

Towards an “Intelligent” Tagging Tool for Blogs

Juraj Frank^{1,2}, Renate Motschnig¹, and Martin Homola^{2,3}

¹ University of Vienna, Research Lab Educational Technologies, Austria

² Comenius University, Faculty of Mathematics, Physics and Informatics,
Bratislava, Slovakia

³ Fondazione Bruno Kessler, Trento, Italy

Abstract. Tagging allows people to effectively organize web resources such as images, bookmarks or blog articles. Things are found easier by browsing tag clouds relying on the tags that have been assigned before. The success is by large determined by the quality and relevance of tags assigned to content – and so it is dependent on people who do the tagging. We investigate mental processes that underlie tagging. In order to improve quality of tagging, we provide guidelines for users of tagging systems and in addition we suggest features that an “intelligent” tagging tool should bear in order to facilitate the tagging process.

Keywords: tagging, categorization, blog, usability.

1 Introduction

Tag is a label associated with something for the purpose of identification. Tagging is a process of assigning such labels or keywords to objects for sake of future identification. It is possible to tag anything what makes sense, however, these days tagging is a popular categorization method for photos, bookmarks or blogs articles. But why is tagging so popular? During the history, first major successful catalogs of web content used simple hierarchical categorization schemes, where one object could usually be a member of one category only. The catalog was exclusively edited by its editors. Tagging is different in two ways. First, tag is much more a label than a folder: apart from classical categories, many labels are possibly associated with each object. Second, Everyone is allowed to tag things. These two little tweaks made tagging more accessible and attractive to everyday people. Tagging is useful to them: it enables to find things by fishing them out of tag clouds and it allows to effectively organize things with little effort. Thus, tagging has become a novel decentralized way of organizing, sharing, and structuring information in the knowledge society [1].

However, several practical problems are associated with tagging. Different people usually use different tags to describe the same topic or type of object (e.g., where one would assign tag *fruit*, someone else assigns the tag *food*). Also, people sometimes forget to assign some particular tag that is of some importance (e.g. *fruit* is missed out for an article about oranges, apples and pears). In order to improve quality of tagging we try to answer the following three research questions:

- What are the basic mental processes that underline tagging?
- How should people be educated and guided during the process of tagging?
- How should tagging systems be improved in order to support the users during the process of tagging?

Literature study shows that both prototype theory [2] and the concept of the basic level categorization [3] are relevant approaches to explain tagging. We have conducted interviews with blog users. Based on this research, we have put together guidelines for tagging and we have identified features that intelligent tagging tools should have in order to improve the quality of tagging.

2 Small Empirical Study

Interviews were conducted with users of the `blog.matfyz.sk` portal.¹ First, a key user was interviewed in a face-to-face session to provide encompassing and detailed feedback on the initial version of the questionnaire. Eight users have completed the questionnaire. The conclusions drawn in this section are based on these interviews, and they provide insight into human motivation and habits concerning blogging and tagging. However, due to the smaller number of participants and special purpose of our portal these conclusions should not be taken as universally applicable and exhaustive. Let us quote some of the responses to the three research issues we consider the most central.

2.1 Why Do People Write Blogs?

The questionnaire opened with the question: “Why do you write blogs?” Typical answers were:

- Because I have to and sometimes because I want to.
- Sometimes I confess and sometimes I write something useful for students, eventually useful for the world.
- I wrote blogs as a part of my course on web-design.

Besides for the obvious motivation connected to the educational purpose of our portal [4], the users have suggested the following: writing an article on topic of interest for oneself or for others; writing something useful. This clearly points to the use of blogs to organize and share knowledge. Interestingly, one user pointed out she was writing blogs in order to confess, therefore blog can be seen as a psychological tool that may help to deal with aspects of one’s personality.

2.2 What Strategy Do People Choose for Tagging?

The survey included several questions about tagging strategy: “What kind of words you usually use as tags?” “What do these words mean?” “Do you use

¹ A community portal used at the Comenius University, Faculty of Mathematics, Physics and Informatics.

names of people as tags?” “Do you use abbreviations as tags?” “Do you use more tags to name the same thing?” “Are the tags that you assign to one blog posting related, or are they unrelated?” Typical answers were:

- I use more tags to name the same thing, but I try to avoid such use.
- I do not use more tags to name the same thing, however I use hyponyms.
- I use the name of a person as a tag, in case it’s a famous person.
- I use abbreviations, definitely. Tag “FMFI” is a much better tag than “Faculty of Mathematics, Physics and Informatics”.

Some of our users try to avoid using more tags to name the same thing and *hyponyms* can be used to avoid this. If the article is about a person, it can be topically tagged by the name of the person. The interesting point suggested by one of our users is that it is reasonable especially if this person is known to others. Clearly this answer makes sense, it is not very likely that others would search using this tag if they do not know the person. People also prefer to use abbreviations if they are commonly used in speech or in writing.

2.3 What Are the Properties of a Good Tag?

We are also concerned with how to determine which tags are good. The following questions were relevant in this respect: “What kind of words do you use as tags?” “What do these words mean?” “What are the properties of a good tag?” “How would you define a well tagged blog article?” “When you choose which words to select as tags, do you consider that good selection may later facilitate search for the article?” We have learned the following answers:

- I am using nouns, sometimes proper nouns and names.
- A good tag should be a word, that comes first to mind to the majority of the users.
- A good tag should be precise, appropriate, intuitive, grammatically correct.
- A well-tagged article should have tags that are appropriate and brief, but not too specific neither general. It shouldn’t include too many synonyms.
- A well-tagged article should have tags, that describe each important part or topic of the article with the most general terms and at the same time with the most appropriate terms.
- Tags are supposed to facilitate later search for the article.

According to the responses of our users, a tag should be a word, usually a noun, which is appropriate in regards to an article. It should be also intuitive and the word that comes first to mind (activated immediately). A well-chosen tag should be grammatically correct and it should enable future retrieval of tagged content, many users agree, that this is one of the goals of tagging.

In this section we have presented the results of a survey we have conducted with users of our blog portal. We have learned some interesting observation and users’ views on blog publishing and tagging. However, we have also learned that some of the users are unaware of how to properly tag the content in order to assure good navigation through articles in the future. In order to build “intelligent” tagging tools, not only users’ but also experts’ opinion needs to be considered.

3 Psychology of Tagging

Categorization is one of the most fundamental and pervasive cognitive activities [5]. Categories are studied since the works of Plato and Aristotle. The classical view on categorization assumes, that each object can be a member of one category. This corresponds with a standard hierarchical categorization scheme, where each category can contain objects and one or more categories, however each object can be associated with only one category. This scheme is usually used to categorize books in a library or in early online catalogs of web content (e.g., Yahoo.com). Although tagging is closely related to this “classical” view on categorization, in current tagging systems we usually assign more than one tag to particular object, hence it belongs to more than one category.

Theories based on the work of Rosch [2], Lakoff [6] and Tversky [3] change the “classical” view on categorization by introducing the prototype theory. Prototype theory assumes that some members of a category can be more central than the others, e.g., *a raven* is more central to a category *bird* than *a penguin*. Systems of categories can be different for each individual, they are based on the experience, so different cultures and individuals can have different categorization systems, which also corresponds with Piaget’s views on constructivism [7].

Considering prototype theory and different levels of categorization, there is one important level of categorization - *basic level*. To illustrate basic level categorization, consider the following question: *What are you sitting on?* People usually say that they’re sitting on a *chair* rather than on a piece of *furniture* (more general category) or on a *kitchen chair* (more specific category). The basic level of categorization is also learned first by children [8]. Categories on the basic level are very informative and easy to imagine (e.g., car, dog). The categories are based on semantic features that are related to our perception of the real world [3]. Understanding human categorization through basic level theory is essential to user-centered design of taxonomies, ontologies and tagging systems [9].

In order to get a better insight on what mental processes are active during tagging, we should first consider some common motivations to tagging. There are basically two categories of motivations: organizational and social [10]. Tagging helps people to manage their information and knowledge, it enables people to express themselves and to have fun. Tagging also gives information management a social dimension by enabling people to share information and perform these activities collaboratively [11].

Inspired by current opinions in the blogging community [12], we hypothesize that the actual process of tagging an object consists of two stages. First, semantic concepts related to the object are activated immediately. Semantic relationships of activated concepts were discovered during cognitive experiments where subjects were asked to write semantic associations with an object. However, not all of activated concepts are suitable as a tag. Therefore a second stage is needed, during which one decides to use a particular concept as a tag or not, considering relations of the concept (tag) to the tagged object.

The process of tagging a blog article can also be divided into multiple sub-tasks, that user needs to perform, so assigned tags are relevant and informative.

The user of the tagging system should be able to indicate the main topic and sub-topics of the blog article and express it by using the correct words. In addition, a brief search for synonyms of these words should be performed. Since added tags will help other users to find the article, thinking about the usual searching behavior of other users may be also useful. Therefore, tags that others may use to find the article should be added as well. Finally, one should spell-check and evaluate the relevancy of each assigned tag to the article.

4 Guidelines for Improving Tagging Systems

Based on our research presented so far, we suggest that tagging systems should follow specific recommendations. This serves to the main objective of our effort, that is, to improve the quality of tag-based navigation, in other words, to facilitate the process of finding the desired information when tags are later used to access content. The recommendations we present here aim to support this goal by increasing the accuracy of tagging carried out by the user. They concentrate on two essential areas: educating the users so that they know how to tag better and enhancing the system in order to support users during the tagging task.

4.1 Educating the Users

Usually a lot of redundant and low-quality tags are assigned as tags, because users are confused and they do not have clear ideas regarding tagging. Users need to understand what exactly tagging is, how it works and what purposes it serves. This information should be provided to them in a suitable way (e.g., a help link near the tagging tool). Apart from that, we can further increase the quality of tagging by simply providing recommendations, best practices or examples of tagging. This is an important usability feature of any system that employs tagging. We suggest the following guidelines to be presented to the users:

Think topic-wise, do not miss any topic. One of the most common mistakes in tagging occurs when people only concentrate on the main/intended topic and fail to realize that the article also covers other topic.

Think specific and general at the same time. An article is only described properly by specific and general words together. This has to do with the importance of base-line categorization with respect to tagging.

Do not be afraid of synonyms. Different people may use different tags than you referring to the same concepts when searching for your article.

Think of what tags others would use to find the article. Certain words are more often used than others, consider how likely is that you would use the word when searching (e.g., more likely one would use *car* than *automobile*). Other commonly used words are names of people and places, etc.

Some tags are better in plural. Many times it is natural to name tags in plural. This is not a universal rule however (e.g., *cars* is a good tag but *home* is better than *homes*). This is more related to the use of these words as categories than to the actual number of things in the picture/article.

Enter multi-word tags correctly. Multi-word tags are also useful (e.g., *modern art*, *flight attendant*, *Philip K. Dick*). Users should not be afraid to use them, on the other hand they should be cautioned how to enter them correctly since different syntax is employed in different tagging systems.

Use established forms consistently. Consider what form of the desired tag is used by others. Which of the synonyms, whether singular or plural form is used, and consider also case in names and acronyms. It is a mistake to miss a tag which is frequently employed as such tags are given prominent place on web sites and hence assure that the content is found by other users.

Follow the community conventions. While these guidelines may be useful in general, it is more important to follow established patterns and community conventions in tagging. Only so the tags assigned by all users of the system can be used to navigate content consistently and efficiently.

Other recommendations are possibly found, however each community may prefer a different strategy and so it is important to follow community conventions above all. These recommendations should be presented to the users in an intuitive way so that they are not overlooked. This is no easy task for the designers. The following is suggested. Help pages that explain tagging conventions ought to be accessible directly from the tagging interface. Assisting tools should be implemented within the system that analyze user's inputs and display warnings/hints when needed. Example should be given by distinguished users which are prominent in the community. Every community has such users. They should be spotted and educated so they would serve as example to the others.

4.2 Enhancing Tagging Systems

Even if we educate users and recommend the best ways to tag their content, if we in addition provide supporting tools that facilitate tagging the outcome shall be greatly improved. In fact it is possible to support computationally a great number of mental tasks that are associated with tagging. The interface of an "intelligent" tagging tool should include features for this sake. These features interact with users in two ways: either they mark up suspicious tags so that the user can easily spot them and correct them if needed or they suggest additional tags to be considered. Below we list the most essential of such features:

- spell check – misspelled tags should be marked up;
- unrelated tags – tags that assumed unrelated to the text of article should be marked up as well;
- missed tags – additional related tags extracted from the text of article missed by the user should be suggested;
- plural form, and possibly other established forms (e.g., special capitalization) may be suggested for certain tags;
- synonyms should be suggested;
- more general concepts derived from tags already used should be suggested, thus increasing the chance of hitting the basic level (see Sect. 3).

Spell check is easily supported by building in one of the open source spell checkers that are available (e.g., Aspell). By applying standard machine learning based term extraction and concept extraction techniques we are able to extract characteristic concepts from the text of the article being tagged. By consequent application of term similarity metrics and comparison with the tags provided by the user we are able to estimate whether some concepts are missing or whether there are some possibly unrelated tags suggested by the user. In addition we will keep track on history of tags that have been already used in the past by different users. If for some tags there are preferred forms, e.g., plurals, capitalization, we will suggest these forms to the user. A ready to use tool to handle synonyms, more general and more specific concepts is WordNet. In addition tagging systems that track the history of used tags may employ ontology learning techniques [13] and may try to incrementally learn such relation between tags that have been previously used. One such method has been proposed in [14].

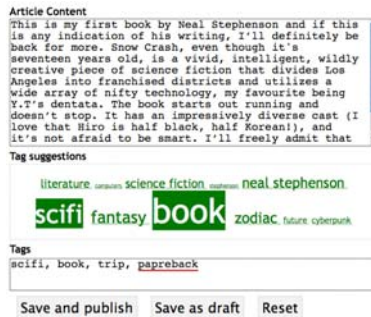


Fig. 1. Prototype of an enhanced tagging tool

In Fig. 1 a prototype of a tagging interface is depicted. This interface includes tag suggestions. Note that suggestions are displayed in enlarged type if they are considered more relevant with respect to the article text by the system. If the user selects some of the suggested tags, these are immediately copied below into the “Tags” input. Since multiple tags can be entered into the “Tags” input, users should be aware of how to separate them – in our case, users can separate multiple tags by using commas. However some systems use other separators, such as space, which may cause difficulties with entering multi-word tags (e.g., *modern art*). In that case, multiple words in one tag can be joined by underscores or hyphens.

5 Conclusion

We present a study in questions such as why humans write blogs, how they tag blog articles, what are the psychological processes that underline tagging, etc. It appears that both prototype theory [2] and basic level categorization [3]

are relevant to explain tagging. Based on this research and based on interviews that we have conducted with blog users, we present guidelines for tagging and we identify essential features for intelligent tagging tools in order to increase quality of the tagging process.

The next step will be to design and implement such an intelligent tagging tool to be used on our portal. We plan to employ ontology learning techniques [13] to discover semantic relations between tags and to record these relations in an ontology. With an ontology in the background, we hope to be able to develop intelligent tag suggestion based on semantic relations between tags.

References

1. Derntl, M., Hampel, T., Motschnig, R., Pitner, T.: Inclusive social tagging: A paradigm for tagging-services in the knowledge society. In: Lytras, M.D., Carroll, J.M., Damiani, E., Tennyson, R.D. (eds.) WSKS 2008. LNCS (LNAI), vol. 5288, pp. 1–10. Springer, Heidelberg (2008)
2. Rosch, E.: Principles of Categorization, pp. 27–48. John Wiley & Sons Inc., Chichester (1978)
3. Tversky, B., Hemenway, K.: Objects, parts, and categories. *Journal of Experimental Psychology: General* 113(2), 169–193 (1984)
4. Homola, M., Kubincová, Z.: Practising web design essentials by iterative blog development within a community portal. In: *Procs. of CSEDU 2009* (2009)
5. Wilson, R.A., Keil, F.C.: *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*. MIT Press, Cambridge (1999)
6. Lakoff, G.: *Women, Fire, and Dangerous Things*. U. Chicago Press (1990)
7. Piaget, J.: *Logique et Connaissance Scientifique*. Gallimard, Paris (1967)
8. Mervis, C.B., Crisafi, M.A.: Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development* 53(1) (1982)
9. Rorissa, A., Iyer, H.: Theories of cognition and image categorization: What category labels reveal about basic level theory. *J. Am. Soc. Inf. Sci. Technol.* 59(9), 1383–1392 (2008)
10. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In: *WWW 2006, Collaborative Web Tagging Workshop* (2006)
11. Smith, G.: *Tagging: people-powered metadata for the social web*. New Riders Publishing, Thousand Oaks (2007)
12. Sinha, R.: A cognitive analysis of tagging, blog article (2005), <http://rashmisinha.com/2005/09/27/a-cognitive-analysis-of-tagging/>
13. Buitelaar, P., Cimiano, P., Magnini, B.: *Ontology learning from text: methods, evaluation and applications*. IOS Press, Amsterdam (2005)
14. Frank, J., Homola, M.: Ontology-driven categorization of blog postings: A scenario. In: *Kognice a umělý život VIII* (2008)