UDC and Folksonomies

Alenka Šauperl

Faculty of Arts, Department of Library and Information Science and Book Studies,
University of Ljubljana, Aškerčeva c. 2,1000 Ljubljana, Slovenia
<alenka.sauperl@ff.uni-lj.si>

Alenka Šauperl is an Associate Professor in the Department of Library and Information Science and Book Studies at the Faculty of Arts, University of Ljubljana. Her teaching and research areas are in the organization of information, including descriptive and subject cataloguing, as well as abstracting. She is the author of several books and many articles in the field of cataloguing and subject indexing.

Šauperl, Alenka. **UDC and Folksonomies.** *Knowledge Organization, 37(4)*, 307-317. 28 references.

ABSTRACT: Social tagging systems, known as "folksonomies," represent an important part of web resource discovery as they enable free and unrestricted browsing through information space. Folksonomies consisting of subject designators (tags) assigned by users, however, have one important drawback: they do not express semantic relationships, either hierarchical or associative, between tags. As a consequence, the use of tags to browse information resources requires moving from one resource to another, based on coincidence and not on the pre-established meaningful or logical connections that may exist between related resources. We suggest that the semantic structure of the Universal Decimal Classification (UDC) may be used in complementing and supporting tag-based browsing. In this work, two specific questions were investigated: 1) Are terms used as tags in folksonomies included in the UDC?; and, 2) Which facets of UDC match the characteristics of documents or information objects that are tagged in folksonomies? A collection of the most popular tags from Amazon, LibraryThing, Delicious, and 43Things was investigated. The universal nature of UDC was examined through the universality of topics and facets covering diverse human interests which are at the same time interconnected and form a rich and intricate semantic structure. The results suggest that UDC-supported folksonomies could be implemented in resource discovery, in particular in library portals and catalogues.

1.0 Introduction

Folksonomy is a form of indexing system. People participating in blogs, social networks, and other shared Web 2.0 systems and services assign tags to different information objects. When these tags are grouped, counted, and used automatically for browsing and searching, a folksonomy is formed. According to Vander Wal (2007) and Smith (2008), the first social tagging system, Del.icio.us, emerged in 2003, and the term appeared in 2004. It quickly attracted the attention of researchers in information and library science. A brief look at LISTA reveals that in 2005, six documents were indexed on that topic. There were 20 documents in 2006, 35 in 2007, 23 in 2008, and seven in the first half of 2009. Folksonomies are popular among Web users, allowing them to tag documents, which involves assigning keywords to resources they find on the Web or submit themselves. This enables them to retrieve documents they have already accessed or find new documents other users have tagged. Browsing is aided by tag clouds, i.e., groups of tags, sorted alphabetically, and presented by size - their size expresses the frequency of their use (see Fig. 1).

Some researchers (Mathes 2004; Noruzi 2006; Munk and Mørk, 2007; Spiteri 2007; Smith 2008) have welcomed folksonomies. Folksonomies allow for the spontaneous, quick and easy assignment of terms, which then serve as search and browse entries. They are democratic because everyone is able to join and contribute. There is no central authority or hierarchy that edits tags, censors, or supervises the system. Users are free to assign any term they choose or make up tags; there is no incorrect tag. Yet this plural system also allows individual user to develop his or her own coding system and use it, disregarding other users, if s/he so chooses. In some systems, the user can keep his or her tags private or make them publicly available. As a consequence, tags can display a rich and contemporary vocabulary.

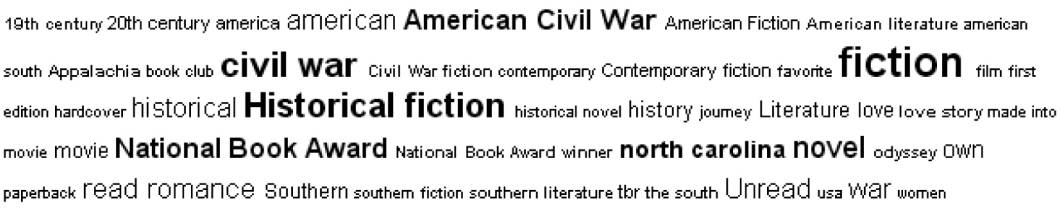

Figure 1. Tag cloud for *Cold Mountain* by Charles Frazier in Library Thing (accessed on 2009-07-07 from http://www.librarything.com/work/2421)

The same researchers (Mathes 2004; Noruzi 2006; Munk and Mørk 2007; Spiteri 2007) have also pointed out some drawbacks in folksonomies, one being that they have no terminology control. They do not manage synonyms and homonyms, and they allow singular and plural forms, as well as different spellings, including intentional or unintentional misspellings. Users can create new terms that may not be understood by the general public. Such terms may quickly become obsolete. Folksonomies do not support any semantic structure among terms. Relationships among terms have to be inferred by individual users. According to Smith (2008) and Steele (2009), this is actually a strength compared to traditional indexing systems. Until recently, many folksonomies did not allow the use of compound terms (or phrases). Searching or browsing was also only possible using one tag or term (not a combination of two or more). There is no limit to the number of tags assigned to one document and their specificity; this can lead to great inconsistencies. Because folksonomies rank tags and documents by popularity, less popular tags (and topics) may be hidden or difficult to find. Additionally, a few tags get used frequently and many just a few times (Zipf Law) (Halpin, Robu, and Shepherd 2007).

Searching and browsing are very different operations, particularly when performed in bibliographic databases of scientific and professional documents. Researchers need comprehensive and preferably complete knowledge of publications in their research area. Searching or browsing by folksonomies in a database when looking for articles supporting one's research is likely to be unreliable, because different and unrelated tags would most probably be used to mark relevant documents. While we have problems with indexing consistency in bibliographic databases, there is at least consistency in terminology. One particular term is consistently used to name a particular concept, as opposed to the many tags inherent in folksonomies. It may be different when one looks for information and does not require comprehensiveness, quality, and reliability. In such cases, any information on the topic may be interesting and useful. Browsing by following hyperlinks may actually bring the serendipitous discovery of highly relevant information.

Advocates of folksonomies believe that social tagging is a useful and more convenient alternative to traditional indexing systems (Furner 2007; Hayman and Lothian 2007). In fact, many libraries have already included social tagging in their library catalogues (e.g., Ann Arbor District Library from Michigan, US at http://www.aadl.org/catalog). This strategy is mainly aimed at attracting users to the library catalogue. It is not likely that users would be any more satisfied with this new search or browse process compared to results that can be obtained from the Web. It is natural to think that keeping track of the tags one has assigned is easier than following traditional indexing systems in a library catalogue. However, this is only true if these tags are few in number, and the user has enough time to navigate the system. We may anticipate that users will have difficulties when tags become too numerous. Over a longer period of time, users are likely to forget terms they have selected and assigned to similar documents (Iivonen (1990) or Olson & Wolfram (2008)). Frustration and dissatisfaction may cause users to abandon social tagging systems. This could be prevented by traditional indexing systems supporting folksonomies, an idea also presented by Binding and Tudhope at the UDC Seminar in The Hague in October 2009, Hayman and Lothian at the IFLA conference in 2007, and Kwan at the annual meeting of the American Society for Information Science and Technology in 2008. Kwan showed how *Library of Congress Subject Headings* could be transformed to support folksonomies. Our question was related: we wanted to know whether the semantic structure of the UDC could be used in complementing and supporting tag-based browsing.

Students of indexing at the Department of Library and Information Science and Book Studies, Faculty of Arts at the University of Ljubljana have examined folksonomies. They have confirmed the strengths and weaknesses of folksonomies that are highlighted

above. They have also seen that most of the concepts are present in the UDC (Demšar et al. 2009; Matoh and Koželj 2009). This paper brings their research a step further. Two specific questions were investigated:

1) Are the terms used as tags in folksonomies included in the UDC?; and,
2) Which characteristics of documents or information objects that are tagged in folksonomies compared to those that can be expressed in the UDC?

A collection of the tags from Amazon (www.amazon.com), Library Thing (www.librarything.com), Del.icio.us (delicious.com), and 43 Things (www.43things.com) was investigated. The universal nature of the UDC was examined through the universality of topics and attributes covering diverse human interests, which are at the same time interconnected and form a rich and intricate semantic structure.

2.0 A Closer Look at Four Folksonomies

Our selection of websites offering folksonomies and tags was not random. How does one develop a random sample from these constantly changing systems of unknown size? The change is constant because users keep adding tags. For example, in 43 Things, 139,786 people have made 324,109 resolutions and added an unknown total number of tags. On the other hand, the change is constant because tags get displayed according to the user's browsing strategy. Even most popular tags change in time because of different events (e.g., tags for Michael Jackson peaked after his death in Del.icio.us) (Del.icio.us trend graphs, 2010). We therefore decided for a purposeful sample of an undefined population of tags in just four of uncountable social tagging systems on the Web. For obvious reasons (this paper is presented in English), the selected systems are in English language.

The site Del.icio.us is a Web service that allows users to save, share, and organize their favourite bookmarks, i.e., addresses of web resources. This site was selected because web resources are not traditional library resources, even if they could potentially be so. The top 197 most popular tags from Del.icio.us were selected for analysis in June 2009. Further, 43 Things' top 198 tags were also analysed. This is a site that provides for users to note and share personal goals. It was selected on purpose to see how well the UDC embraces concepts for non-library materials. Another site we examined, Library Thing, is intended for users to catalogue books and similar traditional library materials. Three works were selected for analysis: *Cold Mountain*, a novel written by Charles Frazier; *The Little Mermaid* by the brothers Grimm; and *The Sound of Music* by Maria von Trapp. In each case, tags associated with the books, sound recordings, and movies were analysed (173 tags). These particular works were selected for two reasons: their popularity in Western culture and their existence in all three forms. The same sample was also analysed in Amazon, the popular Web bookstore (471 tags). Both the Library Thing and Amazon samples included items that would usually be considered library material. We wanted to see whether the UDC covers all the concepts users express in tags and what kind of document attributes get expressed in tags.

Our content analysis (Neuendorf, 2002; Lincoln and Guba, 1985) consisted of categorizing tags. Some categories were expected and prepared in advance (such as place, time, genre etc.). These expected categories were based on the disciplines expressed by the main UDC numbers and groups of auxiliary numbers. Other categories were new (e.g., accessibility, instrument, experience). New categories emerged during the analysis. The Slovenian translation of the UDC Master Reference File (MRF) 2006 was used to identify the appropriate UDC numbers for concepts expressed in tags. Spiteri (2007) performed a similar analysis of tags. Her categorization was based on the seven types of concepts listed in the NISO guidelines for thesaurus construction. These categories are: things, materials, activities, events, properties, disciplines, and measures. These categories are certainly valid; however, for the purpose of our research question, they were not helpful because they do not distinguish usual facets of the UDC - the topic, place, time, etc.

Mapping folksonomies to the UDC posed problems similar to those faced at other mapping projects. Some concepts matched in both systems one-to-one. In other cases, the concept from a folksonomy appeared in several classes of UDC. In such cases, the first class was selected. We were able to understand the concepts from the context, and some unknown concepts were checked by exploring the particular tag in the folksonomy. It is likely that fewer concepts would be found in an automatic matching process.

2.1 Presence of Concepts in Folksonomies and UDC

We expected that more concepts from Amazon and Library thing would be found in UDC and fewer from Del.icio.us and 43 Things. We expected that terms in 43 Things would be present less frequently in the

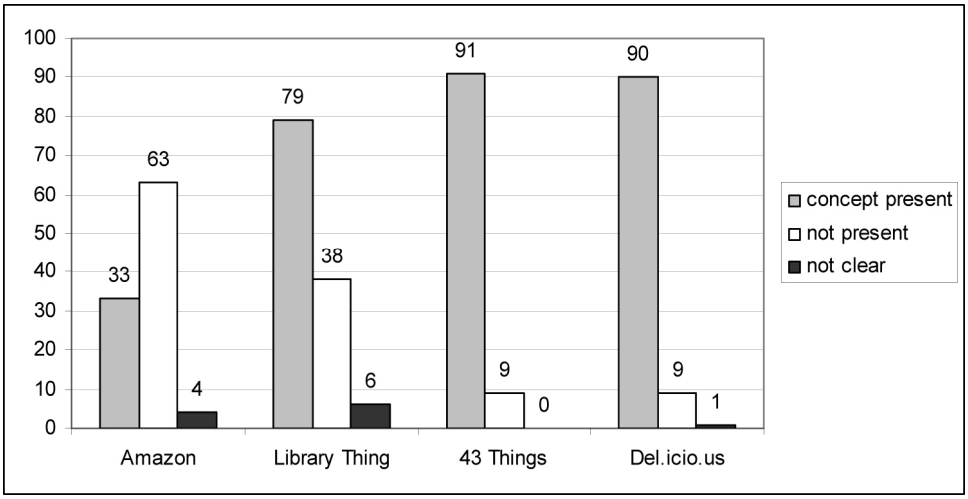

Figure 2. Percentage of concepts from Amazon (n=471), Library Thing (n=173), 43 Things (n=198) and Del.icio.us (n=197), present in UDC

UDC because the site focuses on people's wishes and plans, not on library materials. We expected the opposite from Library Thing and Amazon because we selected traditional library materials for their parts of the study sample. Our expectations proved wrong. More concepts represented in tags of Del.icio.us and 43 Things were present in UDC than concepts represented in Amazon and Library Thing (see Fig. 2). Among the terms not included in the UDC were a large number of names (further discussed in chapter 2.2). These could in fact be expressed in the UDC when associated with a main number. The share of tags we were unable to understand (such as "ada," "e" or "3844) was not very large in any of the analysed systems - the maximum was 6% in Library Thing and" 4% in Amazon, compared to 8% in Amazon reported by Demšar et al. (2009).

The largest number of tags from Del.icio.us could be found in the UDC class 0, followed by classes 3, 6, and 7 (see Table 1). Auxiliary numbers held the third rank. Most tags from this sample were in the area of computer science. This was also observed by Spiteri (2007). In 43 Things, the UDC class 6 contained the largest number of tags. It was closely followed by classes 3 and 7. Class 1 and 6 held the third rank. People seem to be mostly interested in health, arts, and topics of a social and ethical nature. Thirteen cases (7%) could be expressed by auxiliary numbers (such as time or place). Most tags from the Library Thing sample could be represented in the UDC class 7. Excluding concepts not found in UDC, auxiliary numbers rank second. Class 8 is ranked third, which mostly expresses genre. It was followed by classes 0 and 9. *Cold Mountain* and *The Sound of Music* are works concerned with historical topics, namely the Civil War and World War II. Therefore, the association with class 9 is not surprising. Class 0 represents reading (tags like 'reading' or 'to read') and medium (tags such as 'CD' or 'DVD'). Most tags in Library Thing were assigned to the books and movies categories, with the least number assigned to soundtracks. Tags assigned to the book frequently referred to the movie.

Excluding the leading category of concepts not represented in UDC, tags in Amazon were most frequently placed in class 7, expressing music and film. Auxiliary numbers were the second largest group of tags. They express the time or place of the story. The historical nature of two works (*Cold Mountain* and *The Sound of Music*) is expressed with tags in class 9 and is ranked third. Class 8 ties with 9. The large number of non-present concepts is due to the numerous names assigned as tags to these works. They are either names of authors, performers (artists), or literary characters. These (118 of 471 tags) could in fact be expressed in the UDC numbers if associated with a main number.

2.2 Nature of Concepts Expressed in Tags

It was mentioned above that names were frequently assigned as tags for books, movies, and soundtracks (see Fig. 3). Most frequently they were actors' names (e.g., Julie Andrews), followed by owners' names (Martha dvd movie collection), authors (Musker), literary characters (Ariel), names of people that were the topic of the work (biographical treatment of the Von Trapp family), and trade name (Amazon). It is

UDC class	Amazon		Library Thing		43 Things		Delicious		Total	
	n	%	n	%	n	%	n	%	n	%
0	6	1	6	3	21	11	84	43	117	11
1	0	0	2	1	24	12	2	1	28	3
2	0	0	0	0	10	5	1	0	11	1
3	4	1	0	0	38	16	21	11	63	6
5	5	1	1	0	3	2	3	1	12	1
6	0	0	5	0	31	17	21	11	57	5
7	79	17	32	46	32	16	22	11	165	16
8	16	3	14	8	5	3	4	2	39	4
9	15	3	9	5	4	2	3	1	31	3
Auxiliary	32	7	27	16	13	7	17	9	89	9
No	297	63	66	38	17	9	17	9	397	38
Ambiguous	17	4	11	6	0	0	2	1	30	3
Total	471	100	173	100	198	100	197	100	1039	100

Table 1. Number of tags from four folksonomies and their occurrence rate as concepts in Universal Decimal Classification

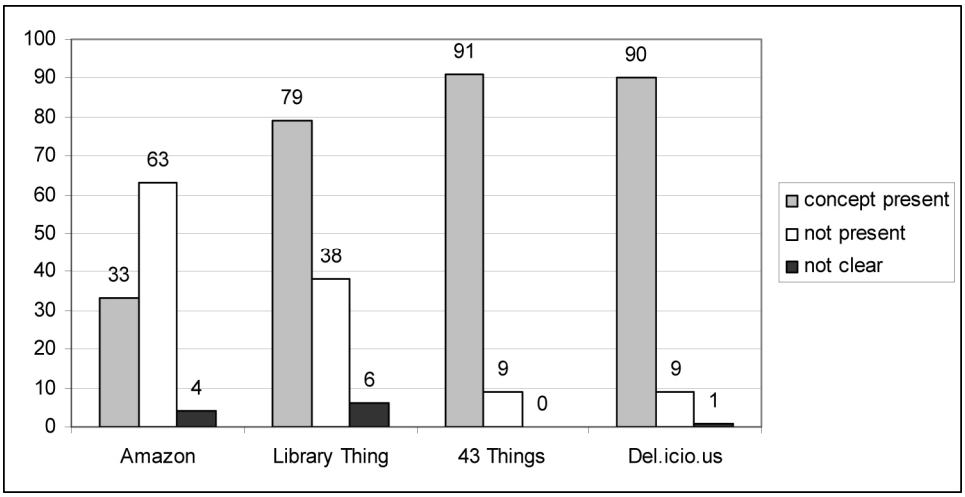

Figure 3. Percentage of names present among all tags from Amazon (118 of 471 tags), Library Thing (26 of 173 tags), 43 Things (7 of 198 tags) and Del.icio.us (16 of 197 tags)

interesting that names of actors in the film version of the story were assigned to books (e.g., Julie Andrews to the book *The Sound of Music*). Ownership was also expressed by pronouns (my dvd), verbs (own), or phrases (never got it). These names appeared in Library Thing and Amazon. In Del.icio.us, software names appeared (Linux), while in 43 Things, only three actors and four trade names appear among the most popular tags. Such names can be important for subject description in library catalogues and UDC, and some other indexing languages have provisions for including them. Curiously, the title of the tagged work was repeated among tags for the same work in about 10% of cases in Amazon and Library Thing. Titles of other works were also mentioned in Amazon and Library Thing, establishing a connection within the bibliographic universe. They may also be expressed in UDC, but only if they are the topic of discussion in the document. Otherwise, the relationship can only be expressed in the area of notes in bibliographic description. It seems that some support for *Functional requirements for bibliographic records* (IFLA 1999) could also be found among tags users assigned to these documents.

In total, topics were most frequently expressed by tags in the sample (see Table 2). In Del.icio.us, tags like "programming," "art," "food" or "education" represent topics. In 43 Things, tags like "career," "education,"

Attribute	Amazon		Library Thing		43 Things		Delicious		TOTAL	
	N	%	N	%	N	%	N	%	N	%
NAMES										
Artist	52	44	9	34	3	43	0	0	64	38
Author	25	21	8	31	0	0	0	0	33	20
unknown role of name	15	13	1	4	0	0	0	0	16	10
Title of work	14	12	3	11	0	0	0	0	17	10
Literary character	9	7	0	0	0	0	0	0	9	5
Biographical	1	1	2	8	0	0	0	0	3	2
Trade name	1	1	1	4	4	57	16	100	22	13
Owner	1	1	2	8	0	0	0	0	3	2
ALL NAMES	118	100	26	100	7	100	16	100	167	100
ALL CATEGORIES										
Name	118	25	26	15	7	3	16	8	167	16
Evaluation	78	17	6	3	0	0	5	2,5	89	9
Form	54	11	35	21	0	0	0	0	89	9
Topic	52	11	21	12	188	95	95	48	356	34
Genre	49	10	16	9	1	1	52	26	118	11
Not clear	17	4	10	6	0	0	22	11	49	5
Audience	16	3	9	5	0	0	1	0,5	26	2
Series/Collection	14	3	4	2	0	0	0	0	18	2
Related work	13	3	4	2	0	0	0	0	17	2
Plan/Action	9	2	7	4	0	0	1	0,5	17	2
Ownership - no name	9	2	4	2	0	0	0	0	13	1
Time	8	2	7	4	2	1	1	0,5	18	2
Gift	8	2	1	1	0	0	0	0	9	1
Place	6	1	9	5	0	0	1	0,5	16	1
Carrier	6	1	8	5	0	0	0	0	14	1
Award	5	1	3	2	0	0	0	0	8	1
Occasion	4	1	0	0	0	0	0	0	4	0,5
Experience	3	1	0	0	0	0	0	0	3	0,16
Instrument	2	0	0	0	0	0	0	0	2	0
Edition	0	0	3	2	0	0	0	0	3	0,16
Accessibility	0	0	0	0	0	0	3	2,5	3	0,16
TOTAL	471	100	173	100	198	100	197	100	1039	100

Table 2. Number of tags expressing different characteristics (attributes) of information objects from four folksonomies

"languages," or "memories" appear. In Library Thing and Amazon, topics are similar because of the sample. Tags, such as "romance," "ocean," "dance," or "american civil war" represent topics in the two systems. Names such as Charles Frasier or adjectives like American are rarely capitalized. Tags can also be misspelled (e.g., Hans Christian Anderson instead of Andersen). Both characteristics (misspelling and capitalization) have been noted as drawbacks of folksonomies by other researchers (Mathes 2004; Noruzi 2006; Munk and Mørk 2007; Spiteri 2007; Steele 2009). Regardless of these drawbacks, the topics can frequently be expressed with the UDC.

While names rank second among most frequently expressed concepts in total, genre ranks third in total and is the second most frequently expressed attribute among Library Thing, Amazon, and Del.icio.us tags, but only appears once in 43 Things. "Novel," "fiction," "biopic," or "adventure" are examples of this attribute. "Folk music" could be perceived as the topic. But when it was assigned to the soundtrack of *Cold Mountain*, it described genre. All terms can be

expressed by UDC. The form of the document is a closely related attribute. It is represented by terms, such as "soundtrack," "movie," or "news". The third related attribute, carrier, expresses the technology supporting the media, e.g., "blue ray," or the media itself, e.g., "DVD". These two attributes may also be expressed by the UDC. Edition is expressed just a few times in Library Thing and Amazon tags. It is also related to the form of document and can refer to the movie version, as opposed to the book. These two codes were kept separate because, in some cases, the movie tag would appear among tags assigned to the book. In such a case, this denotes a different edition of the same work, not the form of the information object it describes. The term edition is used much more loosely in this analysis and does not follow strictly the formal definition from ISBD.

Evaluation and form are attributes used with the same frequency, in total. Evaluation is expressed as "favourites," "oh brother," "singers who can't sing," "classic," "best movies ever," etc. It ranks second in Amazon, fourth in Del.icio.us, and ninth in Library Thing tags. This may be because Amazon sends an invitation to the customer to rate the item a week or so after the purchase. Awards are an attribute closely related to the evaluation. In contrast, with the subjective evaluation of items by users, awards are given by a professional authority and are regarded as an objective evaluation. While subjective evaluation was always discouraged in library catalogues, awards could be part of notes (area 7 of ISBD). None of these attributes are expressed in the UDC. Although it would be possible to construct a UDC number for the Annual Academy Award (Oscar), for example, its assignment to the document would mean that the award is discussed in it, not that it has received that award.

Collection or series is a relatively popular group of tags among users of Library Thing and Amazon, but not used in 43 Things or Del.icio.us. Users express it by assigning actual series names, like Golden Books or Disney Classics, or by naming their own collection, e.g., "Disney DVD" or simply "collection." This attribute can be closely related to ownership when it is expressed with a tag like "my DVDs." This attribute is reported in area 6 of ISBD, but not in subject description in library catalogues.

Neither the collection (series) nor the audience is expressed in 43 Things tags, while audience is mentioned only once in Del.icio.us. They are the sixth most frequent attribute in Library Thing and Amazon. It therefore seems that both attributes are more frequently associated with traditional library materials (printed books and sound or video recordings) compared to information objects that are not traditional library materials (web pages or personal goals). Users can note that the book, soundtrack, or movie is appropriate for a certain audience with tags like "children's book," "family movie," or "toddler". Regarding "toddler" as an example from the category "audience," which was assigned to the soundtrack of *The Sound of Music*, one wonders what the user intended to say. There seems to be no relationship between the two. This is one of the problems associated with folksonomies, noted in scientific literature (Mathes 2004; Noruzi 2006; Munk and Mørk 2007; Spiteri 2007; Steele 2009).

Some tags for sound recordings are categorized by instruments' names ("fiddle," "banjo"). They could easily be represented by a UDC number. The same is true for places, which can either be expressed by general terms ("mountains") or proper names ("North Carolina" or "Blue Ridge Mountains"). Place names were not coded with other names above, but only with this category. Place and time attributes were mentioned in 43 Things, Library Thing, and Amazon, but not in Del.icio.us. Both are represented by auxiliary numbers in UDC. Time was expressed with terms like daily, old time, 1990s, or 1999.

A group of tags expressing intention, plan or action ("gift 4 raven," "CD to review," "must own") or occasion ("mother's day") could hardly be part of a subject or bibliographic description in library catalogues for an obvious reason – they are highly subjective. This attribute only appears among tags in Library Thing and Amazon and cannot be expressed with an UDC number. Experience, expressed with tags such as "silly" or "suspense" is just as subjective and therefore never part of subject description in library catalogues, nor can it be expressed by a UDC number.

A very small number of tags express colour ("color movie"), source of purchase ["donation (T.Nicholas)"], accessibility ("free"), nation ("American Indians"), web address ("URL"), or past actions ("tagged," "seen"). Some of them could easily be expressed by UDC numbers, others not ("URL" or "seen"). These tags are of little importance because of the low frequency of their use.

3.0 Discussion

We expected that topics as the main UDC class numbers and place, time, language, and other auxiliary numbers would appear as attributes among tags in the four analysed systems. In fact, topic and genre

appeared most frequently among tags. Among the four common attributes present in all four analysed services, topic shows the highest frequency. The other three are name, genre, and time. Place and form of document appeared fairly frequently and in at least three systems, which means that they are not limited to traditional library materials. Form of document and medium (carrier) are two attributes related to each other that are both usually part of the bibliographic description and can also be expressed with UDC numbers. The question is whether repetition of the same data in the bibliographic description and subject description part of the catalogue record is reasonable and economically justifiable. Repetition of title proper of the work in about 10% of cases in Amazon and Library Thing seems particularly troublesome. Trant and Bearman (2008) report a similar finding. When they compared user assigned tags to museum documentation, object (45%), primary title (25%), materials (21%), creator (7%), and creation date (2%) were among the most frequent information, repeated in user tags and museum documentation. They did not question whether repeating data was reasonable. Furthermore, surveyed museum professionals in their study believed that user tags were helpful in searching.

We were surprised by the high rank of evaluation tags for Library Thing and Amazon. Librarians carefully avoid such categorisation because it is subjective and could be offensive to the user. It also cannot be expressed with UDC numbers. Another surprise was the large number of names among tags for traditional library materials in Library Thing and Amazon. Names can be added to the UDC numbers. However, the question is whether all the associated different roles names could and should be expressed in the UDC. A recent discussion on the ISKO mailing list revealed that there is a need to distinguish between real persons, literary characters, human, divine, and imaginary beings in the UDC. This is only appropriate for persons or characters that are the topic of the work, e.g., biographical description, literary study. People also have different roles in the intellectual creation of a work, they may be actors, animators, musicians, etc. All these roles are usually expressed in the area 700 of the UNIMARC catalogue record, where it can also be linked to the appropriate name authority files. Authority files then offer uniform headings in case the user searches by a different name. In our study, we had the case of a misspelled author's name - Anderson instead of Andersen. UDC would not be able to help the user in such a case, but a name authority file would. We would therefore join Richard Hartley (2009) in his suggestion to use name authority files in connection with folksonomies.

There are a number of attributes that seem to be important to users and are also part of the UDC. They are the audience, musical instrument, action, and occasion. Audience is usually expressed in coded fields in UNIMARC records. This information may also be included in the UDC number. Purpose (e.g., "gift idea"), intention (e.g., "to review"), or occasion (e.g., "mother's day") could be expressed by an UDC number, but the number would express the topic of the document, not the intended attribute. Because this information can be very subjective, it is not likely it could become part of the UDC in future. This means that folksonomies could accompany, not replace classification systems and subject heading languages in library catalogues.

A number of attributes that are important to users are not part of the UDC. Of those, some are part of the bibliographic description: collection/series, edition, accessibility, and URL. It is clear that the bibliographic description in Amazon is poor and does not allow the user to search by date of publication or edition. Users therefore have to resort to other means. They use tags to overcome the weakness of the system. One wonders whether the designers of Amazon and similar bookstores do this on purpose, to force the user to spend a long time searching and browsing for the appropriate title. During this process, the user is exposed to a large number of other titles. The share of attributes that are important to users but not part of the UDC is larger than the share of attributes that are part of the UDC. They are: award, experience, gift, occasion, and action. Award is a very useful category, as it intends to stimulate interest in the document. It should therefore be part of the bibliographic description but not part of subject description. If included in the subject description, it would mean that the work is about the award instead of indicating that the work received the award.

Ambiguity is a frequent complaint about folksonomies. Neologisms may be culturally biased. There may be some terms among the tags that we do not understand yet they are part of the user's everyday vocabulary. This is why we did not name our category "neologisms," but rather "unclear terms". One percent of tags were unclear in Del.icio.us, 6% in Library Thing, and 4% in Amazon. Unclear tags represented 8% of the total tags for *Cold Mountain* in Amazon in 2007, according to Demšar et al. (2009).

Spiteri (2007) does not analyse concepts she was unable to understand, but found that 10% of tags in Del.icio.us were neologisms, slang, or jargon. We do not know whether these percentages are high or low. It can be expected that they would cause difficulties in searching and browsing. It takes time to include new concepts in established indexing languages. Indexing with UDC or other indexing languages could not help users when searching or browsing for these terms, until they become well known and commonly used terms.

4.0 Conclusion and Summary

Inspired by numerous research reports on folksonomies, we analysed tags in four folksonomies: Del.icio.us, 43 Things, Library Thing, and Amazon. Not all of the analysed systems are intended for categorising usual library materials. However, their selection was intentional: we wanted to see whether UDC is in fact universal and can cover subject description of information objects, beyond library materials. We were not able to develop a random sample. Our sample is actually a very small representation of an unknown population of tags in each of the social tagging systems. We can therefore only claim that the findings hold for the sample but we cannot know whether they would hold for the entire population.

We found that 90% of tags for bookmarks in Del.icio.us and 91% of tags expressing people's goals in 43 Things could be represented by a UDC number. In contrast, only 79% of tags for books, sound recordings, and movies in Library Thing and 63% of tags for the same books, sound recordings, and movies in Amazon could be represented with an UDC number. This low share is not surprising when we analyse the kinds of concepts that are used as tags.

Names are among the most frequent tags in Library Thing and Amazon (16% of all tags in the sample). They could be part of the UDC if an indexer constructs a number with such a purpose. However, this cannot be an automatic process, and it would be more appropriate if authority files were linked to folksonomies to help users in selecting the appropriate form of name. Topic and genre, which rank among the three top categories of tags, could be represented by UDC numbers. Other categories that could also be represented with a UDC number include: form of document or information object, technology supporting its use or media, audience (e.g., children's book), musical instruments, place, and time. These categories represent 60% of all tags in the sample.

Categories which could not be expressed with a UDC number constitute 24% of all tags in the sample. They represent awards, series or collection, edition, evaluation, experience, action, occasion and purpose, availability, ownership, and related work. Some of these categories form part of a bibliographic description (6% of all tags in the sample). However, none of the analysed sites adopted ISBD and, as a consequence, their bibliographic information is not complete. It seems that this information is actually important to users for information objects like books, sound, and video recordings. This finding can contribute to the development of the ISBD or its successors. Among the remaining tags, which could not be represented by UDC numbers, evaluation alone holds a 9% share. It is unlikely that the evaluation would become part of a bibliographic or subject description. However, if this information were included among users' tags, it would probably be helpful to some library users. One could envisage that a user would evaluate library items and another user would find the first user's tags fit his or her literary preferences.

Golub et al. (2009) present a project where user tagging was enhanced by traditional indexing languages (DDC and LCSH). They found that users like to utilise the assistance offered by those indexing languages. We would take their suggestion further in the direction of Smith's (2008) observations of structured reports in Buzzillions.com (2009) and Mefeedia.com. These services identified the most frequently used facets among tags and structured their input according to these facets. We propose that the user is offered a structured form for adding his or her tags. The structure would separate personal and geographic names (not distinguishing real and imaginary persons and places). Authority files for personal and corporate names, and thesauri of geographical names could be offered here to help the user in selecting the appropriate form of name. UDC could be offered to help users in selecting topic, genre, form, and medium. When recording time, users should be offered examples of standardized forms of reporting time. It should not be too difficult to link data from the ISBD area 6 or the UNIMARC field for collection. Suggestions could also be offered regarding awards. On the other hand, evaluation, action, purpose, and experience should remain entirely free of suggestions. We also believe that users should not be forced to use only the suggested terms. They should be able to use either a suggested term or write their own. It would also be appropriate to ask users to suggest similar works. The form should provide space for entering any other terms a user

wishes to enter that are not appropriate to prescribed fields on the form. Our suggestion for a more structured interface with the user is also based on Markey's (2007a, 2007b) observation that people are generally inclined to work on the principle of least effort but are likely to be quite persistent if the system supports them during their work, search, and exploration.

We doubt that folksonomies could contribute significantly to the description of library resources. This thesis is based on three results of our analysis: 1) that a relatively small share of concepts represented by tags in Library Thing and Amazon can be found in UDC, and 2) that a few tags are used frequently and a large number rarely (Zipf's law). The first finding means that librarians and users have different views on the representation of information objects in catalogues and other information resources. Nevertheless, 59% of tags could be expressed with UDC numbers. About half of the 38% tags which could not be expressed with UDC are names. The second finding means that only a small part of the universal knowledge would be accessible through folksonomies, while indexing languages should still provide access to the rest of the knowledge world.

Our finding that a larger proportion of tags used in Del.icio.us and 43 Things can be found in the UDC compared to Library Thing and Amazon tells us that the UDC is indeed universal and could support not only library catalogues but diverse social networks and digital repositories as well. One could envisage either Del.icio.us or some company with a large document repository applying UDC to support navigation through large quantities of documents and information in other formats. The new decision of the UDC Consortium, made public at the UDC Seminar in the Hague in October 2009 to make a multilingual collection of about 2,000 UDC numbers publicly available on the Web may actually stimulate UDC's widespread use (UDC Consortium 2009a, 2009b). It may be appropriate to further invest in the development of the UDC and a free access version. This approach may be the way to connect library and internet communities and bring users back to libraries and library catalogues.

References

Buzzillions.com. 2009. Available http://www.buzzillions.com/

Delicious trend graphs. 2010. Del.icio.us blog. Available http://blog.delicious.com/

Demšar, Blaž, et al. 2009. Folksonomije = Folksonomies. *Šolska knjižnica*, 19: 15-23.

Furner, Jonathan. 2007. User tagging of library resources: toward a framework for system evaluation. In *World Library and Information Congress: 73rd IFLA General Conference and council*, 19-23 August 2007, Durban, South Africa. Available http://archive.ifla.org/IV/ifla73/papers/157-Furner-en.pdf

Golub, Koraljka et al. 2009. Enhancing social tagging with a knowledge organization system. Available http://www.ukoln.ac.uk/projects/enhanced-tagging/dissemination/entag-ifla-v3-final.pdf

Halpin, Harry; Robu, Valentin; and Shepherd, Hana. 2007. The complex dynamics of collaborative tagging. In *16th International World Wide Web Conference*, May 8-12, 2007, Banff, Alberta, Canada. Available http://www2007.org/papers/paper635.pdf

Hartley, Richard J. 2009. Folksonomies to ontologies: the changing nature of controlled vocabularies. In Griffiths, Jillian R. and Craven, Jenny, Eds., *Access, delivery, performance: the future of libraries without walls: a festschrift to celebrate the work of Professor Peter Brophy*. London: Facet. pp. 145-58.

Hayman, Sarah and Lothian, Nick. 2007. Taxonomy directed folksonomies : integrating user tagging and controlled vocabularies for Australian education networks. In *World Library and Information Congress: 73rd IFLA General Conference and council*, 19-23 August 2007, Durban, South Africa. Available http://www.ifla.org.sg/IV/ifla73/papers/157-Hayman_Lothian-en.pdf

IFLA. 1998. *Functional requirements for bibliographic records: Final report*. München: K.G. Saur. Amended and corrected Feb. 2009. Available http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

Iivonen, Mirja. 1990. Interindexer consistency and the indexing environment. *International forum on information and documentation* 15: 16-21.

Kwan Yi. 2008. A conceptual framework for improving information retrieval in folksonomy using Library of Congress Subject Headings. In Griffiths, José-Marie, ed. *ASIST 2008: Proceedings of the 71st ASIS&T Annual Meeting: people transforming information – information transforming people*. Proceedings of the ASIS&T annual meeting 45. Sliver Spring, Md.: ASIS&T, pp. 1-6.

Lincoln, Yvonna S. and Guba, Egon G. 1985. *Naturalistic inquiry*. Beverly Hills, Ca: Sage.

LISTA – Library, Information Science & Technology Abstracts. Available http://www.ebscohost.com/customerSuccess/default.php?id

Markey, Karen. 2007a. Twenty-five years of end-user searching, Part I: Research findings. *Journal of the American Society for Information Science and Technology* 58: 1071-1081.

Markey, Karen. 2007b. Twenty-five years of end-user searching, Part II: Future research directions. *Journal of the American Society for Information Science and Technology* 58: 1123-1130.

Mathes, Adam. 2004. *Folksonomies – cooperative classification and communication through shared metadata* [term paper]. Available http://www.adammathes.com/academic/computer-mediated-ommunication/folksonomies.html.

Matoh, Robert and Koželj, Aljoša. 2009. Predstavitev in analiza dveh folksonomij [Introduction and analysis of two folksonomies]. *Knjižničarske novice* 3. Available http://www.nuk.uni-lj.si/knjiznicarskenovice.

Munk, Timme Bisgaard and Mørk, Kristian. 2007. Folksonomy, the power law & the significance of the least effort. *Knowledge organization* 34: 16-33.

Neuendorf, Kimberly A. 2002. *The content analysis guidebook*. Thousand Oaks, Ca: Sage.

Noruzi, Alireza. 2006. Folksonomies: (un)controlled vocabulary? *Knowledge organization* 33: 199-203.

Olson, Hope A. and Wolfram, Dietmar. 2008. Syntagmatic relationships and indexing consistency on a larger scale. *Journal of documentation* 64: 602-15.

Smith, G. 2008. *Tagging: People-powered metadata for the social web*. Berkeley, CA : New Riders.

Spiteri, Luise, F. 2007. The structure and form of folksonomy tags: The road to the public library catalogue. *Webology* 4: article no.41. Available http://webology.ir/2007/v4n2/a41.html

Steele, Tom. 2009. The new cooperative cataloging. *Library hi tech*, 27: 68-77.

Trant, Jennifer and Bearman, David. 2008. Public and professional vocabularies: Comparing user tagging with museum documents and documentation. In *Networked Knowledge Organization Systems and Services : The 7th European Networked Knowledge Organization Systems (NKOS) Workshop : Workshop at the 12th ECDL Conference,* Aarhus, Denmark, Sept. 19, 2008. Available http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2008/programme.html.

UDC Consortium. 2009a. *UDC Licences*. Available http://www.udcc.org/licence.htm

UDC Consortium. 2009b. *UDC Summary*. Available http://www.udcc.org/udcsummary/php/index.php

Vander Wal, T. 2007. *Folksonomy coinage and definition*. Available http://www.vanderwal.net/folksonomy.html

All URLs were checked February 2010.