Abstract

This paper explores tags and tagging behaviour on health information websites using an empirical, user-oriented, exploratory case study. Taggers and editors were interviewed about tags and tagging, while taggers solved tasks that included applying tags to a website. This qualitative data was combined with quantitative data; a transaction log analysis that included all tags applied to the site during the logging period. Results showed that taggers and editors disagreed both on the purpose of tags and what characterizes a good tag. Taggers applied tags to describe article content, explain articles, request information, and value article content. Some of these show that tags are not only not only topical descriptions, but communicative by intent. This result can potentially inform the design of tagging features.

Introduction

The purpose of this paper is to explore tags and tagging behaviour on a health information website: how “aboutness” is expressed, what facets are used, how the meaning or topical content of tags is related to document content, what vocabulary is used by taggers, and what opinions users have of tags. We chose Cancer.dk case for the study. This Danish information website is intended for cancer patients and their relatives. Knowledge of tagging behaviour will guide the editors of information websites when setting up tagging features. We find it interesting to study tags in a system where metadata is not as visible to users, as compared to other systems with tags, such as LibraryThing or Delicious. Specifically, we seek to answer the following research questions:

1. What characterizes tags on Cancer.dk?
2. How do the tags on Cancer.dk compare to other tagging systems?
3. How do users and editors perceive tags, and how do they intend to use them?
4. Do the views of users and editors correspond both with each other and the nature of tags?

Methods and study designs

This is a user-oriented (Järvelin and Ingwersen 2010) empirical, exploratory case study (Bryman 2012) in which we studied user interaction with tags, tagging, and information behaviour in general on a specific website, Cancer.dk. In this user-oriented study, the user is seen as a part of the system (Järvelin and Ingwersen 2010). With tags, the user also provides metadata, which makes user tagging behaviour crucial in understanding the entire system.

A combination of qualitative and quantitative methods provides a holistic view, which creates a complex picture of tagging behaviour on Cancer.dk (Bergman 2008).
The quantitative data (gained from log files), provides insight into what is going on and makes it possible to explore whether an identifiable phenomenon is frequent or not. We counted tags of various types and compared the number of tags in different categories. The quantitative data provided an overview and identified some patterns. No tags were excluded from the analysis. The qualitative data provided explanations for the behaviour.

Three studies were conducted:

1. **Tagger study**: The tagger study was conducted when the tagging feature was newly launched. Eight cancer patient participants, all aged from 39 to 73, solved tasks using the tagging feature. Five of them applied tags for the first time. Participants were tasked with finding information on Cancer.dk, then applying one or more tags to an article with the requested information. Example: “Find an article about psychological reactions to a cancer diagnosis. Apply one or more tags to it.” Each participant solved four tasks. Two of the tasks differed so that the participants solved three different set tasks. After the tasks, the participants filled out questionnaires and were interviewed. The interviews were semi-structured with the goal of finding out about the patients’ understanding, motivations, and opinions of tags and tagging, (including their purposes for applying tags). In addition to questionnaires and transcribed interviews, we identified the actual tags the participants applied in the transaction log (see Tag study below).

2. **Editor study**: We interviewed three Cancer.dk editors about their experience with tags and tagging. The goal was to find out about their opinions, then compare this to user opinion. The editors were interviewed in December 2012. The tagging feature had been on Cancer.dk for roughly one year, which ensured editors had experience with it at the time of being interviewed. The tagging feature was maintained by its editors and closely connected to Cancer.dk. Editors wrote articles and structured the site in collaboration with medical and social experts. Methodologically, these interviews were carried out the same way as the interviews with the taggers, except by phone; this allowed for limited possibility in observing non-verbal signals. These interviews were a follow-up to a longer project that the participants had been taking part in for over a year. The editors included criticism of both the tags and the tagging feature. We see this as an indication that the telephone interviews gave sufficient information about the editors’ opinions.

3. **Tag study**: A transaction log from the tagging feature gave the broadest data material. The first tag from an external IP was applied on November 30, 2011. 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 called ‘irrelevant tags’. 
They believed that the field for applying tags was used as a search field by mistake. Thus, they changed the tagging feature in September 2012. The main difference, between before and after the change, was the visibility of the field for applying tags. Previously, the field for applying tags was open by default at the top of every article. Subsequently, it was hidden, and users needed to click one of the buttons in the tagging feature to apply a tag. In both cases, the buttons for tagging were visible on top of all articles both before and after the feature change.

**Results – 1 Tagger study**

The participants applied between 5 and 27 tags each, for an average of 12.25 tags. There were 86 unique tags, 98 tags in total. If we look at the unique combinations of user, tag, and URL, there were 92 such combinations.

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

Our analysis indicated 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 general topic of the article; it was sufficient 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.

Six of eight participants mainly focused on tags as topical descriptions; most of their tags corresponded with this view. Examples of their subject description tags were *soja* (soy) and *antihormon* (anti-hormone), both of which applied to articles about these topics.

Other purposes found were: tags to explain the content, tags to evaluate articles, and tags to express requests for additional information (see below). 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.

*Explicative tags* explained an aspect of a word or a specific word within the content of an article. Two of the participants took this a step further and discussed how to apply tags solely to explain difficult wording or other aspects of articles. It was not easy to see this motive in the tags, as explanatory tags look like topical tags. But the willingness
to apply tags that were synonyms to article words could be seen as an indication of explicative tagging. Despite the similarities between topical tags and explicative tags, the motivation was different. Topical tags described what was already in the article, while explicative tags sought to add an explanation and thus additional content to the site (Berendt and Hanser 2007; Ådland and Lykke 2012). None of the participants described practical challenges, like how to know what part of an article a certain tag explains. This would only be a problem if the article concerned similar or opposite concepts.

Opinion tags expressed an opinion on the content of articles. Two participants (4, 8) used tags to value the article content. The tags were: fremragende (excellent), vigtigt (important) and god nyhed (good news). This type of tag is well known from other systems, like Delicious (Golder and Huberman 2006).

Tags applied to make requests did not necessarily relate to the content of the article. Instead, this tag was applied to request information. One participant (3) stated that she had applied such tags. The purpose of this tag was opposite to topical description. 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” (6). She did not want these kinds of tags in the system.

Regarding purposes for using tags, participants focused on topical tags, or at least tags that related to the same content as the article they were applied to. These tags were useful for browsing and searching. For some, good topical tags meant good credibility for the site. Thus, it was a challenge that a number of tags were topically misleading.

Both tags requesting more information and explicative tags could be seen as supplementary tags (Berendt and Hanser 2007; Ådland and Lykke 2012). They added something to the article that was not there before. If these tags were easier to distinguish from other tags, they could aid the editors in improving the site and be less likely to mislead users.

Results – 2 Editor study

Practical use of Cancer.dk differs between editors. 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 looked for something, because she knew the site so well that she always knew what to search for. They had opposite behaviours but similar explanations; both made sense.

Editors E1 and E2 did not use the tags themselves. Again, they said this was because they knew the site so well and thus did not need tags. This was a sufficient 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 requested more information. These tags could inspire them to make changes on articles
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 words that were already in the text, while others did not. This seemed to be the same for the taggers; some taggers applied tags they had already found in the article text, while others saw this as useless and did not.

Editor opinions about the tagging feature changed during the project. At the project’s start, editor E1 was open and positive, and hoped that the tags would improve the search feature and users’ general ability to find information. Within a year’s experience, she felt that neither the feature nor the tags had fulfilled her expectations. Many tags did not describe the content of the articles. Editor E2 was more sceptical from the beginning. After 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 launched. She liked the idea that users could index content, but from the tags she concluded that the users did not understand the tagging feature. The different purposes users had when they applied tags was a challenge for the editors, as users preferred tags that described the topical content of articles.

Some synonyms may be impossible for the editors 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. So, if one is used in the text, and the other one is applied as a tag, both variations are covered. A mix of English and Danish is the same. Breast cancer or brystkæft may be used in the articles, but breastkæft or brystcancer are mixed compounds and less ideal. Yet, as some users use them and search for them, if these kinds of compounds are applied as tags, they may serve as lead-in terms; the text will not be disturbed by compound words that are incorrect because of the language mix.

A good tag gives a summary, or ‘essence’ of the text, according to editor E2, and it should be given from the user’s viewpoint, as an opposing view to editors and health professionals. However, the main point is the content description, so that tags can be used for search and retrieval. Her explanation was broader than just saying that tags are synonyms to article text words, but her examples confirmed that she included synonym tags in her summary-description of good tags. For the editors, a bad tag is one that does not describe the article. This means that a good, clear tag that expresses an opinion about the article or request for information is still a bad tag.

The editors expressed different opinions about how damaging bad tags are. Editor E1 saw them as noise, which can be a challenge in searching and browsing (though 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. This diversity was also observed among the taggers. Some found that bad tags gave bad credibility, while others did not see bad tags as a threat at all.
The different purposes users had when they applied tags was a challenge for the editors. They knew that it was difficult to distinguish between various purposes when looking at the tags only. It was hard to say whether a certain tag was a description of the article, a supplement, a request for more information, etc. This made it hard to use the tags for searching and browsing, as it was difficult to know what lay behind a tag: whether it meant that a given article was about the topic described in the tag, or that an article lacked the topic indicated in the tag.

From the interviews, it is our impression that, in some ways, the editors 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 provide the lead-in terms that users and editors need. Such a solution is, however, contradicted by the editors’ view that tags are essentially the user’s voice in the system. A controlled vocabulary can never replace this, which the editors were also clear about.

Results – 3 Tag study

After 13 months, there were 25,253 tags in the log. This number includes all tags, empty tags, repeated tags, and errors that the user seemed to correct by adding another tag, i.e. every time a user hit the *tilføj*-(submit)-button in the feature for applying tags. There were 8.4 tags per URL.

| Table 1: Aboutness: Tag categories according to the relationship between tag and article content |
|----------------------------------|----------------|----------------|
| **Aboutness categories**         | **Internal tags** | **External tags** |
|                                 | **Number** | **%** | **Number** | **%** |
| General topic                   | 75 | 3.25 | 235 | 1.02 |
| Exact topic                     | 263 | 11.40 | 592 | 2.58 |
| Exact aspect of topic           | 231 | 10.01 | 391 | 1.70 |
| Topic mentioned                 | 957 | 41.48 | 2,086 | 9.09 |
| Indirectly related              | 15 | 0.65 | 195 | 0.85 |
| Too general                     | 40 | 1.73 | 109 | 0.48 |
| Too specific                    | 90 | 3.90 | 2,507 | 10.93 |
| No relation                     | 210 | 9.10 | 9,271 | 40.40 |
| Empty tags                      | 19 | 0.82 | 3,241 | 14.12 |
| Meaningless tags                | 10 | 0.43 | 2,663 | 11.61 |
| Article was deleted             | 397 | 17.21 | 1,656 | 7.22 |
| **Sum**                         | 2,307 | 100 | 22,946 | 100 |
| Summarized aboutness categories |                       |              |
| Related                         | 1,541 | 66.79 | 3,499 | 15.24 |
| Not related                     | 340 | 14.73 | 11,887 | 51.80 |
| Empty/meaningless tags          | 426 | 18.46 | 7,560 | 32.94 |
| **Sum**                         | 2,307 | 100 | 22,946 | 100 |
We analysed all tags that had been applied during a year, dividing them into the following categories:

1. Internal and external tags – an indication of who applied the tag.
2. Lay or professional – does the tag belong to a lay or professional vocabulary?
3. Aboutness – the relationship between the aboutness of the tag and the aboutness of the article.
4. Topical facets – what is the tag 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 group of external taggers was more stable. Their number of tags varied less during the logging period.

*Visibility of tagging feature:* When the tagging feature changed in September 2012, the visibility of the feature was the only change. This resulted in a dramatic decrease in the number of tags. There were variations in the number of tags during the entire logging period, but none as dramatic as this. What this demonstrates is that the visibility of the tagging feature strongly influences the number of tags.

*Aboutness:* 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 an article, (see Table 1). Due to the nature of this material, it is not clear from the logs whether the users were aware of these difficulties. Some tags are good subject descriptions, while many others are not. Some express the needs and experiences of both users and patients, whereas others requested answers. Many more seemed to be misunderstandings and individual thoughts. Mixed together, the tags as a whole were difficult to use and difficult to judge. One possible explanation for the high number of tags that did not relate to the content of the articles could be that users mistook the tagging field for a search field. If so, this is an example of how efforts have to be taken to inform users about how to properly use a website.

*Facets:* The tag content was influenced by the content of Cancer.dk. Many tags repeated words from articles. The tag topics more or less covered the topics covered by Cancer.dk as a whole. Tags from internal taggers covered categories that were more diverse and described article content from various angles. Their tags showed a more even distribution of tag facets compared to those of external taggers. These results conflicted with the expectation that external users can add new viewpoints to the systems. External taggers, in fact, applied more tags that were unrelated to the content of articles.

The internal taggers were predominant when it came to tags about thoughts, feelings, and psychological symptoms. It seems that in this sense, internal taggers were better at choosing tags independently of Cancer.dk than the external users were. This conflicts
with the expectation that end users can add new viewpoints, while those within the system cannot.

Tagging is communication with the system. In an extended, narrow folksonomy like this one (Peters 2009; Vander Wahl 2005), there is no bookmarking feature and thus the users cannot use the tags solely for their own purposes. They cannot use tags to communicate with themselves. The communication is targeted towards the system, the editors, and the organization behind the system.

Conclusions

We found 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 and non-topical tags, which were intended as means of communicating 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.

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

Previous studies frequently extract tags for indexing or retrieval purposes. Our 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. If we look 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.

References


