

Folksonomy: The Manner of Data-processing and Organization

Ilva Paidere

Department of Information and Library Studies
University of Latvia, Faculty of Social Sciences
Lomonosova Street 1A, Riga, LV-1019 (Latvia)
Phone number of contact +371 28369310
Ilva.Paidere@lu.lv

Gunta Dogžina

Department of Information and Library Studies
University of Latvia, Faculty of Social Sciences
Lomonosova Street 1A, Riga, LV-1019 (Latvia)
Phone number of contact +371 28705991
Gunta.Dogzina@lu.lv

Dace Ūdre

Department of Information and Library Studies
University of Latvia, Faculty of Social Sciences
Lomonosova Street 1A, Riga, LV-1019 (Latvia)
Phone number of contact +371 26675366
Dace.Udre@lnb.lv

ABSTRACT

Nowadays folksonomy known as a manner of data-processing and organization within the Internet is very prevalent. The author of the document and other Internet users are allowed to put a keyword – a tag which helps to sort out the necessary data within the process of information gathering. Unfortunately most of the given tags are inappropriate and do not characterize the content of the document causing the possibility to meet the useless information.

The aim of investigation was to explore Latvian websites which offer the possibilities of folksonomy and to determine how informative tags are, as well as to find out how popular among the users the allocation of tags (democratic indexing) is.

KEYWORDS: folksonomy, Web 2.0, tags, websites.

1. KNOWLEDGE ORGANIZATION, KNOWLEDGE ORGANIZATION SYSTEMS AND ITS SIGNIFICANCE

Knowledge organization (KO) has several functions. First of all, it helps to find information which is necessary, second, it marshals information and, third, it gives information about the document. The theoretical meaning of KO is to find out the objective order of existent knowledge. KO has practical meaning, too. Its aim is to provide access to knowledge. (4)

“The term *knowledge organization systems* is intended to encompass all types of schemes for organizing information and promoting knowledge management. Knowledge organization systems include classification and categorization schemes that organize materials at a general level, subject headings that provide more detailed access, and authority files that control variant versions of key information such as geographic names and personal names. Knowledge organization systems also include highly structured vocabularies, such as thesauri, and less traditional schemes, such as semantic networks and

ontologies. Because knowledge organization systems are mechanisms for organizing information, they are at the heart of every library, museum, and archive.” (Hodge, 2000). (2)

B. Hjørland offers other kind of knowledge organization systems (KOS) as bibliometric maps, concept maps, hypertext, synonymy rings, typologies, topic maps, webpacs (WebPAC – Web Public Access Catalog) and folksonomies. (2)

2. FOLKSONOMIES: DEFINITION AND OPTIONS

Nowadays the Internet allows to process information easily by creating self-made metadata. Metadata standards define and describe ways for identifying and organizing the content of data. (1)

T. B. Munk and K. Mørk in their publication “Folksonomy, The Power Law & the Significance of the Least Effort” define folksonomy like this:

“The collections of user-created metadata are called folksonomies. The essence of folksonomies is user-created descriptive metadata as opposed to the traditional sender-determined descriptive metadata in taxonomies and faceted classification.” (3)

The term “folksonomy” occurred in 2003 when it was defined by the information specialist Thomas Vander Wal. The term consists of two words – *folk* and *taxonomy*. The word “taxonomy” comes from the Greek *taxis* and *nomos*. *Taxis* means “classification” and *nomos* means “management”. It can be translated as “folk classification system”. Folksonomy became popular in 2004 when it was defined in a new Internet Web 2.0 solution which has promoted the idea that users are part of the information producers and publishers.

T. Vander Wal explains that “folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one’s own retrieval. The tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information.” (9)

Tagging is also known as social classification or social indexing. Its main principle – the user, rather than expert, can take the matter of indexing to make tags appear on the network immediately.

Since 2004 the number of the sites that support tagging, for example, Flickr.com, Delicious.com, Livejournal.com, has increased. In Latvia it became as something new over the last four years when one of the Latvian sites offered to its users to tag within their Internet diary. Since then, there emerged certain sites (one of the most popular websites in Latvia is Draugiem.lv) that allow users to tag their galleries and diaries.

Among the Latvian sites there is widely recognized Latvian version of open access Internet encyclopedia Wikipedia which enables the creation of categories to accommodate the articles according to topics.

Jessamyn West explains that “when an item (..) has a tag added to it, the tag becomes a clickable link to more

items associated with that tag”. Users can create tag clouds that reveal the most popular tags. These tags has no hierarchy, they has neither related nor wider or narrower terms. (11) This means that searching documents by keywords will retrieve only those documents which have exactly the same keywords assigned to them. There is no control of synonyms.

The value of tagging depends on person’s own vocabulary, and how he is able to understand and add meanings to the information or object. (9)

Websites that support the tagging and principles of folksonomies work on principles of Web 2.0. Participating in them is very simple and designation of data is used in a new way how to find the information, for example, tag clouds. It should be noted that there is several synonyms for term “keywords” like “tags” or “labels”.

Tag cloud is a visual representation of the keywords. Usually it features the keyword list in alphabetical order often used as highlighted keywords with larger letters. It is possible to search by keywords using the alphabetical or popularity list. The tag is attribute of the information material, by clicking on it opens the tagged item. (8)

The vision of the future announces that users will have no problem in finding the necessary information by using options of folksonomy, but, in the absence of finding ways to organize and control the tagging, there is a possibility that the Internet will be flooded up because of unsuitable tagging.

Tagging is also known as democratic indexing. In the study about the used folksonomy in the website Flickr.com P. Rafferty and R. Hilderley argued that the democratic indexing is different from classic indexing by the fact that inside of process of democratic indexing a large role is given to the user, because the user defines the meaning and his own interpretation of the document. In this case it is described in the entire document as a whole, such as document author, size, date, etc. Democratic indexing principle is that the individual himself can play a potentially distinctive interpretation of the document. (5)

As P. Rafferty and R. Hilderley has written, democratic indexing has some novel features – every single item includes descriptive cataloguing and subject indexing based on user perceptions of the item; the ability of individual users to record their private indexes, offering a democratic approach to indexing. (5)

E. Quintarelli and D. Fichter suggest that folksonomies reflect the movement of people away from authoritative, hierarchical taxonomic schemes; the latter reflect an external viewpoint and order that may not necessarily reflect users’ ways of thinking. Folksonomy reflects user’s lexicon not the way of thinking, view, qualification, preconceptions, etc. (6)

3. ADVANTAGES AND DISADVANTAGES OF FOLKSONOMIES

There are different opinions about good and bad points of folksonomy. Mark Suster has said that folksonomy is democratic solution of information classification; he also admits that there are some problems – it is possible that folksonomy is the less nonorganized alternative of all. Power law almost always is better than some intelligent people law. Tagging and folksonomy can not be controlled. Modern folksonomy systems need to develop some regulations. (7)

User-created tags are a good way how to add inadequate metadata to the document easily. It shows users' needs and wishes. But there is no possibility to define the meaning of the tag, so the word which has several meanings can label absolutely different documents with different content, for example, "windows" which are built into the wall of a house or operating system "Windows".

As often as not, searching information by keywords, user finds information which mismatch expected results.

J. West points out that people use many different synonyms or even spellings for the same general concept. So, while one user tags a restaurant with the word "worldsbestsoup", another, while doing the searching, might never imagine that kind of tag to lookup the restaurant. (11)

There are specific relationships in traditional classification terms, for example, thesaurus shows relationships between broader and narrower terms. The same is with synonyms and homonyms within controlled dictionaries. Folksonomy does not control lexical synonymy, morphological synonymy and syntactic synonymy, as a result, user is not allowed to change ending of the word or expect that searching program will offer automatically the correct form of the word. If the item has words or acronyms which describe it, for example, "United Kingdom", "UK", "England", during the processes of classification and indexing, this fact has been considered, but in case of folksonomy, there will be offered only those documents which will be described with particular word.

Democratic indexing is a challenge how to involve users in creating metadata. Metadata created by professionals have high quality but are expensive. User-created metadata are an alternative. Pauline Rafferty and Rob Hilderley considers that there are some advantages of folksonomy or user-generated indexing – that tagging is cheaper and more economical in terms of time and effort than traditional indexing, and that the feedback which can be derived from user generated tagging can facilitate a high level of community interaction that would probably not be possible if decisions have to be made first about the rules governing any controlled taxonomy. (5)

As Jakob Voß admits in his publication "Tagging, Folksonomy & Co – Renaissance of Manual Indexing?" that the art of creating interfaces for developed tagging systems is still in its infancy. Some of expensive

thesaurus is not available even in digital form. The art of development of tagging system is on its dawn. KO will always need manual input, it is costly to manage – however, Wikipedia has shown that groups of volunteers can create large knowledge resources if a common goal and the right toolkit exist. (10)

4. PROCEDURE

A survey was made by sending 100 questionnaires via e-mail and by using website Frype.com.

Analysis of the items was made by studying tag clouds of five websites by choosing 4 most popular keywords (2 for textual items, 2 for images) and evaluating their adequacy.

5. DESCRIPTION OF WEBSITES

Website Draugiem.lv is one of the most popular websites in Latvia, it offers to create a profile, upload information about person, make galleries, blogs, upload videos and music, publish articles and enter into discussion. Main function of this website: never lose contact with your friends. Similar to Frype.com and Facebook.com

Orb.lv is similar to Draugiem.lv but with different design; it was made for people with uncommon way of life – the information in this website makes people think of what they just have read.

Poga.lv is similar to Flickr.com, it offers to upload photos. This website is not so popular between the respondents because it is mostly made for professionals of photography.

Wikipedia is a well known free-access online encyclopedia. In this survey there is analyzed the Latvian version of Wikipedia. There can be uploaded information about everything – movies, music, definitions, biographies, science facts, etc. Wikipedia is popular as a study aid, although users hardly ever think of wherefrom the information has come, and who has written it.

Website Klab.lv is similar to Livejournal.com and Wordpress.com where users can write their own e-diary. It is not as advanced as blog system, but very popular among teens and students.

All aforementioned websites has the option to tag users' own-uploaded information. This kind of opportunity in Latvia appeared only 2-3 years ago when Draugiem.lv started to offer new options of Internet, and sites of the Internet diaries allowed users to tag their diaries. There is small number of websites in Latvia which offers the options of Web 2.0, but those who does, offers these functions in good quality.

6. RESULTS AND ANALYSIS

Considering that there is raising number of computers and the Internet users in Latvia – in 2008 57% of population uses the Internet – it is valuable to know how far the ability to work with IT is evolved, and whether tagging is actual for the Internet users.

The results of the questionnaire (see Figure 1) show that most of the users (94%) are registered in at least one of the websites which allows use the options of Web 2.0. Only 6% have not registered in that kind of websites. They mentioned that they are not interested and do not have need for being registered in those websites. Some of the users are of the opinion that being a part of those websites is like a swaggering of own life and curiosity of other lives.

Users, who support that kind of websites (see Figure 2), mostly are registered in Draugiem.lv (49 respondents), a lesser number of users use information of Wikipedia (41 respondents). Orb.lv is visited by 18 respondents, but only 12 respondents are writing an internet diary at Klab.lv. No one of respondents use Poga.lv. It is possible that respondents do not need a special website for photos if they are not professional photographers because Draugiem.lv and Orb.lv already gives the opportunity to upload photos. It must be noted that most of the respondents use a number of several websites at the same time.

By summarizing the results of questionnaire (see Figure 3), it is obviously that not all of the registered users use the options of Web 2.0. 96% of users had uploaded information (image, text, and video) while 4% have not. Maybe it is because of reason they do not want to publish their private life.

As it can be seen in Figure 4, the most popular website is Draugiem.lv, it has been chosen by 47 respondents. Orb.lv has been chosen by 12 respondents which is the same number of Klab.lv. 2 respondents go for other websites, and there is only 1 respondent who has ever uploaded information in Wikipedia. Maybe it is because of users' incompetence or fear to upload such important information.

Analyzing the results (see Figure 5), it can be seen that the Internet users, which have ever uploaded the information, use the options of tagging (67%). 33% of respondents do not use this option.

When asking what kind of qualities keywords should have (see Figure 6), 18 respondents answered that keywords have to be interesting, in that way increasing the possibility that keyword would stand out against the background of other keywords in a tag cloud. 8 respondents maintained that keywords must interpret user's individuality. Most of the users (23 respondents) stick by their classical understanding of keywords and affirms that keywords should be descriptive, related with the topic of item. 3 respondents permitted that keyword can be unrelated with the topic of item.

The main function of tags is to describe item as explicitly as possible. 53% of respondents follow this principle, but 44% of respondents admit that they choose tag by their own liking and perception, considering this principle only sometimes (see Figure 7). 3% of respondents tag documents ignoring the topic of the document in such a way increasing information "noise" for other users who use these tags.

One of the greatest problems of democratic indexing is the individual point of view when tagging. Some users use symbols and codes as tags, for example, "X1", "XX2", some use "-" or "_" in place of space character. That is so-called private language. (3) Survey shows (see Figure 8) that user's emotional feeling has great importance in the process of tagging. 38% of respondents said that their mood affects tagging process (tag is a describer of their mood). 31% of respondents only sometimes allow their mood to affect the tagging. Same number of respondents eliminated a possibility that tag could describe their emotional world.

Tagging is a good and easy way how to organize information. As it is shown in Figure 9, all respondents like tagging in general and admit it as useful (82%), only 18% of users find tagging as useless and unnecessary. Mostly it is said by users who do not tag information.

One of the problems of tagging is that information retrieval can be much slower because of inadequate tags. It takes lot of time to find necessary information and screen the exact document. Still most of the users (see Figure 10) use chance to search information by keywords (74%), the rest part (26%) does not use this option. Some of them use searching only by controlled keywords which are added by information specialists.

The questionnaire demonstrates that majority of users (86%) only sometimes has found the exact information they wanted (see Figure 11). Only 14% of respondents have had opportunity to use correct tags and find exact information. Zero of respondents never have had a situation when tags disagree with information item.

7. ANALYSIS OF WEBSITES

Analysis of five websites has been done by choosing four most popular tags from the tag clouds of every website (2 for textual items, 2 for images) and evaluating their adequacy to items by following criteria:

- Word adequacy to the status of tag;
- Emotional aspect – to what degree emotions affect tagging process;
- Popularity of tag;
- Originality of tag;
- Number of tags for the item.

Adequacy of the keyword was appointed if it describes the concrete information of the item. Emotional aspect was appointed if user makes connection between tagging and his emotions. Popularity of the keyword was appointed if the tag is in the tag cloud and by evaluating the size of the tag in the tag cloud – large-sized tags are more popular. Unpopular tags are determined by searching them in searcher. Originality of the tags was determined by user's grammar and vocabulary – if the user uses standard labels, or if he uses his own special vocabulary. Number of tags which are added to the item shows how explicit user has tried to describe the document – if number is bigger; document has been

analyzed more detailed. Detailed elaboration of tags is determined by how widely the document has been analyzed, for example, “sports” – wide, “football” – detailed.

In Latvian websites there is small number of inadequate keywords, but there is tendency of creating codes as “mcx” and “xxx” which can be understood only for gentry.

The keywords mostly are generalized. There are situations when user creates tag not by analyzing the document, but by expressing himself as an author, for example, creating a tag, which is a user’s name or other identity mark.

Sometimes users try to express their emotions and individuality, trying to create stylish tags, for example, traditional tag “skateboard” is placed with “sk8” which is common among skaters.

Some of the users create tags not only by using nouns; they create tags from verbs and interjections as “ha ha”, “oh”, “huh” etc.

There are situations when user creates a tag, trying to describe whole document with this one tag, but there are situations when users try to add all possible tags to one document, avoiding to use word sets, for example, in case of “martial arts” they use two separated words “martial” and “arts”. That increases information mess for other users when one image has more than 20 tags and half of them are inadequate.

The analysis of images proves that most of the users often use keywords which do not describe the image, they mostly follow their associations, for example, tag “nurse” describes an image with a girl who has rubber gloves, in this case creating an illusion that girl is a nurse. It can be explained with different perception and understanding of the meanings.

There could be great mess in the tag cloud because of differences of languages, as well as singular and plural forms. Users choose varied forms of keywords, for example, an image can be described by keywords “photo”, “photos” or “picture”.

Unlike the mentioned problems with keywords, Wikipedia offers category assigning. Website performs referential function; therefore, categories must be concrete and completely describing the document. During the process of analysis it had been verified. Negative aspect of the website is that there are categories which do not contain any information.

The analysis has shown that folksonomies of the websites largely depend on users’ individuality, not on users’ ability to describe completely the content of item.

8. CONCLUSIONS

Folksonomy is one of the KOS which has grown out of the Web 2.0 development. This KOS is not as organized and controlled as others, but it is going to be controlled, thereby trying to decrease information “noise” which presents difficulties for information retrieval.

Folksonomy is the result of personal free tagging of information and objects. User, by his own perception, evaluates the content of the item and creates a tag for it, in that way organizing information in the Internet.

The knowledge and the individual perception of the user are very different, therefore the tags are subjective, inadequate or do not fully describe the content of information. To describe the content of information, user must be objective, he must imagine by which keywords other users could search the item, as well as he needs to be neutral and eliminating emotional aspect, which is done while doing classical indexing.

Even though democratic indexing is good and fast way how to describe an item, still it does not define meanings, control synonyms and morphological forms of the words, so there is a great mess in the tag cloud, because one term is described with several forms of word.

Folksonomy is a new alternative for organizing information cheaply, because classical library classification systems need finance and specialists for creating thesaurus. Democratic indexing needs only a computer and the Internet. This indexing does not lose its unique for its lacks, because initially it is meant for Internet users, not for information specialists. The results of questionnaire show that most of the Internet users like tagging and it seems that tagging is a very useful option, not mentioning the problems which appear when searching information by user-created keywords.

The results of questionnaire and keyword analysis verify that users often try to interpret their individuality while doing the tagging, as well as they turn meaning to associations and similarities, thus creating tag with one meaning to the item with other meaning.

It is difficult to analyze keywords because users have different perceptions of information located in items, so attention is paid to specific occasion, not to the problem in general.

Authors of the survey think that hypothesis – Internet user created keywords are general and do not describe items particularly enough – has been proved, because, as internet users admit, side effects and individuality has great sense while doing the tagging, wherewith losing the ability to be objective and neutral to marking items.

9. REFERENCES

1. **Goldberga, Anita.** *Digitālā vide un standarti* [online]. 2004-02-05 [cited 2008-02-21]. Available: www.km.gov.lv/ui/faili/2/MABSTK.ppt
2. **Hjørland, Birger.** *Knowledge organization systems (KOS)* [online]. 2008-29-02 [cited 2008-03-01]. Available: http://www.db.dk/bh/Lifeboat_KO/CONCEPTS/knowledge_organization_systems.htm
3. **Munk, Timme Bisgaard, Mørk, Kristian.** Folksonomy, The Power Law & The Significance of the Least Effort. *Knowledge Organization* : International Journal, Vol.34, 2007, p. 16-33. ISSN 0943-7444.

4. **Pagrode, Agnese.** *Zināšanu organizācijas sistēmas bibliotēkzinātnē un informācijas zinātnē* : maģistra darbs. Latvijas Universitātes Bibliotēkzinātnes un informācijas zinātnes nodaļa. Rīga, 2006.

5. **Rafferty, Pauline, Hilderley, Rob.** *Flickr and Democratic Indexing* : dialogic approaches to indexing [online]. The Ninth International ISKO Conference, July 4-7, 2006 [cited 2008-03-03]. Available: <http://www.emeraldinsight.com/10.1108/00012530710817591>

6. **Spiteri, Louise F.** Structure and form of folksonomy tags: The road to the public library catalogue [online]. *Webology*, Vol. 4, No. 2, 2007 [cited 2008-02-25]. Available: <http://www.webology.ir/2007/v4n2/a41.html>

7. Suster, Mark. Folksonomy : a short description of what it is and how it works [online]. *AIIM E-DOC Magazine*, 2006 [cited 2008-02-25]. Available: http://www.edocmagazine.com/article_new.asp?ID=32214

8. *Tag cloud* [online]. Wikipedia, free encyclopedia, 2008 [cited 2008-02-25]. Available: http://en.wikipedia.org/wiki/Tag_cloud

9. **Vander Wal, Thomas.** *Folksonomy* [online]. vanderwal.net, 2007-02-02 [cited 2008-02-21]. Available: <http://vanderwal.net/folksonomy.html>

10. **Voß, Jakob.** *Tagging, Folksonomy & Co – Renaissance of Manual Indexing?* [online]. January 2007 [cited 2008-02-25]. Available: http://arxiv.org/PS_cache/cs/pdf/0701/0701072v2.pdf

11. **West, Jessamyn.** Subject Headings 2.0 : Folksonomies and Tags [online]. *Library Media Connection*, April/May 2007 [cited 2008-02-25]. Available form: EBSCOhost Avademic Search Complete.

10. ATTACHMENTS

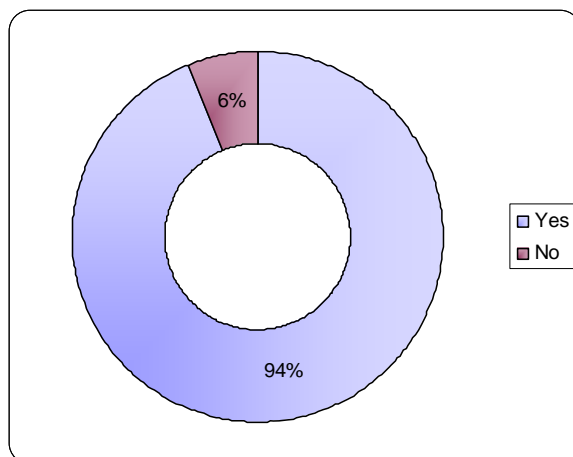


Figure 1: Do you visit the Internet websites in which users can upload their own information like image, text and video?

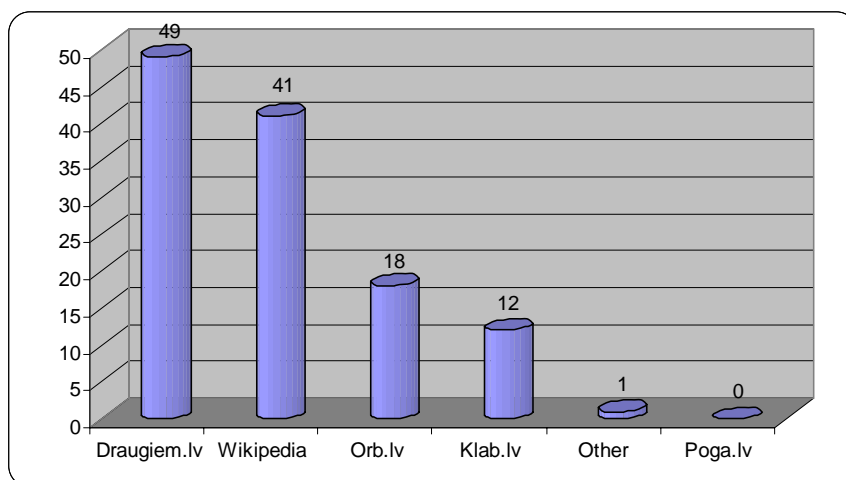


Figure 2: Which of these Internet websites do you visit?

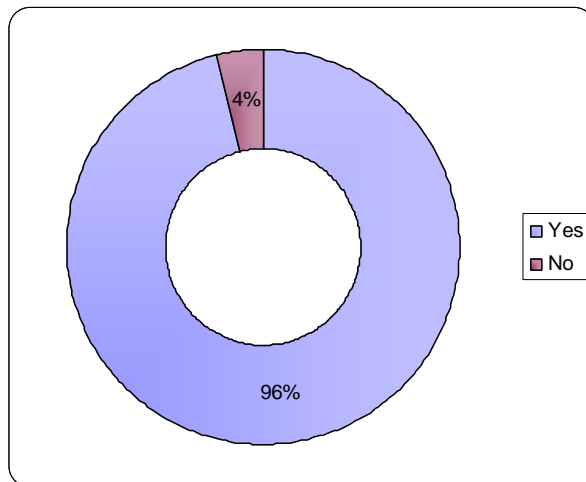


Figure 3: Have you ever uploaded the information on the Internet websites?

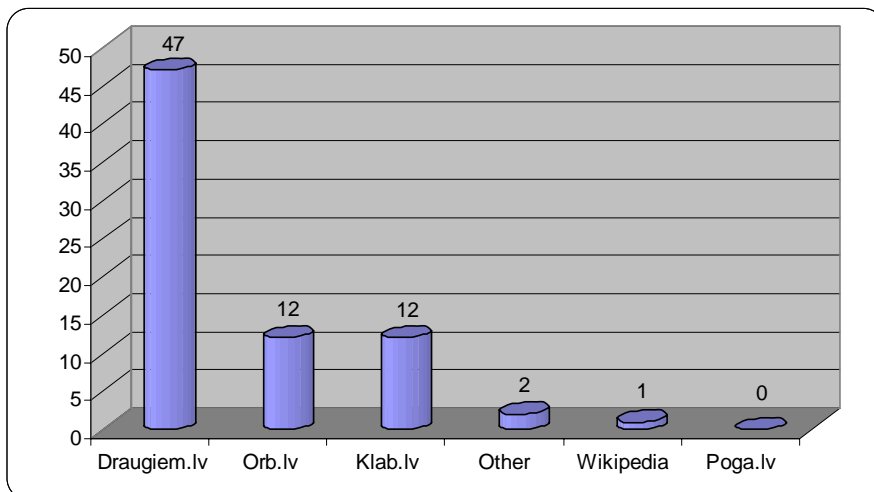


Figure 4: On which Internet websites have you uploaded the information?

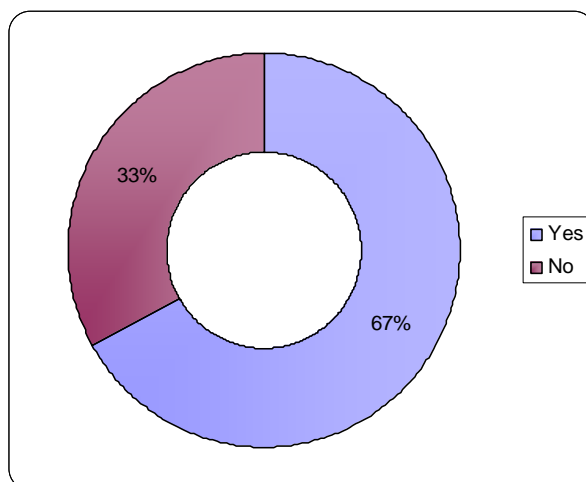


Figure 5: Do you use the options of tagging when you upload the information on the Internet websites?

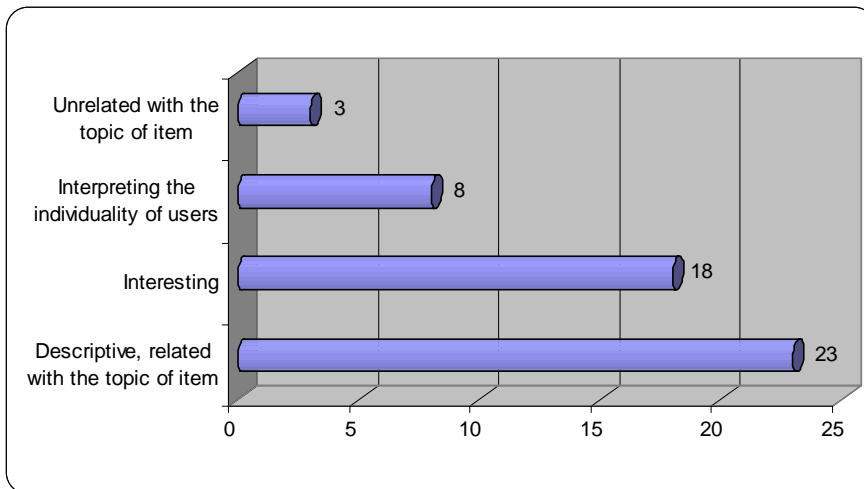


Figure 6: What kind of qualities tags should have?

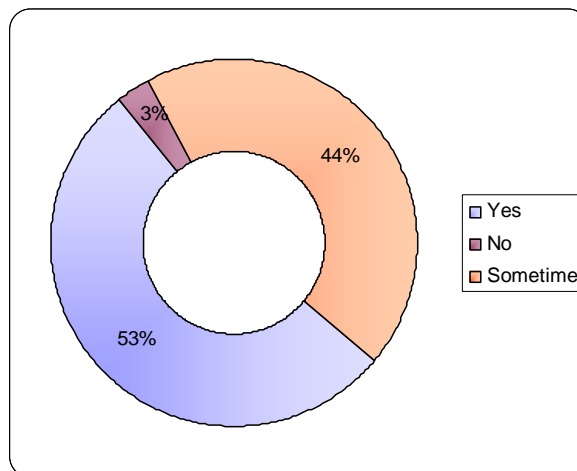


Figure 7: Do you take in account the topic of the item when you do tagging?

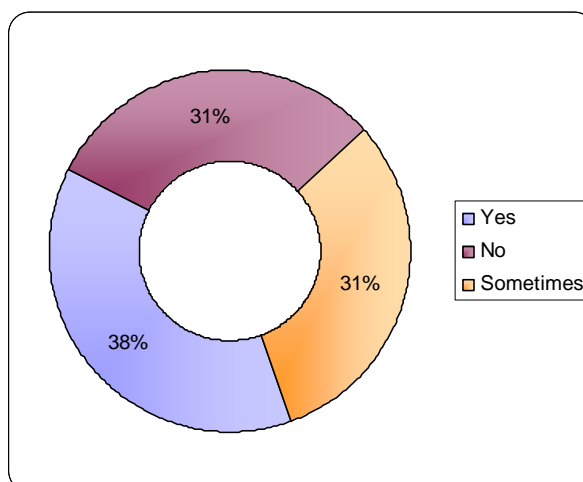


Figure 8: Do your added tags reflect your emotional feeling, too?

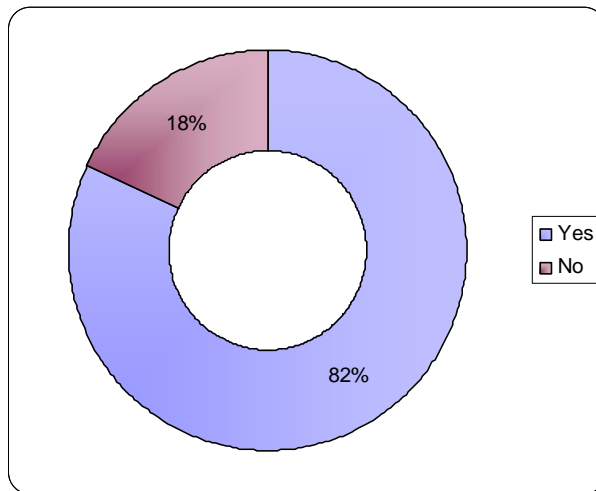


Figure 9: Do you enjoy and do you admit useful the option to do tagging?

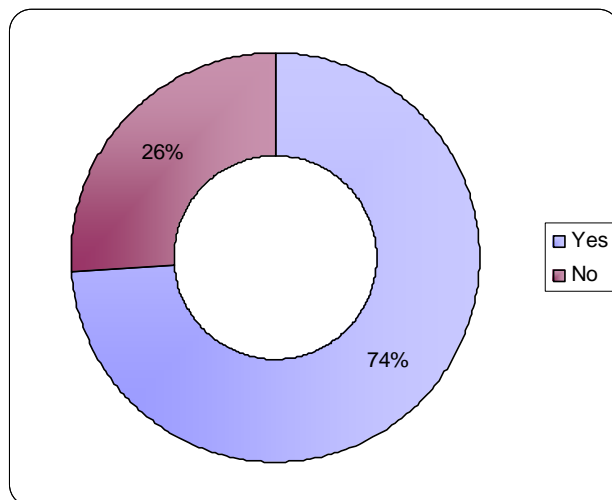


Figure 10: Do you use the tags for selecting the information?

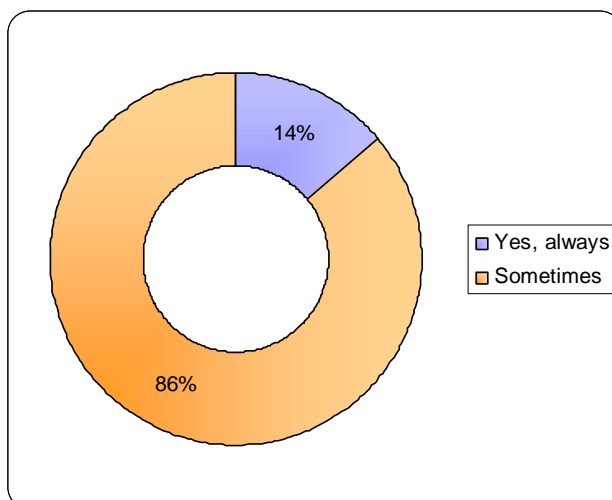


Figure 11: Do the tags always correspond to the necessary information which associates with the tags?