When interacting with Cancer.dk, users could not click on tags directly from the start page. They had to enter and article and find the tagging feature, and from there they could click on tags or also choose the tag browsing page for more tag-based browsing or search. Thus, interacting with tags could be their second click after entering Cancer.dk, and then tags were available for use through the tagging feature at all articles.

The analysed data do not show whether tags were useful when users clicked on them. 2.1% of all pageviews happened because the user clicked on a tag that led this certain page. The data do not show how frequently the users considered the tag as a good link to an interesting article, or not. It is likely to believe that success of tags as links to articles depend on whether the tag is a good description of the aboutness of the article or not. The interaction with tags as cues that guide users further in a process of information interaction, may have led to articles not relevant to the user. The tags may have influenced the user when interacting with the site, in an unsatisfying way. A second session of interviews with users could have given answers here. Unfortunately, this was not possible as part of the project.

The partly conflicting expectations to tags make it hard to define what a good tag and a bad tag is. But the least controversial conclusion would be to say that a god tag is a tag that relate to the content of the article. In addition to this, one could say that the most useful tag is a tag that add terms to the article that are not already there.

10.3.1.1 Tags as the only gateway to Cancer.dk

The comparison in October 2012, between the tags that far and all other words at Cancer.dk, can also bring a perspective on the quality of tags. Here, 173 search terms had been used to access pages in Cancer.dk where the term was a tag word on that

page but no terms in the search appeared anywhere else on the page. The number is low. Since 78% of all tags were searched for, the low number indicate that many users apply tags with words that are already on Caner.dk and even in the specific article.

In settings where you browse or search only tags, it is important to have tags applied to as many documents as possible. In settings where you browse or search tags *and* other data, it is not as important that all documents have tags. But tags that add new words to the documents, are always most valuable, if the new words are related to the content of the document. Then they can make new relevant documents retrievable.

In broad folksonomies, tags are applied not only to describe documents, but as a part of each users' subcollection of tagged documents (Peters, 2009). Thus, each tag, wheter it contains words from the article or not, have a role to play for retrieval in each users' own subcollection. In narrow folksonomies, there are no subcollections. The tags are still useful for browsing and to give descriptions when the article is retrieved, but for search a tag is more valuable if it adds new words to the article that are also relevant to the content.

The editors and the user test persons were not clear on whether they liked tags that repeated words from the article, or not. Especially user test persons did not make a clear distinction between the usefulness of tags that were already in the article, and tags that were not. Some of them simply did not understand Cancer.dk and its functions good enough. There is also reason to believe that this may be a hard thing to learn users. Because tags that repeat document words play an important role in some broad folksonomies, users may believe that they do in Cancer.dk too. Thus, general understanding of tagging features is not enough.

77, or 44%, of the 173 tags that are used to retireve articles, referred to the content of the articles they were applied to. 92, or 53%, did not. Thus, the tag that really add something to Cancer.dk are misleading slightly more often than they are helpful. It is hard to say whether this is as expected result or not. It has been found that tags are more helpful in search when they are not the only data in the index (Morrison, 2008; Pera et al., 2009). Thus, the expectations to tags alone was not high. On the other hand, these 173 tags could be better, since both user test persons and editors mainly agree that tags should be related to the content of the articles.

10.3.2 INTERNAL AND EXTERNAL USERS

(See research questions 3: What are the users' and editors' view and behaviour towards tags?)

Internal and external taggers differ when it comes to empty tags and tags that consists of numbers or codes. External users applied 99.36% of the empty tags and 99.75% of the numbers, codes and meaningless words. For comparison, external users applied

90.86% of the total number of tags, a slightly lower share of tags (see Table 3). Thus, internal taggers applied less empty and meaningless tags. This is probably because the internal users knew Cancer.dk better and had a better understanding of the tagging feature and their purpose of tagging. It is hard to compare internal taggers directly with elite users of Twitter (see chapter 4.3.3.6). But there seems to be similarities between these groups. Elite Twitter users seems to users " occupy central positions in Twitter networks" (D'heer et al., 2017, chapter 4.3). Internal taggers on Cancer.dk are better at applying topical tags. It is not the same, but both groups are more successful compared to other groups using Twitter and Cancer.dk.

Some facets with tags with meaningful content are also dominated by external taggers. Most tags with names, e-mail addresses and what seems like usernames, are from external taggers. Many times, this looks like mistakes where people have tried to enter their e-mail or a username as a part of an attempt to communicate in some way. Names are sometimes names of doctors or others mentioned on the site, but most times, I do not know why names appear as tags. They may be mistakes.

In December 2011, chief editor Øyan made some intitial observations of the tags. One was the amount of empty tags (meaning tags with no content, see also chapter 7.2.2). That month 13% of the tags were empty, 315 of 2,502 tags (see Table 2).

Empty tags, and tags with a content that give no meaning, can be seen as noise. They show an activity, an attempt to do *something*, but it is usually unclear what. Some of these tags are numbers or codes that do give meaning if you have the key to understand them. Some are usernames, some are lottery numbers, and some may be postal codes. Some may be misspellings that made it impossible to recognize the intended word. As the tagging feature did not give the users any means to edit or delete tags, editors had to deal with tagging mistakes. It was quite clear that they saw no challenges with deleting these tags. They also edited misspelled tags, with the same complete naturalness. They saw them as bad tags, useless tags, or misleading tags.

There is a possibility that some users would like to use tags with no meaning for individual information retrieval. If nobody else see a meaning with your tag, you will most likely be the only one who control which articles have this tag applied to it. This purpose when applying tags is also possible for other types of tags, like e-mail addresses. None of the interviewed users suggested such a purpose when applying tags. But if this happened, the deleting end editing policy for these kinds of tags may have interrupted these attempts.

There are no obvious indications in the data that this type of tagging happened. If it did, it is my best guess that it was a rather marginal way of applying tags. None of the participants in the usability study mentioned this type of tags. However, more experienced taggers could have applied such tags. On the other hand, more experienced taggers would probably have access to other tools for this purpose.

10.3.3 PURPOSES WHEN APPLYING TAGS

(See research questions 3.1: What are the purposes of users when applying tags?)

The participants in this study were not typical taggers in the way Rainie describes them (2007). They were patients targeted by Cancer.dk and were representatives of general users and tagging behaviour. Their topical and explanatory tags were applied to help themselves and others finding information. With fixed tasks, the study did not seek to fulfill the information needs of cancer patients. However, there is an interest in ones' own illness. More visible was a willingness to help other patients, which was also observed in the preliminary studies (see chapter 5). This is most visible in the explicative tags, but also topical tags and opinion tags had a helping purpose. The willingness to supply content indicates that the participants did not see tags as solely metadata. They saw them as a way to communicate. This is challenging, as it was not always clear from the tags what they seeked to communicate.

When it comes to purposes for applying and using tags, participants had a focus on **topical tags** or at least tags that relate to the same content as the article to which they are applied. These tags are useful for browsing and searching. To some, good topical tags mean good credibility for the site. Thus, it is a challenge that 17% of the tags are topical misleading.

When **tags express wishes** for more information, like the name of a cancer type applied to a random article, taggers decide to use the tagging feature as another way to express a wish. The challenge is that nobody else knows what the relationship between tag and article is. This causes problems when other users search or browse with tags or when users view tags in the article. The tags may mislead users into believing that the article has a content that is not there. There is no way for the user to know that another user applied a tag to name topics that is not in this or other articles. It also causes problems if Cancer.dk editors want to follow up on wishes, because they are mixed with a large amount of other tags. Thus, this attempt to communicate through tags fails.

Opinion tags also have a potential to confuse when searching and browsing tags. However, words that describe *opinions* about article content are usually easy to distinguish from tags that describe topical content. The tag *viktigt* (important) does not describe article content topically, and thus it will not mislead. Thus, these tags are less challenging, and they may function well as communication between taggers or between taggers and editors.

The **explicative** tags are less challenging. They relate to the content of the article and differ from topical tags only by the motive of the tagger. This causes less problems when searching, browsing and viewing tags. Explicative tags can be seen as the participants' most concrete answer to the consumer vocabulary problem. In the simplest form, an explicative tag adds lay words to an article. The participants, who preferred to learn the professional words, agreed that this could be useful for many patients. However, the connection between a word in the article, and the explanation

in a tag, may be too loose. If you do not know that the tag explains what you do not understand, the explanation is useless.

Both the tags that express wishes for more information and explicative tags can be seen as supplementing tags (Ådland & Lykke, 2012; Berendt & Hanser, 2007). They add something to the article that was not there before. If these tags were easier to distinguish automatically from other tags, they would not mislead users. Instead, they could be a help for the editors in their work to improve the site, and even inform the external users.

10.3.4 TAGS AS COMMUNICATION

(See research questions 3: What are the users' and editors' view and behaviour towards tags?
and research question 3.1: What are the purposes of users when applying tags?)

In my project, interviews with users shed a little light on the personal front. They shared about their background and their purposes when applying tags. Interviews and user studies also include face-to-face interaction between researchers and users, and between researchers and website editors. It is necessary to include this in order to obtain an overall picture of what is really going on. The participants in the user studies did not only interact with the system. They interacted with the researchers. In their information interaction, they were more or less conscious that users and editors of Cancer.dk would see their applied tags. They had a four-fold audience. The researchers in the room, the system they interacted with, the editors and other users of this system. Not all the users were aware of all the members of this audience, but they knew there was an audience.

The participants' awareness of their audience is visible in different ways. They were of course aware of the researchers who met with them. Their talk about applying tags to help others or to explain things, reveal that they know other users see their tags. Tags that express wishes for more information reveal that they knew somebody working with Cancer.dk could see their tags.

Seeing tagging behaviour and information interaction as performance implies a view on tags as communication. A performance is communicative. Cancer.dk is a tool for communication, mainly for the editors of the site, on behalf of the Danish Cancer Society. With the tagging feature, the site was opened for users to communicate openly too, on every page where the tagging feature was present. The four purposes that the users identified are communicative (see chapter 8.3.3): When applying tags as topical descriptions, the taggers often have other users in mind, and want to inform them about the topical content of the article, or help them find articles. When applying explicative tags, opinion tags and tags to express wishes for more information, the

tags are attempts to communicate more directly with other users, edtotors, or the system itself.

Editors and taggers became each other's audience, and they had expectations to each other's' communication. A main expectation was that tags should describe the topical content of the documents to which they were applied. This is in the core of a challenge with tags: If one expects them to be descriptions of documents, one must conclude that many tags are wrong.

The performance metaphor may give an impression of vanity or self-promotion. But if so, it seems to be hidden behind the purposes to communicate, to help, explain, and to share knowledge. "A performer presents an idealized view of the situation", states Goffman (1959, p. 35). This *idealization* may result in a wish to show off one's ability to explain or help others or the system. To apply tags in order to help others is a volutarly deed and there is no reason to suspect that taggers do it solely to show off their knowledge or good will. But idealization may still play a role. To many taggers, there is also a consciousness that somebody will read the tags. Other users can be informed or learn something, or editors are expected to act on the communicative tags. Sometimes there is also a fear: "Did I do something wrong?" "Will people believe that I am stupid?" "Will the system accept this tag?" When some editors and taggers saw tags as an integrated part of Cancer.dk, and worried about how some tags could mislead users, they worried about the credibility of Cancer.dk. This worry is also an attempt to do impression management on behalf of Cancer.dk.

Before the tagging feature was live, the editors expected tags to give new lead-in terms to the site, and vocabulary that reflected the users' vocabulary. When interviewed, they saw tags more as noise and there was a worry concerning Cancer.dk's credibility, because of the high number of tags that did not meet their expectations.

To sum up, taggers need a context to provide good tags. Cancer.dk did not provide this context. The high number of tags that did not fulfil the expectations taggers had to tags, indicate that taggers did not know how to performe on Cancer.dk. Thus, the setting itself had to give them hints and help. This was a challenge. A tagging feature cannot take up too much place, not on the site and not in the users' mind. Thus, instructions and labels need to be short and easy to read and grasp.

When the setting failed to guide users into applying the wanted tags, tags more similar to Twitter tags was one result. Some of the taggers wanted a dialog, attemted to communicate through tags. The tags that editors wanted was similar to subject headings, but also had similarities with tags in systems like Delicious and LibraryThing. But they also got tags with the communicative aspects that are more common in Twitter hashtags.

10.4 CONCLUSIONS

The partly conflicting expectations users have towards tags, makes it hard to define what a good tag and a bad tag is. From an indexing point of view, the most useful tag adds terms to the article that are not already there. And the least controversial conclusion from taggers and editors, is that a good tag describes the topical content of the article.

It is not easy to include a tagging feature on an information site like Cancer.dk. In this case, there is no lack of tags, but there is no 'control' of their content and usefulness. The success of a tagging feature depends on not only the tagging feature itself, but also the users and how the feature corresponds with their user behaviour and their interaction with the whole site.

There is a disagreement between taggers and editors on what a good tag should be like and for what purposes users should apply tags. This explains some of the 'wrong', irrelevant, non-topical tags: The users wanted to communicate with the system through tags, as opposed to describing the content. But misunderstandings and mistakes may also explain 'wrong' tags. Editors saw topical descriptive tags as most useful and found the tagging feature attracted too many tags that did not meet this criterion.

The results can inform the design of tagging features; visibility is essential to attract tags, and the information surrounding tagging needs testing. Adding at tagging feature as part of an information web site is a chanlleng. Because such a tagging feature is only one of many features on the site, users may find it hard to use it correctly. The disagreement between the user groups can also inform tagging features: the tags applied within the system will be influenced by who has permission to apply tags. A general audience like the users of Cancer.dk will most likely be influenced by hashtags in social media when applying tags. Thus, if less communicative tags are wanted, taggers need instructions or guidance in order to apply more desctiptive and less communicative tags.

Previous studies frequently extract tags for indexing or retrieval purposes. My study, in which all tags are included, confirm these studies as productive. The communicative aspects of tags found in Cancer.dk indicate that taggers do not necessarily distinguish between tags in different systems. When looking at systems like Twitter, tags are communicative by intent; they add information to the tweet and do not necessarily cover the topical content of the tweet. However, when moved to an information website, this behaviour is unwelcome.

In this project, I studed a certain tagging feature on a certain website within a certain subject, cancer information. This case can not be repeated, not on Cancer.dk, and not anywhere else. But the results give a picture of what tags and tagging can be in such

a setting, and what role or script users end up with. Within Gofman's model, the script and the applied tags remind me of a theatre of the absurd.

The result is that the tags altogether were hard to use for information retrieval or -interaction after they were applied. But at the time when each tag was applied, it meant something for the individual who applied it, and was a part of their information interaction. The user learnt something or got lost. Thus, the tags are traces form a plurality of information behaviour.

LITERATURE LIST

Ådland, M. K., & Lykke, M. (2012). Social Tagging in Support of Cancer Patients' Information Interaction. In G. Widen & K. Holmberg (Eds.), *Social Information Research*. Emerald Group Publishing Limited.

Ames, M., & Naaman, M. (2007). Why we tag: motivations for annotation in mobile and online media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 971–980.

Ankem, K. (2005). Types of information needs among cancer patients: A systematic review. *LIBRES: Library & Information Science Research Electronic Journal*, *15*(2), 1.

Ankem, K. (2006). Factors influencing information needs among cancer patients: A meta-analysis. *Library & Information Science Research*, *28*(1), 7–23. https://doi.org/16/j.lisr.2005.11.003

Bar-Ilan, J., Zhitomirsky-Geffet, M., Miller, Y., & Shoham, S. (2010). The effects of background information and social interaction on image tagging. *Journal of the American Society for Information Science and Technology*, *61*(5), 940–951. https://doi.org/10.1002/asi.21306

Barité, M. (2017). *Literary warrant (IEKO)*. https://www.isko.org/cyclo/literary_warrant#1

Basile, V., Peroni, S., Tamburini, F., & Vitali, F. (2015). Topical tags vs non-topical tags: Towards a bipartite classification? *Journal of Information Science*, *41*(4), 486–505. https://doi.org/10.1177/0165551515585283

Benbunan-Fich, R., & Koufaris, M. (2013). Public contributions to private-collective systems: the case of social bookmarking. *Internet Research*, *23*(2), 183–203. https://doi.org/10.1108/10662241311313312

Berendt, B., & Hanser, C. (2007, March 26). Tags are not metadata, but just more content-to some people. *ICWSM'07*. International conference on weblogs and social media, Colorado, USA. http://icwsm.org/papers/2--Berendt-Hanser.pdf

Bergman, M. M. (2008). *Advances in mixed methods research: theories and applications*. Sage Publications.

Broughton, V., & Budzak, D. (2010). Adventures in tagging: local history in East London. *Aslib Proceedings*, *62*(4/5), 502–513. https://doi.org/10.1108/00012531011074735

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Bullingham, L., & Vasconcelos, A. C. (2013). 'The presentation of self in the online world': Goffman and the study of online identities. *Journal of Information Science*, *39*(1), 101–112. https://doi.org/10.1177/0165551512470051

Carman, N. (2009). *LibraryThing Tags and Library of Congress Subject Headings: a Comparison of Science Fiction and Fantasy Works* [Master thesis, School of Information Management]. http://researcharchive.vuw.ac.nz/handle/10063/1272?show=full

Case, D. O. (2012). *Looking for information: a survey of research on information seeking, needs, and behavior* (3rd ed.). Emerald.

Chowdhury, G. G. (2010). *Introduction to modern information retrieval*. Facet.

D'heer, E., Evelien. Dheer@ugent. b., Vandersmissen, B., Baptist. vandersmissen@ugent. b., De Neve, W., Wesley. deneve@ugent. b., Verdegem, P., Pieter. verdegem@ugent. b., & Van de Walle, R., Rik. vandewalle@ugent. b. (2017). What are we missing? An empirical exploration in the structural biases of hashtag-based sampling on Twitter. *First Monday*, *22*(2), 1–1. lls.

Engholm, G., Kejs, A. M. T., & Christensen, N. C. S. (2011). *Kræft i tal* [Text]. http://www.cancer.dk/Hjaelp+viden/fakta+om+kraeft/kraeft+i+tal/

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219–245.

Fox, S., & Jones, S. (2009). *The Social Life of Health Information: Americans' pursuit of health takes place within a widening network of both online and offline sources.* Pew Internet & American life project. http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Health_2009.pdf

Fugmann, R. (1993). *Subject analysis and indexing : theoretical foundation and practical advice* (Vol. 1). Indeks Verlag.

Furnas, G. W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C., & Naaman, M. (2006). Why do tagging systems work? *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, 36–39.

Gerolimos, M. (2013). Tagging for Libraries: A Review of the Effectiveness of Tagging Systems for Library Catalogs. *Journal of Library Metadata*, *13*(1), 36–58. https://doi.org/10.1080/19386389.2013.778730

Glenton, C., Nilsen, E., & Carlsen, B. (2006). Lay perceptions of evidence-based information: A qualitative evaluation of a website for back pain sufferers. *BMC Health Services Research*, *6*(1), 34. https://doi.org/10.1186/1472-6963-6-34

Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.

Goffman, E. (1983). The Interaction Order: American Sociological Association, 1982 Presidential Address. *American Sociological Review*, *48*(1), 1–17. https://doi.org/10.2307/2095141

Goffman, E. (2010). *Relations in public: microstudies of the public order*. Transaction Publishers.

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208.

Golub, K., Moon, J., Tudhope, D., Jones, C., Matthews, B., PuzoD, B., & Nielsen, M. L. (2009). EnTag: Enhancing social tagging for discovery. *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 163–172. ACM.

Guy, M., & Tonkin, E. (2006). Folksonomies. *D-Lib Magazine*, *12*(1). https://doi.org/10.1045/january2006-guy

Hajibayova, L., & Jacob, E. K. (2015). Factors Influencing User-Generated Vocabularies: How Basic are Basic Level Terms? *Knowledge Organization*, *42*(2), 102–112.

Haunschild, R., r. haunschild@fkf. mpg. d., Leydesdorff, L., loet@leydesdorff. ne., Bornmann, L., bornmann@gv. mpg. d., Hellsten, I., i. r. hellsten@uva. n., & Marx, W., w. marx@fkf. mpg. d. (2019). Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags. *Journal of Informetrics*, *13*(2), 695–707. lls.

Hemsley, J., jjhemsle@syr. ed., Erickson, I., imericks@syr. ed., Jarrahi, M. H., jarrahi@unc. ed., & Karami, A., karami@mailbox. sc. ed. (2020). Digital nomads, coworking, and other expressions of mobile work on Twitter. *First Monday*, *3/2*, 1–1. lls.

Heymann, P., & Garcia-Molina, H. (2009). *Contrasting Controlled Vocabulary and Tagging: Do Experts Choose the Right Names to Label the Wrong Things?* 1–4. http://ilpubs.stanford.edu:8090/955/

Hjortsæter, E. 1951-. (2005). *Emneordskatalogisering: innholdsanalyse, emnerepresentasjon og lagring: Vol. 2005 nr 2*. Høgskolen i Oslo.

Hunter, J. (2009). Collaborative semantic tagging and annotation systems. In B. Cronin (Ed.), *Annual review of information science and technology* (Vol. 43, pp. 187–239). Published by Knowledge Industry Publications for American Society for Information Science.

Hutchins, W. J. (1975). *Languages of indexing and classification: a linguistic study of structures and functions* (Vol. 3). Peter Peregrinus.

Hutchins, W. J. (1977). On the problem of 'aboutness in document analysis. *Journal of Informatics*, *1*(1), 17–35.

Hyett, N., Kenny, A., & Dickson-Swift, V. (2014). Methodology or method? A critical review of qualitative case study reports. *International Journal of Qualitative Studies on Health and Well-Being*, *9*. https://doi.org/10.3402/qhw.v9.23606

Ingwersen, P., & Järvelin, K. (2005). *The turn: Integration of information seeking and retrieval in context*. Springer.

International Organization for Standardization. (2011). *Information and documentation: Thesauri and interoperability with other vocabularies*. ISO.

Iyer, H., & Bungo, L. (2011). An examination of semantic relationships between professionally assigned metadata and user-generated tags for popular literature in complementary and alternative medicine. *Information Research*, *16*(3), 26–26.

Järvelin, K., & Ingwersen, P. (2010). User-Oriented and Cognitive Models of Information Retrieval. In *Encyclopedia of Library and Information Sciences* (Third edition, pp. 5521–5534).

Jensen, T. W. (2011). *Kognition og konstruktion : to tendenser i humaniora og den offentlige debat*. Samfundslitteratur.

Jen-Yuan Yeh, & Shihn-Yuarn Chen. (2016). Sentence-Level Opinion Analysis for Chinese News Documents Based on Sentiment Information of Social Tags. *Journal of Digital Information Management*, *14*(1), 52–62.

Joorabchi, A., English, M., & Mahdi, A. E. (2015). Automatic mapping of user tags to Wikipedia concepts: The case of a Q&A website – StackOverflow. *Journal of Information Science*, *41*(5), 570–583. https://doi.org/10.1177/0165551515586669

Kalbach, J. (2007). *Designing Web navigation*. O'Reilly.

Keselman, A., Smith, C. A., Divita, G., Kim, H., Browne, A. C., Leroy, G., & Zeng-Treitler, Q. (2008). Consumer health concepts that do not map to the UMLS: Where do they fit? *Journal of the American Medical Informatics Association*, *15*(4), 496–505. https://doi.org/10.1197/jamia.M2599

Kipp, M. E. I. (2005). Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator, and Intermediary Keywords. *Canadian Journal of Information & Library Sciences*, *29*(4), 419–436.

Kipp, M. E. I. (2011). Tagging of Biomedical Articles on CiteULike: A Comparison of User, Author and Professional Indexing. *Knowledge Organization*, *38*(3), 245–261.

Kipp, M. E. I., & Campbell, D. G. (2006). Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. *Proceedings of the American Society for Information Science and Technology*, *43*(1), 1–18. https://doi.org/10.1002/meet.14504301178

Kipp, M. E. I., & Campbell, D. G. (2010). Searching with Tags: Do Tags Help Users Find Things? *Knowledge Organization*, *37*(4), 239–255.

Kvale, S. 1938-2008, & Brinkmann, S. (2009). *Interviews: learning the craft of qualitative research interviewing*. Sage.

LaFrance, M. (1989). The quality of expertise: Implications of expert-novice differences for knowledge acquisition. *ACM SIGART Bulletin*, *108*, 6–14.

Lancaster, F. W. (1977). Vocabulary Control in Information Retrieval Systems. *Advances in Librarianship*, *7*, 1–40.

Lancaster, F. W. (2003a). Abstracts: types and functions. In *Indexing and abstracting in theory and practice* (third ed, pp. 100–112). University of Illinois.

Lancaster, F. W. (2003b). *Indexing and abstracting in theory and practice* (third ed). University of Illinois.

Leydon, G. M., Boulton, M., Moynihan, C., Jones, A., Mossman, J., Boudioni, M., & McPherson, K. (2000). Cancer patients' information needs and information seeking behaviour: In depth interview study. *BMJ*, *320*(7239), 909–913. https://doi.org/10.1136/bmj.320.7239.909

Lin, Y.-L., yllin@mis. nsysu. edu. t., Brusilovsky, P., & He, D. (2016). Finding cultural heritage images through a Dual-Perspective Navigation Framework. *Information Processing & Management*, *52*(5), 820–839. https://doi.org/10.1016/j.ipm.2016.03.003

Matsuyama, R. K., Kuhn, L. A., Molisani, A., & Wilson-Genderson, M. C. (2013). Cancer patients' information needs the first nine months after diagnosis.

Patient Education and Counseling, *90*(1), 96–102. https://doi.org/10.1016/j.pec.2012.09.009

Melenhorst, M., Grootveld, M., & Veenstra, M. (2008). Tag-based information retrieval of educational videos. *EBU Technical Review Q*, *2*.

Morahan-Martin, J. M. (2004). How internet users find, evaluate, and use online health information: A cross-cultural review. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *7*(5), 497–510.

Morrison, J. P. (2008). Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing & Management*, *44*(4), 1562–1579. https://doi.org/10.1016/j.ipm.2007.12.010

Muller, M. J., Ehrlich, K., & Farrell, S. (2006). Social tagging and self-tagging for impression management. *Submitted as an Interactive Poster to CSCW*. http://domino.watson.ibm.com/cambridge/research.nsf/c9ef590d6d00291a85257141004a5c19/27658d7dcf7e8cce852572330070244d/$FILE/TR2006-2.pdf

Munk, T. B., & Mørk, K. (2007). Folksonomy, the power law & the significance of the least effort. *Knowledge Organization*, *34*(1), 16–33.

Nicholas, D., Huntington, P., Williams, P., & Dobrowolski, T. (2004). Re-appraising information seeking behaviour in a digital environment: Bouncers, checkers, returnees and the like. *Journal of Documentation*, *60*(1), 24–43.

Patel, V. L., Arocha, J. F., & Kushniruk, A. W. (2002). Patients' and physicians' understanding of health and biomedical concepts: Relationship to the design of EMR systems. *Journal of Biomedical Informatics*, *35*(1), 8–16. https://doi.org/10.1016/S1532-0464(02)00002-3

Pera, M. S., Lund, W., & Ng, Y.-K. (2009). A sophisticated library search strategy using folksonomies and similarity matching. *Journal of the American Society for Information Science and Technology*, *60*(7), 1392–1406.

Peters, I. (2009). *Folksonomies: Indexing and retrieval in Web 2.0*. De Gruyter.

Peterson, E. (2009). Patron preferences for folksonomy tags: Research findings when both hierarchical subject headings and folksonomy tags are used. *Evidence Based Library & Information Practice*, *4*(1), 53–56.

Poikonen, T., & Vakkari, P. (2009). Lay persons' and professionals' nutrition-related vocabularies and their matching to a general and a specific thesaurus. *Journal of Information Science*, *35*(2), 232–243.

Price, S. L. (2007). Moving toward web-scale: Adapting semantic components for use in large collections. *Proceedings of FDIA 2007: BCS IRSG Symposium: Future Directions in Information Access Editors: Andrew*, 35–41. http://irsg.bcs.org/proceedings/proceedings-fdia2007.pdf

Price, S. L., Nielsen, M. L., Delcambre, L. M. L., & Vedsted, P. (2007). Semantic components enhance retrieval of domain-specific documents. *Proceedings of the Sixteenth ACM Conference on Conference on Information and*

Knowledge Management, 429–438. http://portal.acm.org/citation.cfm?id=1321502

Price, S. L., Nielsen, M. L., Delcambre, L. M. L., Vedsted, P., & Steinhauer, J. (2009). Using semantic components to search for domain-specific documents: An evaluation from the system perspective and the user perspective. *Information Systems*, *34*(8), 724–752. https://doi.org/10.1016/j.is.2009.04.005

Quintarelli, E. (2005). Folksonomies: power to the people. *Incontro ISKO Italia - UniMIB*. Incontro ISKO Italia - UniMIB, Milano. http://www.iskoi.org/doc/folksonomies.htm

Rainie, L. (2007). *28% of online Americans have used the internet to tag content* (Pew Internet & American Life Project). Pew Internet & American life project. http://www.pewinternet.org/~/media//Files/Reports/2007/PIP_Tagging.pdf.pdf

Ranganathan, S. R. (1967). *Prolegomena to library classification*. Asia Publishing House.

Ransom, N., & Rafferty, P. (2011). Facets of user-assigned tags and their effectiveness in image retrieval. *Journal of Documentation*, *67*(6), 1038–1066. https://doi.org/10.1108/00220411111183582

Rasborg, T., & Nielsen, I. G. (2008). *Fokusgruppeundersøgelse af cancer.dk: Udarbejdet for Kræftens Bekæmpelse*. Userneeds.

Rasborg, T., & Toft, I. (2008). *Strukturel kortsortering af cancer.dk: Udarbejdet for Kræftens Bekæmpelse*. Userneeds.

Rolla, P. J. (2009). User Tags versus Subject Headings: Can User-Supplied Data Improve Subject Access to Library Collections? *Library Resources & Technical Services*, *53*(3), 174–184. https://doi.org/Article

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H. R., Dobrowolski, T., & Tenopir, C. (2008). The Google generation: The information behaviour of the researcher of the future. *Perspectives*, *60*(4), 290–310.

Rozmovits, L., & Ziebland, S. (2004). What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. *Patient Education and Counseling*, *53*(1), 57–64. https://doi.org/10.1016/S0738-3991(03)00116-2

Ruocco, M., ruocco@idi. ntnu. n., & Ramampiaro, H., heri@idi. ntnu. n. (2015). Geo-temporal distribution of tag terms for event-related image retrieval. *Information Processing & Management*, *51*(1), 92–110. https://doi.org/10.1016/j.ipm.2014.09.001

Saracevic, T. (1996). Relevance reconsidered. *Proceedings of the Second Conference on Conceptions of Library and Information Science (CoLIS 2)*, 201–218.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 484.

Shultz, C., hedgimom@aol. co. (2011). Cadavers or Corpses: Comparing User-Created Metadata. *PNLA Quarterly*, *75*(4), 82–111.

Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a Tool for Health Research: A Systematic Review. *Am J Public Health*, *107*(1), e1–e8. https://doi.org/10.2105/AJPH.2016.303512

Smith, C. A., Stavri, P. Z., & Chapman, W. W. (2002). In their own words?: A terminological analysis of e-mail to a cancer information service. In I. S. Kohane (Ed.), *Bio*medical informatics: One discipline: Proceedings* (pp. 697–701). Hanley & Belfus.

Soohyung Joo, & Kipp, M. E. I. (2015). Exploring the structure of library and information science Web space based on multivariate analysis of social tags. *Information Research*, *20*(4), 128–148.

Spiteri, L. F. (2007). The structure and form of folksonomy tags: the road to the public library catalog. *Information Technology and Libraries*, *26*(3), 13–25.

Spiteri, L. F. (2009). The Impact of Social Cataloging Sites on the Construction of Bibliographic Records in the Public Library Catalog. *Cataloging & Classification Quarterly*, *47*(1), 52–73. https://doi.org/10.1080/01639370802451991

Stake, R. E. (1995). *The art of case study research*. Sage.

Steinhauer, J., Delcambre, L. M. L., Lykke, M., & Ådland, M. K. (2013). Do User (Browse and Click) Sessions Relate to Their Questions in a Domain-Specific Collection? In T. Aalberg, C. Papatheodorou, M. Dobreva, G. Tsakonas, & C. J. Farrugia (Eds.), *Research and Advanced Technology for Digital*

Libraries (pp. 96–107). Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-40501-3_10

Steinhauer, J., Delcambre, L. M. L., Maier, D., Lykke, M., & Tran, V. H. (2011). Tags in domain-specific sites: new information? *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, 109–112. https://doi.org/10.1145/1998076.1998096

Svenonius, E. (2000). *The intellectual foundation of information organization*. MIT Press.

Taylor, A. (2012). User relevance criteria choices and the information search process. *Information Processing & Management*, *48*(1), 136–153.

Thomas, M., Caudle, D. M., & Schmitz, C. (2010). Trashy tags: problematic tags in Library Thing. *New Library World*, *111*(5/6), 223–235. https://doi.org/10.1108/03074801011044098

Thomas, M., Caudle, D. M., & Schmitz, C. M. (2009). To tag or not to tag? *Library Hi Tech*, *27*(3), 411–434.

Toms, E. G. (2002). Information interaction: Providing a framework for information architecture. *Journal of the American Society for Information Science and Technology*, *53*(10), 855–862. https://doi.org/10.1002/asi.10094

Uden-Kraan, C. F. van, Drossaert, C. H., Taal, E., Seydel, E. R., & Laar, M. A. van de. (2008). Self-Reported Differences in Empowerment Between Lurkers and Posters in Online Patient Support Groups. *Journal of Medical Internet Research*, *10*(2), e18. https://doi.org/10.2196/jmir.992

Userneeds. (2008a). *Webstatus: Brugertilfredshedsundersøgelse*. Userneeds.

Userneeds. (2008b). *Inholdsundersøgelse: Cancer.dk: Udarbejdet for Kræftens Bekæmpelse*. Userneeds.

Userneeds. (2008c). *Terminologisk kortsortering af cancer.dk: Udarbejdet for Kræftens Bekæmpelse*. Userneeds.

Vander Wal, T. (2005). *Explaining and showing broad and narrow folksonomies: Personal InfoCloud*. http://personalinfocloud.com/2005/02/explaining_and_.html

Vickery, B. C. (1966). *Faceted classification schemes*.

Voorbij, H. (2012). The value of LibraryThing tags for academic libraries. *Online Information Review*, *36*(2), 196–217. https://doi.org/10.1108/14684521211229039

Wetterstrom, M. (2008). The Complementarity of Tags and LCSH — A Tagging Experiment and Investigation into Added Value in a New Zealand Library Context. *New Zealand Library & Information Management Journal*, *50*(4), 296–310. https://doi.org/Article

Williams, P., & Rowlands, I. (2007). *Information behaviour of the researcher of the future: The literature on young people and their information behaviour: Work package II*. British Library. http://www.ucl.ac.uk/infostudies/research/ciber/downloads/GG%20Work%20Package%20II.pdf

Yi, K., & Chan, L. M. (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of Documentation*, *65*(6), 872–900. https://doi.org/10.1108/00220410910998906

Yin, R. K. (2003). *Case study research: design and methods* (3rd ed., Vol. 5). Sage.

Yin, R. K. . (2014). *Case study research: Design and methods* (5th ed.). SAGE.

Zeng, Kogan, S., Ash, N., & Greenes, R. A. (2001). Patient and clinician vocabulary: How different are they? *Studies in Health Technology and Informatics*, *84*(Pt 1), 399–403.

Zeng, Kogan, S., Ash, N., Greenes, R. A., & Boxwala, A. A. (2002). Characteristics of consumer terminology for health information Retrieval. *Methods Inf Med*, *41*, 289–98.

Zhitomirsky-Geffet, M., & Daya, Y. (2015). Mining query subtopics from social tags. *Information Research*, *20*(2), 1–15.

10.5 E-MAILS

E-mail from Tor Øyan We 30.12.2011

>-----Original Message-----
>From: Tor Øyan <tor@cancer.dk>
>Sent: Wednesday, November 30, 2011 12:40 PM
>To: 'lmd@cs.pdx.edu' <lmd@cs.pdx.edu>; 'Jeremy Steinhauer'
><jeremy.steinhauer@gmail.com>; 'Jeremy Steinhauer'
><jsteinha@cs.pdx.edu>; 'Marianne Lykke' <mlykke@hum.aau.dk>; Marit
>Kristine Ådland <Marit-Kristine.Adland@hioa.no>
>Cc: Mette Tandrup Hansen <mth@cancer.dk>; Pernille Vigild Laursen
><pevl@cancer.dk>; Hanne Sandvang <has@cancer.dk>
>Subject: Tagging live!
>
>Hi FIRE team
>

>Finally we're live - get the champagne!
>
>The last wait was for getting the second frontend server updated and for our
>hosting provider to make a fresh dump of the tagging base, so that you could
>get a snapshot of the tagging done by us at the Cancer Society before going
>live. The database dump should (soon) be available for Jeremy on the FTP
>servers. A funny detail: It looks as if the first external tag was; selleribøf =
>celerysteak
>
>Right now I'm busy setting up the new search pages on all our subsites.
>
>Hooray
>Tor :-)

E-mail from Tor Øyan Fr 09.12.2011 13:53

>-----Original Message-----
>From: Tor Øyan <tor@cancer.dk>
>Sent: Friday, December 9, 2011 1:53 PM
>To: Marit Kristine Ådland <Marit-Kristine.Adland@hioa.no>
>Cc: 'Marianne Lykke' <mlykke@hum.aau.dk>
>Subject: SV: Undersøgelse av nøgleord
>
>Hej Marit Kristine
>
>Godt at høre, at du har fået noget ud af dit besøg. Roserne er sendt videre til
>Gurli.
>
>Jeg har også gjort et par observationer - alene på grundlag af de tags, der
>dukker op:
>
>Den ene er, at vi får ret mange tags, som blot er helt tomme. Dette sker, hvis
>man klikker 'Tilføj' og tilføjer, uden at skrive noget i 'Tilføj' feltet. Min
>hypotese er, at nogle brugere tolker 'Tilføj' knappen, som en knap, man kan
>bruge til at tilføje siden til sine favoritter. Og så kommer de altså til at tilføje et
>tomt felt som en tag i stedet. (Jeg sletter disse tomme tags løbende)
>
>Den anden er, at vi får en del tags, som ikke på nogen måde hænger sammen
>med indholdet på den side, der er tagget. Min hypotese er her, at nogle
>brugere tolker tilføj feltet nederst på siden som et søgefelt. De tror, at de
>fyrer en søgning af, men det, de faktisk gør, er at tagge siden med et tag, der
>er fuldstændigt 'random' i forhold til sidens indhold. Og brugeren har jo heller
>ikke selv nogen mulighed for at slette tagget efterfølgende (Jeg har ikke
>systematisk været inde og slette denne type tags. Det er også meget
>besværligt fordi jeg så skal ind og se, hvilke sider, der er tagget. Men det ser

>ud til, at jeg bliver nødt til at rydde i det, for det forekommer ret tit. Vil dog
>være meget tilbageholden med at slette, med mindre det er fuldstædig
>oplagt, at der slet ikke er nogen sammenhæng overhovedet.)
>
>God weekend!
>Mvh Tor

APPENDICES

10.6 APPENDIX I – TAGGER STUDY

PRÆ-TEST SPØRGESKEMA (PRE-TEST QUESTIONNAIRE)

1. Alder:	
21-30	□
31-40	□
41-50	□
51-60	□
61-70	□
71-80	□
81 eller ældre	□
2. Køn:	
Kvinde	□
Mand	□
3. Hvilket årstal fik du første gang konstateret kræft (skriv årstallet)	

4. Modersmål (skriv f.eks. dansk, tysk)	
5. Hvor mange år har du brugt Internettet (f.eks. søgemaskiner som Google)?	
Mindre end 1 år	□
Mellem 1 og 3	□
Mellem 4 og 6	□
7 eller flere	□
6. Hvor lang tid har du anvendt Cancer.dk?	
Aldrig	□
Mindre end 1 måned	□
Mellem 1 og 6 måneder	□
Mellem 7 og 12 måneder	□
1 år eller mere	□
7. Hvor ofte bruger du Cancer.dk?	
Aldrig	□
En eller to gange om ugen	□
En eller to gange om måneden	□

En eller to gange om året	□
En eller flere gange om dagen	□
8. Hvordan bruger du Cancer.dk? Du er velkommen til at afkrydse flere muligheder	
Søger efter information	□
Læser Nyheder	□
Læser Blog om kræft	□
Bruger cancerforum.dk	□
Bruger Kræftens Bekæmpelses Netbutik	□
Indrapporterer om Fejl I behandlingen (Patientsikkerhed)	□
Andre. Skriv gerne andre måder, som du anvender Cancer.dk	□
9. Kjente du til fenomenet tagging på nettet før du ble orientert om denne undersøkelsen?	
Ja	□
Nei	□

10. Dersom ja i spørsmål 9: Har du selv *brukt* tagger (klikket på dem eller søkt/browset)?		
Ja, en eller noen få ganger	□	
Ja, flere ganger	□	
Nei	□	
11. Dersom ja i sprøsmål 9: Har du selv *laget* tagger?		
Ja, en eller noen få ganger	□	
Ja, flere ganger	□	
Nei	□	
12. Bruker du med sosiale medier? Du er velkommen til at afkrydse flere muligheder og legge til flere alternativer		
	Leser	Skriver/bidrar selv
Facebook	□	□
Cancerfourm	□	□
LinkedIn	□	□
Youtube	□	□
MySpace	□	□
Wiki	□	□
Blogs	□	□
Annet:	□	□

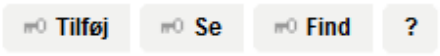
OPGAVER (TASKS)

DELTAGER: ___ DECEMBER 2011

Tasks, version 1

Opgaver

Brug websiden Cancer.dk til at gennemføre følgende opgaver. For nogle af opgaverne kan der være flere websider på Cancer.dk, der giver svaret. Det er tilstrækkeligt, at du vælger en webside for hver opgave.



Brug linket 'Tilføj' til at tildele nøgleord, linket 'Se' til kigge på nøgleord, og linket 'Find' til at søge nøgleord.

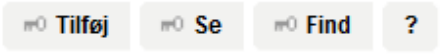
1. Find listen af nyheder eller nyhedsartikler på cancer.dk. Klik på en nyhedsartikel, som du finder spændende. Tilføj et eller flere nøgleord til artiklen.
2. Find en artikel om, hvordan Kræftens Bekæmpelse anvender de midler, som de modtager fra sponsorer og donorer. Tilføj et eller flere nøgleord til den.
3. Find en artikel om vaccination mod kræft. Tilføj et eller flere nøgleord til den.
4. Genfind en af de artikler, som du har tagget. Brug søgeknappen på cancer.dk. Hvis du har lyst, må du gerne tilføje et eller flere ekstra nøgleord til websiden.

DELTAGER: ___ DECEMBER 2011

Tasks, version 2

Opgaver

Brug websiden Cancer.dk til at gennemføre følgende opgaver. For nogle af opgaverne kan der være flere websider på Cancer.dk, der giver svaret. Det er tilstrækkeligt, at du vælger en webside for hver opgave.



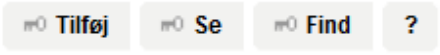
Brug linket 'Tilføj' til at tildele nøgleord, linket 'Se' til kigge på nøgleord, og linket 'Find' til at søge nøgleord.

1. Find listen af nyheder eller nyhedsartikler på cancer.dk. Klik på en nyhedsartikel, som du finder spændende. Tilføj et eller flere nøgleord til artiklen.
2. Find information om soja og kræft. Tilføj et eller flere nøgleord til artiklen.
3. Find en artikel om psykologiske reaktioner til det at modtage en kræftdiagnose. Tilføj et eller flere nøgleord til artiklen.
4. Genfind en af de artikler, som du har tagget. Brug søgeknappen på cancer.dk. Hvis du har lyst, må du gerne tilføje et eller flere ekstra nøgleord til websiden.

Tasks, version 3

Opgaver

Brug websiden Cancer.dk til at gennemføre følgende opgaver. For nogle af opgaverne kan der være flere websider på Cancer.dk, der giver svaret. Det er tilstrækkeligt, at du vælger en webside for hver opgave.



Brug linket 'Tilføj' til at tildele nøgleord, linket 'Se' til at kigge på nøgleord, og linket 'Find' til at søge nøgleord.

1. Find listen af nyheder eller nyhedsartikler på cancer.dk. Klik på en nyhedsartikel, som du finder spændende. Tilføj et eller flere nøgleord til artiklen.
2. Find en artikel om indflydelse af arbejdsmiljø på risikoen for at få kræft. Tilføj et eller flere nøgleord til artiklen.
3. Cancer.dk indeholder også madopskrifter. Find en opskrift for noget, som du synes ser fristende ud. Tilføj et eller flere nøgleord til artiklen.
4. Genfind en af de artikler, som du har tagget. Brug søgeknappen på cancer.dk. Hvis du har lyst, må du gerne tilføje et eller flere ekstra nøgleord til websiden.

POST-TEST INTERVIEW

[Questions I want to ask are numbered 1, 2, ... Core questions are in **bold**, the other ones are for starting the interview and for follow-up if needed.

Aspects I want test subjects to comment on, or options I will suggest for test subjects if they struggle to find words, are listed with a, b, c...]

Introduction

We will talk a little about your experiences when completing the tasks using cancer.dk. As you know, your actions on the screen have been captured in a log, but we are now interested in how and why you did what you did.

We will not ask whether you completed the tasks as you were instructed or not, we presume that you did what you did for a good reason.

We will focus on tagging, but also touch on the other things you did in addition to using tags.

If it is ok with you, we will record the interview.

Interview guide

Purpose: How useful are the tagging features?

Interview questions:

1. Tell me a little about how easy it was to complete the tasks?
2. How easy was it to use the tagging features?
   Aspects:

   a. Find the feature.
   b. Understand the feature.
   c. Apply a tag.
   d. See tags.
   e. Search, including searching tags.
   f. Browse tags.

Purpose: How do people tag?

Interview questions:

3. **How easy was it to choose the words you used when applying tags?**
4. **Did you use any sources when formulating your tags? Was this important / useful?**
   Some source options:

   a. Words from title or rest of the article?
   b. Words in tags that were already applied?
   c. Words describing in which situation the article may be needed?
   d. Words describing the subject of the article?
   e. Other sources?
   f. Words *not* present in article or already applied as tags?

Purpose: How do people use tags? How do they judge the quality of tags?

Interview questions:

5. **In the last task you were asked to use the search field to re-find one of the articles you tagged. How did you do that? Which search terms did you use?**
   Some options:
   a. A word you remember from the title?
   b. A word you remember from the text?
   c. A word you applied as a tag?
   d. A word somebody else had applied as tag?
   e. Other words?
6. **What would you describe as good properties for a tag, based on your experience with tagging?**
   Some options:
   a. Precise content description?
   b. Description of who would use the article?
   c. Easily understandable words?
   d. Words from title or text?
7. Do you think tags are useful?
   a. Your own tags or all tags?
   b. For whom?
8. Do you think you will use tags and/or apply tags at cancer.dk later? Why/Why not?

9. (If yes for question 8:) What do you think that you will use tagging/tags for?
   Some options:
   a. To make it easier for *you* to re-find articles
   b. To make it easier for *others* to find articles
   c. To help improve cancer.dk
   d. To add search words to articles
   e. To add information to the article, to add information about the topic to other users
   f. To communicate new ideas, information, comments to other patients and their relatives
   g. To refer other users to other related topics

10.7 APPENDIX II– EDITOR STYDY

INTERVIEW GUIDE

Introduction:

- **Brief conversation on participation and audio recording.**
- **I am interested in your honest opinion about tags and how the feature is implemented.**

Starting points for conversation (Ask for examples everywhere)

General questions:

1. Tell me a little about your daily work: Writing and editing articles, other things (that relate to tags).

2. Do you use Cancer.dk professionally? Do you ever use Cancer.dk as an end-user?
    a. Do you use it from home, from work or both? If from home, what do you do?

3. What was your general opinion on tagging in the beginning? What is your opinion now?

Questions about tags and how they are used:

4. How would you describe a *good* or *ideal* tag?
    a. Would *good* tags be different on different parts of Cancer.dk?
    b. Depending on the subject's answer, ask them about types of tags:
        i. Subject descriptor
        ii. Tags that explain
        iii. Tags that value or judge an article
        iv. Tags that express a desire for more information
    c. What is your opinion of these tag types?

5. How do you believe *good* tags can be used?
    a. Depending on the answer, suggest types of *tag usage* like:
        i. Finding information – search and/or browsing
        ii. Finding out how other people used the article
    b. What is your opinion on these tag usage functions?
    c. Will bad tags influence this/these types of usage(s)? <maybe out? and move to 6.d.>

6. How would you describe a typical tag that is *bad*?
    a. Do *bad* tags fall into certain types?

   b. Would bad tags be different on different parts of Cancer.dk?
   c. What damage do you believe bad tags can do?
   d. Will bad tags influence the types of usage(s) described above (in 5.a.)?

7. Are there tags that are neither good nor bad – perhaps tags that help some people some of the time but may not be helpful to other people?

Questions about tagging:

8. Do you ever use tags (professionally or as an end-user)? How?
   a. Look at tags
   b. Search for tags
   c. Browse with tags (by clicking on them)
   d. (also include parts of tagging feature and types of tags)

9. Do you have an opinion about how people formulate their tags?
   a. Do you think that they use any sources of information like words that appear in the article itself?

10. Do you apply tags? Do you use any sources? If so, which sources? Google Analytics? User questions or comments? Others?
    a. Do you think your choice of tags or your sources differ from what end users do?

11. Do end-user tags differ from terms that you (as a staff member) would apply to documents?
    a. If so, how?

Questions about tagging policy:

12. Do you have an opinion about who should be tagging at Cancer.dk? Users or staff members or both?

13. What do you think of today's policy of tag editing?
    a. Do you have access to edit and delete tags? Have you ever deleted or edited tags?
    b. Should more/fewer/other tags be deleted or edited by the staff?
    c. Should other tags be deleted or kept in the system? Do you agree with todays policy for editing tags?

Questions about the way tags are implemented (over time) in cancer.dk:

14. What is your opinion of the tagging features on the cancer.dk site?
    a. Interface and visibility of feature to apply tags
    b. Interface and visibility of tags
    c. Interface and visibility of tag cloud
    d. Interface and visibility of tag browsing page

15. What is your opinion of the latest change in the tagging feature, where the bottom tagging field is removed?

16. Have other parts of Cancer.dk changed or been influenced as a consequence of tags or the tagging feature?
    a. How and why?
    b. Are these changes beneficial?

17. What would a perfect tagging feature be like?

ISSN (online): 2246-123X
ISBN (online): 978-87-7210-505-5

AALBORG UNIVERSITY PRESS