

An evaluation of enhancing social tagging with a knowledge organization system

**Brian Matthews, Catherine Jones,
Bartłomiej Puzoń**

STFC Rutherford Appleton Laboratory
OX11 0QX

Chilton, Didcot, United Kingdom
+44 (0)1235 446648

{catherine.jones,brian.matthews}@stfc.ac.uk

Jim Moon, Douglas Tudhope

Faculty of Advanced Technology,
University of Glamorgan
CF37 1DL

Pontypridd, United Kingdom
+44 (0)1443 482271

{jnjmoon,dstudhope}@glam.ac.uk}

Koraljka Golub

UKOLN, University of Bath
BA2 7AY

Bath, United Kingdom
+44 (0)1225 383691

k.golub@ukoln.ac.uk

Marianne Lykke Nielsen

Royal School of Library and
Information Science
Fredrik Bajers Vej 7K

9220 Aalborg Øst, Denmark
+45 (0)98 157922

mln@db.dk

Abstract

Traditional subject indexing and classification are considered infeasible in many digital collections. Automated means and social tagging are often suggested as the two possible solutions. Both, however, have disadvantages and, depending on the purpose of use or context, require additional manual input. This study investigates ways of enhancing social tagging via knowledge organization systems, with a view to improving the quality of tags for increased information discovery and retrieval performance. Benefits of using both social tags and controlled terms are also explored, including enriching knowledge organization systems with new concepts.

Keywords: folksonomy; social tagging; knowledge organization system; controlled vocabulary

1. Introduction

Knowledge organization systems have been used as tools for information discovery and retrieval in libraries and abstracting and indexing services, some for more than a century. Their benefits for improved information retrieval in the digital environment have been well acknowledged and recognized. They have devices to reduce the ambiguity of natural language when describing and retrieving items, and to allow access via browsing and navigation. However, there are costs associated with use of knowledge organization systems – manual indexing or classification are a significant resource, especially when performed by trained indexers.

Social tagging, as supported within community-based tools such as (Flickr, 2008) and (Del.icio.us, 2008), currently attracts much attention and are seen as a key feature of Web 2.0 services. They hold the promise of reducing indexing costs by drawing end-users into contributing, adding value as part of their interaction with information services. . However, social tagging is tailored to making it easy for end-users to describe information items and to have access to other users' descriptions rather than information discovery and retrieval in mind. The resulting collections of tags ("folksonomies") are completely uncontrolled, lacking even basic control of word forms such as spelling variants, synonyms and disambiguation of homonyms (Spiteri, 2007; Guy and Tonkin, 2006). Many users use tags only to characterise own documents, and not to help the community (Tonkin et al., 2008). On the other hand, user assigned tags could cover aspects that are not available in a knowledge organization system, especially when it comes to new concepts; as such, they could help update the knowledge organization system.

The EnTag¹ project explored the combination and comparison of controlled and folksonomy approaches to semantic interoperability in the context of repositories and digital collections. The aim is to evaluate the effect on both indexing and retrieval when using only social tagging versus when using social tagging in combination with a knowledge organization system. In this project, we describe the aims and methodology of the project, the results of one of the case studies in detail, and discuss the general conclusions of the project arising from both studies.

2. Related work

The need for knowledge organization systems in relation to folksonomies has been reported in the literature. (Weller, 2007) compares ontologies and folksonomies, suggesting that they are not to be seen as rivals but complement each other. (Noruzi, 2007) provides seven arguments for why a folksonomy-based system should use a thesaurus, emphasizing that there is no way to maintain consistency over time or across folksonomy users without a thesaurus. For (Connotea, 2008), a service for organizing references, recently an add-on tool has been developed that allows taggers to select terms from a knowledge organization system (Entity Describer, 2007).

(Smith, 2007) explores the connection between folksonomies and Library of Congress Subject Headings (LCSH) and describes advantages and disadvantages of each. She suggests that their product (called LibraryThing for Libraries) may “provide a compromise between the constraints of controlled vocabulary and the relative wilderness of the folksonomy”. (Hayman, 2007) argues for combining the best of the two worlds and describes its application on Australian collection of education resources (www.education.au). In their collection, knowledge organization systems are used for metadata creation and searching, and in order to keep pace with user needs, folksonomies are being explored. Users can tag resources by choosing from an established taxonomy or by entering their own terms. Users’ own terms will be used later to feed back into the taxonomy to improve its quality. The Library of Congress is collaborating with Flickr, in order to enhance bibliographic records for its images by end-user tags (Raymond 2008).

3. Overall approach

The aim of EnTag was to investigate the combination and comparison of controlled and folksonomy approaches to semantic interoperability in the context of repositories and digital collections, and had the specific objectives to:

- investigate indexing aspects when using only social tagging versus when using social tagging in combination with a controlled vocabulary;
- investigate above in two different contexts: tagging by readers and tagging by authors; and,
- investigate the influence of only social tagging versus social tagging with a controlled vocabulary on retrieval.

Thus the main focus of investigation was into the effectiveness of an enhanced tagging system. The enhanced system, with the capability of offering suggestions via a knowledge organization system, was compared against free social tagging. Two different contexts were explored: tagging by readers and tagging by authors. For each of these a separate demonstrator was developed, one operating on data extracted from Intute (Intute, 2009), and the other operating over Science and Technology Facilities Council (STFC)’s repository (STFC ePublication Archive, 2008) in which tagging will be conducted by authors submitting papers to the repository. A user study will be conducted for each demonstrator, which will allow a general comparison of a repository versus digital collection context, a different knowledge organization system, interface, and user community. In the EnTag Intute demonstrator, the Dewey Decimal Classification (DDC) forms the knowledge organization system, and (ACM Computing Classification Scheme, 1998) was used in the STFC study.

¹ *Enhanced Tagging for Discovery (EnTag)* project funded by the UK Joint Information System Committee (JISC), September 2007-October 2008 <http://www.ukoln.ac.uk/projects/entag/>.

4. Methodology

4.1 Intute Demonstrator Study

The Intute subject gateway is a UK-based database of hand selected and catalogued Web documents for education and research. For each document the catalogue record comprises title, description, controlled keywords, uncontrolled keywords, classification, type, URL, format, etc. In the study, 11,042 Intute catalogue records were chosen, all from the area of politics. Records were classified by the DDC, IBSS (International Bibliography of the Social Sciences, 2009) thesaurus and HASSET (Humanities and Social Science Electronic Thesaurus, 2009). Uncontrolled keywords mostly comprise names of countries.

The Intute study involved 28 participants who completed the study. They were all politics students at British universities, with one from the European University Institute. A call for participation was put together. They were recruited mainly from a written call for participation throughout the country. Each participant was given 4 tasks, and in each task 15 documents were to be tagged – 60 in total. Each task covered one topic of relevance to the politics student. Two tasks were controlled and two tasks free. In order to reduce the learning influence, tasks were rotated. A hypothetical group project scenario was outlined as a rationale and motivation for the tagging activity (by users as reader/searchers).

In each task, the participant was first to search for documents and then tag 15 of them. In controlled task, they were told to choose the top 15 documents, while in free tasks they could choose any documents they found relevant. In the case that a URL had become unavailable, the instruction was to move on to the following document. Tagging instructions specified that tagging each document should on average take between 5 and 10 minutes. They were to describe as many aspects and topics they thought appropriate for the task. They were also reminded to open the URL, but need not follow further internal links within a Web site. In case of very long documents, they were to focus on its abstract, introduction, conclusion, headings and table of contents. Additional instruction was added to tasks for Enhanced Tagger, to try to consider the suggestions from the controlled vocabulary.

Topics for the controlled tasks were suggested by a subject expert, PhD student in politics, who also evaluated whether there were at least 20 documents in the database relevant to the topics. The controlled task for Simple Tagger was on the topic of European integration. The controlled task for Enhanced Tagger was on the topic of peacekeeping.

After signing the participation form and completing a pre-study questionnaire, the Instructions document was sent out. It was the main document that each participant was given. It introduced the study and described each step the participant was supposed to do. The major steps comprised the following: 1) Technical requirements for using the system (with reference to the Settings document); 2) Learning the system (reference to the Training document); 3) Task 1; 4) Task 2; 5) Task 3; 6) Task 4; 7) Final questionnaire; 8) Emailing the results. Before starting the study itself, each participant was given a Training document through which she was to learn the system and try out tagging. The Settings document described how to enable scripting in Internet Explorer and Firefox browsers, and how to zoom the screen display for better viewing in Firefox.

The main method of data collection was logging the steps the participants conducted in the demonstrator. In order to help contextualize and explain the results better, questionnaires were also used. Apart from the pre-study questionnaire for collecting background information about the participants, the participant was to complete a post-task questionnaire, after every task, and a post-study questionnaire, after finishing all the tasks.

4.2 The STFC Study

STFC provides an institutional repository of publications for the staff and users of its UK laboratories, known as *ePubs*. Most staff who are scientifically active and publish paper will deposit papers (or records of papers) within the repository. Thus STFC targeted users as *authors* of papers from within its own staff.

The STFC study involved 10 participants who are scientifically active and deposit papers (or records of papers) within the STFC repository. They all worked within one discipline for consistency and ease of comparison, each have published refereed papers, and have deposited a number (> 10) papers within the ePubs repository, and can thus be considered as regular authors and depositors. The discipline chosen computer science and information technology, for which there was an easily available controlled vocabulary: the ACM Computing Classification Scheme. Since its main purpose is to classify papers which are submitted to various ACM journals, it has a widespread awareness and authority within the computing community. It was first imported into the SKOS thesaurus format (Simple Knowledge Organization System (SKOS), 2009), and then into the tool.

The sample of authors being by necessity small, this study sought to explore the purpose and ways of tagging from an author's perspective in a qualitative rather than statistical manner. In contrast to the Intute study focused on users, the participants were closely acquainted with the content and context of the papers being tagged and thus provided feedback on the comparative usefulness of free and controlled tags available to them. It was presumed that authors would like their work to be accurately indexed so that it can be accurately catalogued and retrieved from a search, and also knowing the work, they can accurately pick appropriate terms. The exercise sought to address a number of questions on the improvement of tagging from an author's perspective:

- What do they feel the purpose of tags is, and how would they like them to be used?
- Whether they consider using controlled vocabularies a worthwhile exercise over and above free tagging. Is the overhead of browsing and selecting from a controlled vocabulary worth the more accurate and consistent tags? Or is does freedom which free tags offer to allow them to pick their own terms give a more satisfactory result in their terms? (accuracy and efficiency)
- Do they consider the controlled vocabulary makes them consider tags which they would not have otherwise considered? (exhaustivity)
- Do they choose terms at a deep level in the hierarchy? (specificity)
- Is the controlled vocabulary structure (hierarchy, related terms) intuitive and easy to use?
- Is the user interface intuitive and easy to use?
- User satisfaction: is this something which they would like to see provided as a feature of the input system to ePubs?
- Are there any further tools and uses for tags (controlled or free) which the author would like to see?

The evaluation used the following methodology:

1. An author was invited to a session of approximately 40 minutes.
2. A worksheet was provided for the author, with notes and guidance on the task.
3. The author was invited to use the tagging interface to select tags for a number of papers which he or she has authored or co-authored (6 would be a reasonable number). This should take no more than about 20 minutes.
4. The author was free to select appropriate descriptive tags from any of the three approaches offered (free text, controlled vocabulary, tag cloud) as their preference.
5. At a later point in the trial, the observers suggested that they try a different ways of tagging using a different component of the tool in order to gauge their response to a different approach to their normal preference.
6. The session was observed by a member of the EnTag team, using an observer guidance sheet.
7. The member of the EnTag team will then discuss the exercise with the author to record impressions of the exercise.

The main method of data collection was logging the steps the participants conducted in the demonstrator. The Tagger tool collected information as follows.

- Tags collected on the paper

- Length of time to tag a paper.
- Proportion of free to controlled tags
- Number of tags selected
- Depth of hierarchy of tags selected
- Tags deleted.

The observers looked for user behaviour, whether the papers already had keywords assigned and general observations.

There are a number of known weaknesses to this approach:

- Small sample size: the number of available authors in the discipline willing to take part in the exercise may be too small to draw worthwhile conclusions
- The ePubs system already had a simple text field where free-text keywords can be entered, and a number of the papers considered in this trial already had keywords, potentially confusing the data. We left this data in the system as it was found to add another dimension to the data collected.
- Small number of papers tagged: the number papers each authors can tag in the exercise maybe too small to draw worthwhile conclusions.
- Inappropriate controlled vocabulary: the selected controlled vocabulary may not be appropriate for the nature of the papers which the authors are tagging.
- Unrepresentative selection of discipline: the computing and IT specialists may be early adopters of technology and also too close to the subject and too familiar with the concepts and thus unrepresentative of wider academic disciplines.
- The fact that all the use of the tool is done in one sitting, so rather than repeated uses as would happen in real life.

Nevertheless, it was felt that the results of the study were illuminating and useful.

5. Implementation

For each part of the project, separate demonstrators were implemented.

5.1 Intute demonstrator

The EnTag simple social tagging system is based upon ideas used in other Social Tagging Websites. Intute provided part of their Social Sciences database which the users can search to find abstracts of web resources that might interest them. The URL to the web resource is provided so that the user can view the actual document. The EnTag software permits a user to tag Intute resources. The EnTag Enhanced social tagging system provides the same capabilities as the Simple system with the additional feature of automatic tagging suggestions based on the Dewey Decimal Classification (DDC). Visual Studio, C# and ASP .NET were used for the development.

The prototype proved extremely robust; no errors were reported in the EnTag software for the duration of the user study. Some users did experience a few problems with the dated hardware and software configurations they were using. Each of the reported problems was resolved within a few hours, although, from one or two user comments, it appears that some configuration problems were not reported. A modern social tagging system requires quite a high bandwidth and the relatively complex interface of the EnTag system needs a reasonably high screen resolution.

A description of the Intute demonstrator follows. There are three major interfaces: searching, basic tagging, and enhanced tagging. The searching interface offers a tag cloud with tags linked to documents that they index, names of taggers linked to documents they indexed, as well as a free-text search box where searching can be limited to tags, title and description fields. Once a document is selected from search results, a tagging interface appears. Two types of tagging interfaces are developed: the basic, with tagging features usual in popular tagging services, and the enhanced one, with options from the knowledge organization system. Both tagging interfaces have the following options from which to select tags: a global tag cloud (an alphabetical list of all

tags in the system, with different font sizes relative to popularity); tags assigned by a specific user; and list of own tags. A user may also type in a tag. The enhanced interface (Fig. 1) involves suggestions from the knowledge organization system, presented in three frames. In the first frame, a list of DDC classes are automatically suggested based on matches with a tag entered by the user (by any means). If the user selects a class its narrower and broader classes are shown in the second frame, allowing interactive browsing of the hierarchical context. Simultaneously, in the third frame a tag-cloud like list of DDC captions, relative index and LCSH mapped terms is presented as a source of suggestions from which the user may optionally select.



Figure 1: Enhanced tagging interface.

An example of how enhanced tagging is implemented follows. After searching for the term “slavery” in all fields, a list of documents is returned. We choose the document with title “Slavery in New York”, an online exhibition on history of slavery in New York State (<http://www.slaveryinnewyork.org/>). We find the document useful and want to tag it. By clicking on “Tag” button, the enhanced tagging page opens. We enter tag “slavery” and click on “Suggest” button to get suggestions based on DDC terminology. The first frame (Possible Suggested Matches) then displays matching classes, which in our example includes:

- Slavery and emancipation
- Emancipation
 - Discriminatory practices and slavery
 - Extension of slavery
 - Slavery
- Etc.

Since the document is about the slave trade and emancipation movement, we click on “Slavery and emancipation”, which results in the second frame (Browse Suggestion Hierarchy) as shown in Figure 1, with “Slavery and emancipation” highlighted, and its broader and narrower classes. The third frame lists optional Tagging Suggestions. The terms chosen in this scenario are:

- Antislavery movements - United States
- Emancipation
- Slave trade
- Slavery

Thus, from the initial “slavery” tag, three other tags are derived via the knowledge organization system.

5.2 STFC demonstrator

The STFC demonstrator was an Apache Cocoon application using combined Java and XML; the underlying database was Oracle. This linked dynamically to the ePubs intuitional repository so that the user did a search in the repository and in a specially adapted edit mode, there was an option to enter the tagger system. The tagger application was then started and the title and abstract of the work was transferred together with any existing free tags. The controlled vocabulary was imported into SKOS and then into the tool. There is the potential to have more than one controlled vocabulary within the tool.

The Tagger interface is supplied in conjunction with the ePubs metadata editing tool so that tags can be entered for a specific publication by its authors. Figure 2 shows a typical tagging screen.

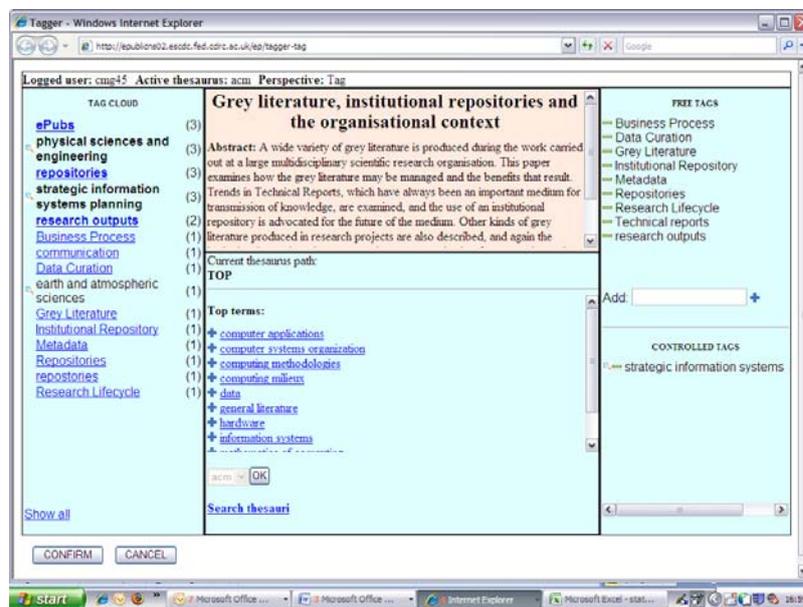


Figure 2. STFC interface.

- The screen is divided into four main areas.
1. At the top-centre, the title and abstract of the publication selected for tagging is displayed.
 2. At the bottom-centre, a browse interface for the thesaurus is shown. The chosen thesaurus can be changed by choosing from a drop down menu (only the ACM scheme is available for the trial). This has the following features
 - a. The top-level terms are shown initially,
 - b. The hierarchy can be expanded by clicking on a term (shows narrower terms and related terms)
 - c. The current path to the top of the hierarchy is shown as a “breadcrumb” trail along the top of the hierarchy. This can be used to backtrack.
 - d. Terms can be selected for adding to the “controlled tags” list by clicking on the “+” symbol to the left of each term.
 - e. The “search thesauri” link brings up a window where free-text terms can be entered which allow the thesaurus to be searched.
 3. To the left a “tag-cloud” is displayed, ranked in order of frequency of use of the tag. These can be selected by either clicking on them (if they are free-text terms) or clicking on the spyglass symbol displayed to their left (if they arise from the controlled vocabulary). This will enter them into the “free-text” add term box on the right of the screen, where they can be

accepted as a tag for this paper. The tag cloud as a default shows only the terms used by this particular author. By clicking on “show all”, the tag cloud of all authors can be displayed.

- To the right, the current selected free-text (above) and controlled vocabulary terms (below) are shown. These can be deselected by clicking on the “-“ sign to their left. In the centre of this panel, there is a free-text box, which the user can enter free-text terms and click on the “+” symbol to the right to enter them as tags for this paper. Multiple terms can be entered at once by separating them with commas.

Once a suitable selection of tags has been made using the tool, the user can accept them by clicking on the “Confirm” button at the bottom of the screen.

6. Results

Details of the results of the Intute demonstrator are presented in (Golub et. al., 2009). In this paper, we present the detailed results of the STFC case study.

We present an analysis of the statistics gathered in the STFC trial. As we have only ten participants, we can identify them individually using letters. All the STFC participants were male, and had significant experience in IT, with half with over twenty years experience, only two having less than ten years, and one having over thirty. Three were not native English speakers but were fluent in speaking, reading and writing English. All participants had published at least 10 items each, with a minimum of 11 and a maximum of 117. None of the participants identified themselves as frequent taggers, although one did use controlled vocabulary as part of his role. Only two participants had any knowledge of the ACM thesaurus before the tests, though several others were aware that it existed.

Number of tags. Over the study, on average 6 tags per item were assigned and 67% of the total number of tags assigned were free text. Figure 3 shows the number of tags and percentage free text for all our participants. There are no obvious trends. Analysing this same data using the length of IT experience, as in Figure 4, shows an interesting outcome where the most and least experienced staff add more free text tags than those in the middle of the range.

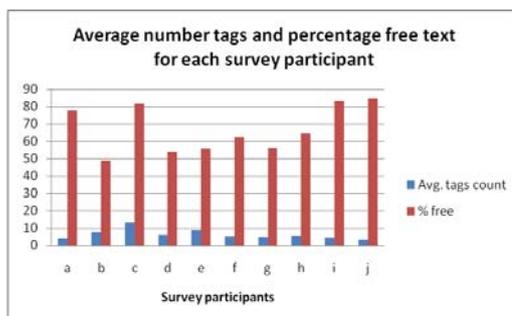


Figure 3

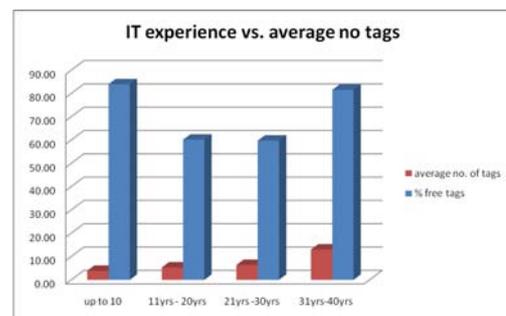


Figure 4

Choosing tags. Figure 5 shows the average time in seconds for choosing tags for a paper for the first four papers. A downward trend can be observed showing that as the tester becomes more comfortable with the tool and the vocabulary, then the length of time taken reduces. The downward trend is reversed when we add in the final two items (as in Figure 6) where we would have asked the testers to work in a way that was not their observed preferred way.

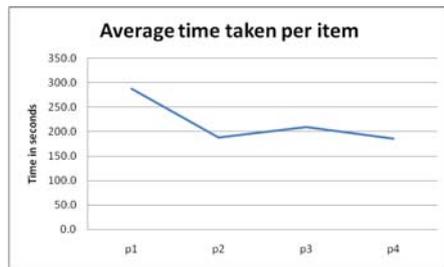


Figure 5

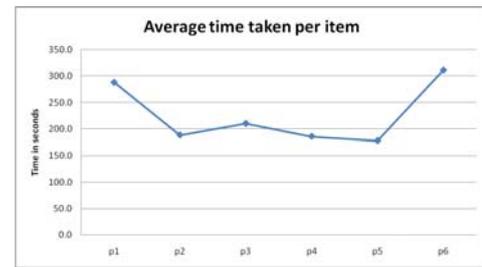


Figure 6

6.1.1 Observations

We record some key observations which were gained during the trials.

Preferred Style. It soon became clear that each individual had a clearly preferred style. Nine out of ten subjects carefully read the instructions before starting. Four subjects chose to use the controlled vocabulary most of the time, carefully browsing and selecting terms from the thesaurus. The browsing which they undertook is typically quite careful, with much backtracking up and down the hierarchy, and across related terms to find the most accurate terms. These subjects saw adding free-text terms as a last resort when they could not find terms from the controlled terms. One subject began with free-text entry, but soon switched to systematic searching of the vocabulary. He then found the Thesaurus Search feature and used that for the majority of searches.

Another subject found the “search thesaurus” option early and used this almost completely, trying to search the controlled vocabulary. He would carefully select a keyword to search with, and when this failed to find a desired term, would adjust the search. He would also do some browsing, and some unsuccessful interrogation of the tag cloud, but only entered free-text terms when prompted. Two other subjects also clearly preferred the use of the thesaurus search option for the majority of their selections.

A further subject entered free-text terms more or less immediately from the beginning, rapidly thinking up and entering terms without trying to look them up from the controlled vocabulary. Another used mainly free-text, and while he considered the tag cloud a good deal, could not see why he would use it.

Use of the tag cloud was limited and was used late in the process in almost all cases. Sometimes this was recognised by the subject and sometimes not. Usually, the observers had to point out the tag cloud and propose that they use it to select terms for later runs of the trial. Others found the tag cloud too large and confusing.

Thesaurus. There was a strong tendency in almost all subjects to select terms from the bottom of hierarchies. One subject entered the same term twice after accessing it from different routes through the vocabulary. One subject reported that he felt that he did not understand how the controlled vocabulary organised terms from just observing the top levels. One subject entered organisational and project keywords, which are specific to the organisation, and would not be part of a general thesaurus, but could be supplied as a specialist set of terms for the specific organisation. One subject deliberately picked singular rather than plural terms and was careful about capitalisation.

6.1.2 Post-task Questioning

After each task, the participants completed a post-task question and answer session.

General Impressions of the tool. Five of the subjects understood what the tool was intended to achieve, and why it was beneficial to provide such keywords to index their papers. Six of the subjects found the tool quite straightforward and useful to use – “perfectly usable” one said. One said that the tool needed to be much easier to use than this. He felt that the it “looks

like a tool for an information professional”, although he could see the tool used, if it had automated features. One subject queried whether there was any benefit to author of adding any subject keywords? He felt that the benefit to the user in retrieval would need to be emphasised to get it to be used. Two subjects said that they would be prepared to use the tool if the vocabulary was suitable fit to the main focus of their work. One subject said that he would be willing to use if we can prove that a searcher would actually use these tags in locating items

Preferred style. Those that began with a methodical search through the keywords saw that there was an advantage in providing a standard description, and saw free-text tags as being complementary when you cannot find a keyword. One subject said that he did not want to put in a free-text tag in case it was in the controlled vocabulary. One said that the use of a standard vocabulary “made you think” about the exact subject of the work in hand. One said that there was a tendency to “crowbar” the meaning desired into the terms provided by the controlled vocabulary. One said that there were limitations of vocabulary search: “navigation through the tree is fraught” while searching locates the right place and then can explore. They have got to be able to find terms in the vocabulary. He would prefer to free text tag if not well covered but controlled vocabulary gives a structure and this helps searching and retrieval.

One subject said that he wanted to use the advantages of the controlled vocabulary, and found searching it faster than browsing, whilst one other said that he liked the “intellectual challenge of finding things in browse” – rather than using search. One subject who used mainly free-text tags said that controlled vocabulary was only useful if you know it – need more explanation of the meaning of the terms

Suggestions for changes and additions. One said that the full tag cloud was too general and would like to see it made more specific for his purposes. Four subjects proposed that automatic keyword extraction – from titles, abstracts or full text - would be a useful addition to make the task easier, and keep user work to a minimum. One subject said that they would be more likely to use the tool if it was in combination with a automated system as this would save duplicating work. Another subject said that they would like a keyword “suggester” tool in conjunction with the free-text entry, and another said similarly, that he would like to be able to find matches from the controlled vocabulary from the free-text term added. Another subject suggested that a suggestion tool for the tag cloud might be useful, while one suggested making the tag cloud alphabetic rather than usage ranking. One did not like the phrase “Tag Cloud” – would prefer “popular” or “common tags”. One subject felt that the “show all” in the tag cloud didn’t suggest that this would show everyone else’s tags. He would be interested in selecting a group of people and seeing their tags rather than everyone’s.

Comments on the Thesaurus chosen. Three of the subjects recognised the vocabulary and knew where it came from. Some had used it to classify journal submissions, though only one said he was familiar with it. Five subjects said they found the thesaurus quite theoretical and abstract, and felt a more practically oriented thesaurus would more accurately reflect their work. One said that ACM was fine, but it was hard to choose a precise term, and it would be useful to provide more information to use, such as definition. Two others said that definitions of terms would be helpful. One said that he felt he needed to understand the ACM scheme in more detail before he could use it effectively. Three subjects said they would consider that a choice of vocabularies was needed, to cover the computing aspects and the application areas. One subject felt that the thesaurus chosen was rather out of date.

Comments on the user interface. Two subjects felt that the presentation of the Tag Cloud was confusing, with a different treatment of free-text and controlled vocabulary search, a non-obvious selection mechanism and no search. There was a feeling that there were too many terms seen at once – one other subject said that it was too big to comprehend. Another subject said that they would expect some classification/grouping/hierarchy in the tag cloud. One other subject used the Tag Cloud a good deal, but could not see the value of using other people’s terms. One subject would like a more visual browsing interface as he found that the searching through the hierarchy awkward and unintuitive. Would like to see a more contextual view of the place in the thesaurus he was searching within. Another subject was similar, wanting to see more than one

layer of the hierarchy at once, and a more visual search, if it could be achieved in a non-cluttered manner. Similarly, one felt that it needed a better visual way of showing the hierarchy so that one can see the global view whilst navigating a small local view.

7. Conclusions

7.1 Conclusions on the STFC demonstrator

From these observations, we determined some general conclusions about the use of controlled vocabulary and free-text input for tagging.

- There was general sentiment amongst the depositors that choosing terms from a controlled vocabulary was a “Good Thing” and better than choosing their own terms. The subjects could overall see the value of adding terms for information retrieval purposes, and could see the advantages of consistency of retrieval if the terms used are from an authoritative source.
- Most claimed that they would be willing to use a tool similar to the one provided, albeit with some reservations and suggestions on the interface.
- Several subjects pointed out that usability would be greatly enhanced by provided some automatic assistance in suggesting tags, to find suitable tags in the vocabulary or the cloud, from either known information on the paper, or else from free-text input. This would speed up the process and make it more accurate.
- ACM was not seen as a good enough for the purposes of this group. Either the coverage was not seen as adequate, it is seen as theoretical, or is not up to date. However, some recognised that as a controlled vocabulary used by a publisher it had its place.
- The Tag cloud was not a success. Most did not use it, and those that considered it either found it confusing to use or presenting too many options. It is interesting to speculate why this may be. Whilst an experienced group of IT professionals, they also claimed not to be major users of tag in most applications, and perhaps a larger sample could determine whether different groups with different types of experience would use the tag cloud more. However, there were also clear problems with the realisation of the tag cloud concept in this tool, and a revised interface may have more potential to be used more effectively.
- Most people were conscious of consistency and retrieval and used capitalisation and plural/singular in terms consistently. There may be a bias here – they were aware that they were being observed. However, this is consistent with the methodical approach that this group took to the tagging task.

Most people have a strong preference for the way they interact with the system, and how they used the variety of tools on offer, and have a clear preference not to use it in other ways.

We would also make some recommendations about the design of the tool:

- The search should to look at more than one source of information at a time – controlled vocabulary and tag cloud. Also a search from free-text input.
- The tag cloud needs some work on it. People seemed to suggest that it needed structure (subject/friends/etc) to be of use. People were more interested in their own terms than others.
- A significant minority wanted a more graphical/visual representation of the controlled vocabulary – probably to help them navigate.
- Spelling checker functionality for the free text at input stage?
- Help and definitions of terms should be supplied if possible.
- Automatic assistance in suggesting appropriate terms.

Most depositors had a strong preference for the way they interact with the system. The findings suggest three main groups:

- 1) Free text taggers: they enter many free-text terms, and don't care about the use of the controlled vocabulary.

- 2) Thesaurus browsers: they systematically browse the hierarchy of the controlled vocabulary, and only enter free-text term when the controlled vocabulary does not have a term they are comfortable with.
- 3) Thesaurus searchers: they prefer to interact with the controlled vocabulary via the search tool, then move to browsing and only enter free-text term when the controlled vocabulary does not have a term they are comfortable with.

We would speculate that another category of people would start with the tag cloud and mainly select tags from there. However, this approach was not represented in our group.

7.2 Comparison of Intute and STFC Results

The subject groups of searchers/readers (Intute) and authors/depositors (STFC) clearly have different roles (although they can overlap). Also regular depositors tend to be more mature than the searchers, due to more experience within a discipline being required before authoring papers. Nevertheless, a number of similarities between the Intute and STFC users could be identified. In both contexts the following was the case:

- Users appreciated the benefits of consistency and vocabulary control and were potentially willing to engage with the tagging system;
- There was evidence of support for automated suggestions if they are appropriate and relevant;
- The quality and appropriateness of the controlled vocabulary proved to be important;
- The main tag cloud proved problematic to use effectively; and,
- The user interface proved important along with the visual presentation and interaction sequence.

7.3 Overall Conclusions

The Intute study showed that most reader tags are added by typing them in directly, as is common in social tagging applications; of the other features used, the most frequent one is DDC suggestions, and another tagger's cloud. That the participants appreciate the DDC suggestions was also seen from their comments. In the STFC study some participants prefer typing their own tags in, while others use the ACM Computing Classification Scheme suggestions. They overall see that choosing terms from a controlled vocabulary is important for consistency in retrieval.

The main tag cloud, common in social tagging applications, is little used in both studies. Part of the problem is that there are too many options from which to choose and excessive scrolling due to the size of the cloud. This could be reduced by improved search/browsing and personalised ranking, clustering or filtering. However, it is not clear that displaying all tags is useful for retrieval purposes with large collections. Further work is required to investigate the best presentation and use of tag-clouds, including as tag clouds based on communities of friends/colleagues.

The Intute study also showed that in enhanced tagging, while users appreciated the 'direct' suggestions (third pane) and made some use of the disambiguation interface element (first pane), they did not browse the Dewey hierarchy very much (middle pane). Further work is needed to explore when browsing functionality is desirable in this context.

In the Intute study both simple and enhanced tagging provided additional entry points beyond the original subject indexing, as well as compared to free text search engines. There was some evidence that, in particular, controlled suggestions provided additional access points beyond the literal text.

User experience and task completion showed that both demonstrators were usable with little prior training. However comments showed that the Intute interface, particularly in the Enhanced Tagger was experienced as complex. By design the interface was cluttered because a variety of tagging features was to be tested. An operational system should have a simpler, less cluttered user interface, focusing on the key functionality and with user interaction streamlined.

Most participants in both studies claimed that they would be willing to use similar tools in real life. Because the results also showed the importance of controlled vocabulary suggestions, both at the time of tagging (indexing) and retrieval, it is recommended that social tagging be allowed in repositories and other digital collections, supported by controlled vocabularies.

7.4 Further work

Further studies would involve the following:

- Qualitatively analyze user tags, for example, whether they are different from those assigned by librarians in the Intute database, what topical facets are represented in the tags etc.
- Further investigate user tagging styles and motivation for tagging. For users as searchers/readers, the rationale is less straightforward than with authors. In some situations, where a user is part of a natural community engaged in a joint endeavour (as in the scenario suggested in the Intute study), tagging content will serve for mutual benefit. In other cases, users may be motivated to act as good (informed) citizens and tag based on their desire to share expertise or enthusiasm. One aspect of this could be a consideration of the different types of tagging activity. For example, tags might express the genre or utility of a document for a user's purposes. To the extent that others share the same perspective, non-subject based tags might serve as useful access points for others, in addition to their potential in personalising access to a collection.
- Conduct a retrieval test with appropriate performance measures of retrieval effectiveness. This is complex due to the interconnectedness of indexing, searching, motivation and evaluation.
- Conduct extended (longitudinal) evaluation in live settings to gain information of such systems in real life.
- Investigate different types of vocabularies and different domains. A different type of vocabulary, such as a thesaurus, may have different effects as a basis of tagging suggestions.

Furthermore, a number of enhancements to the functionality of the tools were proposed. There should be greater quality of automatic suggestions: they should be user-oriented as regards terminology, level of specificity, perspective and currency. The Intute demonstrator implemented a very crude form of automatic classification, in that the title of a document selected for tagging was fed through to the DDC matching system. The top ranked match yielded suggestions which automatically appeared in the suggested tag cloud, which worked remarkably well in some cases, but not in others. A more sophisticated automatic classification system would add considerable value to the quality of suggestions.

An auto-completion feature could also be provided which could include interactive term disambiguation and visualization. Some participants wanted a more graphical/visual representation of the controlled vocabulary, others asked for more information about the semantic meaning of a term in the controlled vocabulary and even other people's tags. Also, one could further study whether it is useful to structure the suggestions in a faceted check list of controlled terms.

Broadly speaking, this study has investigated the seeding of folksonomic tag clouds with controlled vocabulary elements. The resulting hybrid shows potential to increase access points to collections of information. Preliminary analysis suggests that vocabulary-based suggestions have potential to prompt taggers to escape the literal text, provided the suggestions are relevant and user oriented. This requires further analysis and work, including personalised filtering of suggestions and selection of appropriate terminology for a concept, the expansion and adaptation of existing controlled vocabularies with social tagging data and the investigation of user-oriented methodologies for the development and construction of controlled vocabularies. If the hypothesis is borne out then vocabulary-based suggestions may potentially both encourage the description of resources by more topics (increased exhaustivity) and may also afford the capability of describing resources at a higher level of generalisation (the activity of classification). In other words, vocabulary-based suggestions might encourage both indexing and classification activities.

Acknowledgements

This work is funded by UK's Joint Information Systems Committee (JISC) Capital Programme. Thanks are due to Intute for supplying an extract of their social science collection and to OCLC Research for providing the Dewey Decimal Classification.

References

- ACM Computing Classification Scheme (1998). Retrieved May 1, 2009, from <http://www.acm.org/about/class/>
- Connotea. (2008). Retrieved March 19, 2008, from <http://www.connotea.org/>.
- Del.icio.us. (2008). Retrieved March 19, 2008, from <http://del.icio.us/>.
- Entity Describer. (2007). Retrieved March 19, 2008, from <http://www.connotea.org/wiki/EntityDescriber> .
- Flickr. (2008). Retrieved March 19, 2008, from <http://www.flickr.com/>.
- Golub, K. Jones, C., Matthews, B., Puzon, B., Nielsen, M. L., Moon, J., Tudhope, D. (2009). EnTag: Enhancing Social Tagging for Discovery. To appear, Joint Conference on Digital Libraries, JCDL 2009, Austin, Texas, June 15-19 2009.
- Guy, M. and Tonkin, E. (2006). Folksonomies: Tidying up tags? *D-Lib Magazine*, January 2006. Retrieved March 19, 2008, <http://www.dlib.org/dlib/january06/guy/01guy.html> .
- Humanities and Social Science Electronic Thesaurus (HASSET). (2009) Retrieved May 1, 2009, from <http://www.data-archive.ac.uk/search/hassetAbout.asp>
- Hayman, S. (2007). Folksonomies and tagging: New developments in social bookmarking. *Ark Group Conference: Developing and Improving Classification Schemes 27-29 June, Rydges World Square, Sydney*. Retrieved March 19, 2008, from <http://www.educationau.edu.au/jahia/webdav/site/myjahiasite/shared/papers/arkhayman.pdf> .
- International Bibliography of the Social Sciences Thesaurus (2009). Retrieved May 1, 2009, from <http://www.lse.ac.uk/collections/IBSS/about/thesaurus.htm>
- Intute. (2009) . Retrieved May 1, 2009, from <http://www.intute.ac.uk/>
- Noruzi, A. (2007). Folksonomies: Why do we need controlled vocabulary? Editorial. *Webology*, Vol. 4, No. 2. Retrieved March 19, 2008, from <http://www.webology.ir/2007/v4n2/editorial12.html> .
- Raymond, M. (2008). My friend Flickr: A match made in photo heaven. *Library of Congress Blog*. Retrieved March 19, 2008, from <http://www.loc.gov/blog/?p=233> .
- Simple Knowledge Organization System (SKOS) (2009). Retrieved May 1, 2009, from <http://www.w3.org/2004/02/skos/>
- Smith, T. (2007). Cataloging and you: Measuring the efficacy of a folksonomy for subject analysis . In Lussky, Joan, Eds. *Proceedings 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*, Milwaukee, Wisconsin. Retrieved March 19, 2008, from <http://dlist.sir.arizona.edu/2061/> .
- Spiteri, L. F. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*, Volume 4, Number 2, June, 2007. <http://www.webology.ir/2007/v4n2/a41.html>
- STFC ePublication Archive (ePubs). (2008). Retrieved May 1, 2000, from <http://epubs.stfc.ac.uk/>
- Tonkin, E., Corrado, E. M., Moulasion, H. L., Kipp, M. E. I., Resmini, A., Pfeiffer, H. D., Zhang, Q. (2008). Collaborative and social tagging networks. *Ariadne* 54. Retrieved March 19, 2008, from <http://www.ariadne.ac.uk/issue54/tonkin-et-al/> .

This paper is a pre-print version presented at the ISKO UK 2009 conference, 22-23 June, prior to peer review and editing. For published proceedings see special issue of Aslib Proceedings journal.

Weller, K. (2007). Folksonomies and ontologies: two new players in indexing and knowledge organization. *Online Information 2007*, 108-115.