http://strathprints.strath.ac.uk/

Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool

George Macgregor & Emma McCulloch
Centre for Digital Library Research,
Department of Computer & Information Sciences,
University of Strathclyde

Abstract

Purpose – *The purpose of the paper is to provide an overview of the collaborative tagging phenomenon and explore some of the reasons for its emergence. The paper reviews the related literature and discusses some of the problems associated with, and the potential of, collaborative tagging approaches for knowledge organisation and general resource discovery.*

Design/methodology/approach – *A definition of controlled vocabularies is proposed and used to assess the efficacy of collaborative tagging. An exposition of the collaborative tagging model is provided and a review of the major contributions to the tagging literature is presented.*

Findings – *There are numerous difficulties with collaborative tagging systems (e.g. low precision, lack of collocation, etc.) that originate from the absence of properties that characterise controlled vocabularies. However, such systems can not be dismissed. Librarians and information professionals have lessons to learn from the interactive and social aspects exemplified by collaborative tagging systems, as well as their success in engaging users with information management. The future co-existence of controlled vocabularies and collaborative tagging is predicted, with each appropriate for use within distinct information contexts: formal and informal.*

Research limitations/implications – *Librarians and information professional researchers should be taking a lead role in research aimed at assessing the efficacy of collaborative tagging in relation to information storage, organisation, and retrieval, and to influence the future development of collaborative tagging systems.*

Practical implications – *The paper indicates clear areas where digital libraries and repositories could innovate in order to better engage users with information.*

Originality/value – *At time of writing there were no literature reviews summarising the main contributions to the collaborative tagging research or debate.*

Keywords Classification, Collaborative Tagging, Controlled languages, Information management, Information retrieval, Knowledge organisation

Paper type General review

Introduction

Metadata aids the identification, description, management and location of information resources in both digital and non-digital environments. Within the digital environment, the use of metadata to enhance resource discovery continues to be indispensable, particularly within specific communities of practice such as digital libraries or repositories. Metadata can enhance the process of resource discovery by disclosing sufficient information about a resource to enable users or intelligent agents to discriminate between what is relevant and what is irrelevant to a specific information need. Metadata also facilitates approaches to searching or browsing that are simply unfeasible using existing post-coordinate systems (Dawson, 2004). For example, it enables the location of resources on the subject of 'Adam Smith', rather than those written by 'Adam Smith'.

To facilitate retrieval by subject, information resources are manually assigned subject headings according to their content or, to use cataloguing parlance, 'aboutness'. Such subject descriptors are commonly known as *index terms* and these are derived from a larger set of index terms known as an *indexing language*. An indexing language constitutes a defined set of terms (or classes) utilising established conventions for ordering and combining terms. The order and arrangement of these terms affect the specificity and exhaustivity of the indexing language. Thus, terms assigned to resources that are exhaustive will result in high recall at the expense of precision. Conversely, terms that are too specific will result in high precision, but lower recall (Maltby, 1975). To ensure effective indexing and to maintain the overall efficacy of the retrieval system, it is necessary to apply some degree of control to the indexing process. By controlling the indexing process using a so-called *controlled vocabulary*, index terms are standardised and similar or related resources are collocated for ease of discovery by the user (Lancaster, 1972).

Although they yield many benefits, the pre-eminence of controlled vocabularies has recently been challenged by the appearance of 'collaborative tagging' in a variety of prominent Web-based services (del.icio.us: http://del.icio.us/, CiteULike: http://www.citeulike.org/, Flickr: http://www.flickr.com/, etc.). Collaborative tagging has emerged as a means of organising information resources on the Web and is contradictory to the ethos of controlled vocabularies. The use of controlled vocabularies - in conjunction with the wider activity of 'high quality' metadata creation (i.e. cataloguing) - remains a skilled process normally undertaken by highly trained information professionals. By contrast, collaborative tagging permits any user to assign keywords (or 'tags') to Web content (Golder and Huberman, 2005). The purpose of this brief paper is therefore to provide an overview of the collaborative tagging phenomenon, why it has arisen, the emerging literature, and to highlight the problems and the potential of such approaches for knowledge organisation and general resource discovery. Since many of the difficulties associated with collaborative tagging can only be understood via a comparative analysis with controlled vocabularies, we begin by defining the essential properties of controlled vocabularies to which we will refer later in the paper.

Defining Controlled Vocabularies

Although similar to an authority list, a controlled vocabulary differs in that it generally incorporates some form of semantic and hierarchical structure (Lancaster, 2003). This structure - and the control exerted over vocabulary - performs several functions:

- It controls the use of synonyms (and near-synonyms) by establishing a single form of the term. This ensures that indexers apply the same terms to describe the same or similar concepts, thus reducing the probability that relevant resources will be missed during a user search (Ranganathan, 1967) (e.g. 'car', 'automobile', 'motorcar', or 'motor vehicle', etc.).

- It discriminates between homonyms, allowing the indexer to resolve clashes of meaning that arise when several terms assume the same form but assume distinct meanings (e.g. 'Java' the programming language, or 'Java' the coffee, or 'Java' the island belonging to the large south east Asian archipelago of Indonesia). By controlling homonymy, the probability of noise in users' results sets is reduced (Ibid.). By virtue of eliminating homonymy, any other problems associated with homographs - where terms may assume the characteristics of homonyms, but have different pronunciation – are addressed (e.g. 'bass' the musical instrument, or 'bass' the marine fish of the family Serranidae). Terms that are spelled identically but have different meanings when pronounced differently (i.e. heteronyms) are also resolved (e.g. 'reading' the act of comprehending written or printed characters, or 'Reading' the town in Berkshire, England, UK).

- It controls lexical anomalies by minimising any superfluous vocabulary or grammatical variations that could potentially create further noise in the users' results set (Chamis, 1991; Garshol, 2004) (e.g. removing vocabulary that is superfluous to describing the intellectual content of the resource, such as leading articles, prepositions, conjunctions, etc., or ensuring consistency in spelling variants, singular and plural forms, verb tenses, and other grammatical variations).

- As noted above, it unites similar terms, or systematically refers the indexer to closely related alternatives, in order to ensure that similar or related resources are collocated. This is normally achieved by displaying the 'genus/species' relationship between terms within some form of semantic hierarchical structure, thus indicating when a subordinate class is a species of the super-ordinate class within which it is hierarchically nested (Maltby, 1975) (e.g. 'Leninism' is a species of 'communism', which in turn, is a species of 'political ideology').

- Where appropriate, syntactic relationships (i.e. non-hierarchical relationships) are accommodated (e.g. 'language' is syntactically related to 'indexing', even though they are not strictly hierarchically related. That is, the relationship between 'language' and 'indexing' only arises when a compound class of 'indexing language' is created).

- The structure also facilitates the use of codes or notation which can then be associated with terms. Such notation is mnemonic, predictable, and language independent (Broughton, 2004). In the physical environment, such notation also assists in the filing, storage and organisation of resources in libraries or information centres (Vickery, 1971).

Lancaster (2003) identifies and defines three major manifestations of controlled vocabulary: bibliographic classification schemes, subject heading lists and thesauri.

The Controlled Vocabulary 'Problem'

Traditional classification methods have long been employed in online services. The BUBL Information Service (http://bubl.ac.uk/) organises its content according to Dewey Decimal Classification (DDC). Renardus (http://www.renardus.org/) employs DDC to help users navigate selected multilingual subject gateways, demonstrating that standard schemes such as DDC have great potential for interoperability and scalability, as well as knowledge organisation and resource discovery. Many digital library services, such as Scotland's Culture service (http://www.scotlandsculture.org/), have resources indexed using Library of Congress Subject Headings (LCSH), while Artifact (http://www.artifact.ac.uk/) employs the Art and Architecture Thesaurus (AAT). Although providing many benefits and opportunities for innovative searching or browsing and interoperability, it has long been recognised that traditional controlled vocabularies (in their various permutations) are not always adequate for online resource discovery. Mai (2004) has summarised difficulties of knowledge representation within established bibliographic classification schemes and Nicholson et al (2001) have identified factors including a lack of, or excessive, specificity in the subject areas of some controlled vocabularies as being an impediment to the adequate description of online collections within specific contexts. The need for some services to implement in-house modifications, their general dependency on significant investments of time, money, training, expertise and professional intervention further discourages their wider adoption within particular communities of practice.

The fundamental obstacle preventing wider deployment of controlled vocabularies is that the proliferation of digital libraries and the Web precedes the ability of any one authority to use traditional methods of metadata creation and indexing. While metadata creation is valuable and indispensable within particular communities of practice, it can be costly to implement and can present significant scaling difficulties (Duval et al., 2002). Advances in research of automatic metadata generation applications is increasing (Greenberg, 2004) and indicates that issues of scaling, efficiency and cost can potentially be ameliorated. It is purported by some researchers that such gains in efficiency, were they to be achieved, would allow information professionals to dedicate their efforts on those intellectually demanding metadata activities necessitating some form of human mediation (i.e. assigning controlled index terms) (Anderson and Perez-Carball, 2001; Greenberg et al., 2006). Until such time automatic applications are fully realised, describing or indexing the corpus of information available on the Web will remain beyond the scope of any one authority. The emergence of 'collaborative tagging' is therefore considered by some as a useful way in which to supersede the subject indexing role of the information professional and to facilitate resource discovery and knowledge organisation over the Web (Quintarelli, 2005; Shirky, 2005a).

Collaborative Tagging

'Collaborative tagging' describes a practice whereby users assign uncontrolled keywords to information resources. Such tags are used to enable the organisation of information within a personal information space, but are also shared, thus allowing the browsing and searching of tags attached to information resources by other users. It also allows users to tag their information resources with those tags that exemplify popularity. The popularity of tags is determined by their level of use and the most popular are often depicted as a 'tag cloud' (see Figure 1). Tags are generally single terms, however the assignation of multiple tags to a single resource can be accommodated by omitting essential syntax or punctuation and by using symbols to combine terms (e.g. information+management).

Figure 1 – Portion of 'tag cloud' as displayed by the collaborative tagging system, 'del.icio.us'.

The collaborative and ad hoc nature of tagging systems dictates that they lack the essential properties characterising controlled vocabularies (as defined earlier). No control is exerted in collaborative

tagging systems over synonyms or near synonyms, homonyms and homographs, and the numerous lexical anomalies that can emerge in an uncontrolled environment. The probability of noise in a user's result set is therefore very high. The corollary dictates that this impacts negatively upon retrieval precision, as well as limiting the ability to collocate similar or related resources. The inconsistent and ambiguous assignation of tags, and the user proclivity towards exhaustive tags (e.g. 'marketing', 'technology', etc.), popular tags and personal tags (e.g. 'me', 'toread', etc.) further compromises precision and contributes to high levels of recall and noise also.

Some of the most prominent services incorporating tagging include del.icio.us (http://del.icio.us/), a collaborative bookmarks manager that operates by inviting users to organise their 'favourites' in a collaborative environment; Flickr (http://www.flickr.com/), a Web based photograph management application; and CiteULike (http://www.citeulike.org/), a tool for managing and sharing academic papers. Each of these services boasts features geared towards simplifying the process of organising a variety of media, in addition to mechanisms facilitating the future retrieval of such items. Of the aforementioned services, del.icio.us is arguably the most developed and possibly the most collaborative. For example, it combines information gathered from unique identifiers (i.e. the URL) with information gathered about the most popular tags used for that URL. This allows del.icio.us to suggest possible tags when users are bookmarking new resources or to provide users with a list of 'common tags' (i.e. popular tags that are assigned to the same resource by multiple users). These common tags can then be used in a subsequent user search strategy. Although Flickr is often discussed as part of the tagging phenomenon, the discrete nature of uploaded objects prohibits such a 'close knit society' and thus 'collaborative tagging' - as distinct from 'tagging'- is not made possible.

Collaborative Tagging for Knowledge Organisation and Resource Discovery: Debate and Research

Several authors have documented their thoughts on collaborative tagging but few have done so via the scholarly literature. Discussion of collaborative tagging has instead been most active within the Web blogging community. Vander Wal (2005) and Mathes (2004) have discussed the potential benefits of tagging (as opposed to collaborative tagging) for personal information management (PIM). Vander Wal (2005) has observed that in tagging systems there exists a powerful PIM tool, allowing users to index their information resources with their own vocabulary. Tagging for PIM, however, has inspired far less debate since the benefits for users – although yet to be empirically tested – is quite palpable, understandable, and is not dissimilar to that of file naming or email filtering. Debate has thus concentrated on the use of collaborative tagging for general resource discovery and knowledge organisation on the Web, much of which has been abstract in nature.

Recent Debate

Shirky (2005a; 2005b; 2005c) has suggested that the emergence of collaborative tagging on the Web is a 'forced move' and hypothesises that tagging will soon supersede controlled vocabularies for the purposes of resource discovery and knowledge organisation. In support of this hypothesis, Shirky (2005a; 2005c) posits the 'exclusive' nature of existing controlled vocabularies as impeding their overall usability and suggests that current schemes are incapable of reflecting the transient nature of knowledge and therefore the demands of the modern information user. Shirky suggests that collaborative tagging is inclusive; there is no vocabulary authority imposing a controlled top-down view of knowledge. All users can participate and contribute their own personal vocabularies to generate a collaboratively built 'bottom-up' vocabulary which more accurately reflects users' conceptual model of the world around them. The perceived economic advantages of collaborative tagging have also been noted by Shirky (2005a; 2005c). He suggests that the economic advantages will further entrench the practice and make it the preferred strategy for service providers and users in the future. The potential cost reductions available by encouraging communities to undertake indexing themselves, as opposed to relying on professional intervention, undoubtedly contributes to the appeal of collaborative tagging and this particular argument has also been forwarded by other commentators (Quintarelli, 2005; Sterling, 2005).

Davis (2005) has explicitly questioned the economies that can be achieved using tagging. He has argued that any economies achieved in indexing or classifying resources are simply moved onto the price of resource discovery for users, since the lack of collocation increases the number of locations that users have to explore before satisfying their information need. Davis states that the historical purpose of controlled vocabularies has not altered and notes that high costs have always been incurred

by a very small number of information professionals in order to reduce the discovery costs for a large number of users. Merholz (2005) has elucidated by providing anecdotal examples from the online reference management service, Connotea (http://www.connotea.org/). Merholz reveals that a query on the subject of 'Avian Flu', for example, exposes twenty six terms that have been used to describe essentially the same concept.

However, the issue of collocation is considered unimportant by Shirky (2005a). He maintains that the lexical ambiguities inherent in tagging should be permitted to distend since it is through this property that a true representation of knowledge is derived. Whilst cataloguers or indexers will attempt to keep similar or related concepts together, Shirky argues that it is impossible to 'collapse' such terms without loosing the essence of what each term conceptually denotes. He therefore states that it is impossible to disentangle terms such as 'queer', 'gay' or 'homosexual' since their meanings are very distinctive and collapsing them together is to misunderstand their conceptual properties. However, Shirky does not discuss how such an approach would scale or impact upon general resource discovery by subject.

Mathes (2004) and Quintarelli (2005) have argued that collaborative tagging can prove beneficial for users' search strategies, providing an increased number of entry points and a measure of serendipity unattainable using controlled vocabularies. Mathes postulates that the serendipitous nature of collaborative tagging, although not necessarily conducive to known-item retrieval or goal-directed browsing, complements *non*-goal-directed searching and browsing by introducing the user to potentially invaluable resources that would otherwise have been undiscoverable. Mathes concludes however that proving or disproving such a hypothesis would require exhaustive large scale qualitative and ethnographic end-user research.

The cognitive processes experienced by users of a collaborative tagging system have been explored by Sinha (2005). She argues that collaborative tagging utilises existing cognitive processes without adding to the cognitive load experienced by the user. She proposes a rudimentary cognitive model of the tagging process and highlights the ability of immediate tagging feedback to circumvent the condition of so-called **'**post activation analysis paralysis'. According to Sinha, such a condition places the user in a state of cognitive paralysis and is triggered when he/she attempts to tag an information resource to ensure future re-findability. Sinha suggests that collaborative tagging reduces the cognitive load experienced by the user because the intellectually onerous task of deciding how a particular resource should be tagged is removed by using system feedback and by observing how others have tagged similar items. Sinha's hypothesis and conclusions remain untested.

Collaborative Tagging Research

In one of the few research studies to date, Golder and Huberman (2005) analysed data gathered from del.icio.us to better understand the structure of tagging systems, such as user activity, tag frequencies, the nature of tags used, and so forth. They found that the users of collaborative tagging systems exhibited much variety in the sets of tags they employ. The frequency of tag use and what the tags themselves described was also found to vary greatly between users. However, the data also suggested that there existed some measure of regularity in the tags being assigned by users. On this basis, Golder and Huberman proposed a 'dynamical model' of collaborative tagging in which it is possible to predict stable tagging patterns. Their proposed hypothesis remains untested.

Finally, Guy and Tonkin (2006) conducted a small-scale study to assess the 'tag literacy' of users and suggest how such literacy might impact upon the utility of the tagging approach. Their study involved the analysis of randomly sampled tags from Flickr and del.icio.us. Guy and Tonkin found that 40% and 28% of tags were erroneous in Flickr and del.icio.us respectively. That is, tags were either mis-spelt, from a language not included in their multilingual dictionary software, in a form that that the dictionary could not decode, or were composed of multiple words or a combination of languages. They also found 8% of Flickr tags and 11% of del.icio.us tags to be plural forms and that there existed clear evidence of users deploying the use of various symbols (such as #) at the beginning of tags to influence system filing. Guy and Tonkin consequently propose various system specific strategies for improving the quality of tags (e.g. spelling error checking, suggestion of synonyms, etc.) and encouraging users to observe certain collaborative tagging conventions.

Conclusion: Future Research and the Future of Collaborative Tagging

Clearly there are numerous difficulties with collaborative tagging, which many proponents have recently been forced to acknowledge. As noted here, most of these difficulties (e.g. low precision, lack of collocation, etc.) originate from the absence of those properties that have come to characterise controlled vocabularies. Commentators on collaborative tagging, such as Quintarelli (2005), consider precision to be unimportant; however it remains difficult to accept that such systems – as an instrument of general resource discovery – will scale and sustain user confidence over the long-term unless they can demonstrate otherwise.

Given some of these basic inadequacies, it is easy to appreciate why tagging has been derided or largely ignored by the LIS community; there appear to be too many irreconcilable problems inherent in the 'mass indexing' ethos to envisage it ever superseding more established methods of indexing for knowledge organisation and general resource discovery. Be that as it may, collaborative tagging systems allow users to participate in exciting, highly interactive services and they demonstrate a possible role for users in knowledge organisation and the construction of controlled vocabularies for general resource discovery. There is now momentum behind the development and application of collaborative tagging systems - as the recent acquisition of Flickr and del.icio.us by Yahoo! perhaps demonstrates - and it is quite possible that automated techniques will be deployed to 'clean up' and mitigate some of the aforementioned difficulties. However, collaborative tagging systems capable of truly interpreting 'the linear unwinding of language' (Foucault, 1977) – as controlled vocabularies and taxonomic classifications do – should not be expected within the foreseeable future.

It is curious to note that during the period in which collaborative tagging has emerged, a reaffirmation of controlled vocabularies has arisen in parallel. The requirement for improved information organisation and management within the corporate sector has facilitated the increased deployment and development of corporate taxonomies (Cruz, 2004; Delphi Group, 2004; Kremer et al, 2005). Similarly, the need for improved subject interoperability within and outside the burgeoning number of distributed digital libraries and digital repositories has also been drawn into sharp focus (McCulloch, 2004; Zeng & Chan, 2004). The need for lexical control, hierarchical structure and associated coding is essential for attaining meaningful subject interoperability across distributed systems (perhaps using different dialects or languages), as well as maintaining the efficacy of subject searching on local systems. To this end librarians and information professionals should be more proactive in extolling the benefits of controlled vocabularies and dispelling the view that controlled vocabularies are inherently non-user friendly. Controlled vocabularies (or taxonomies) are information tools; a means to an end. For a tool to be useful one has to understand how it operates in order to take advantage of what it offers; such an investment could be considered a 'one-off' cost, after which the cost (e.g. time, effort, etc.) of discovery declines. Conversely, collaborative tagging harbours few rules and therefore its use as a 'tool' can be quite limiting in particular contexts. In stark contrast to a controlled vocabulary, an information literacy session with tagging will not enable a user to improve his/her chances of discovering relevant resources and satisfy an information need since the rules of discovery are not sufficiently predictable nor are they learnable. It therefore becomes a choice between a 'perpetual discovery cost' and a 'one-off cost'.

Equally, collaborative tagging can not be entirely dismissed by librarians or information professionals in the manner that tagging proponents dismiss controlled vocabularies. There are positive lessons to be learned from the interactivity and social aspects exemplified by collaborative tagging systems. Even if their utility for high precision information retrieval is minimal, they succeed in engaging users with information and online communities, and prove useful within PIM contexts. The need to engage users in the development of controlled vocabularies has been recognised by vocabulary experts (Abbott, 2004; Mai, 2004) and collaborative tagging systems could potentially provide a base model for such approaches. Ultimately the dichotomous co-existence of controlled vocabularies and collaborative tagging systems will emerge; with each appropriate for use within distinct information contexts: formal (e.g. academic tasks, industrial research, corporate knowledge management, etc.) and informal (e.g. recreational research, PIM, exploring exhaustive subject areas prior to formal exploration, etc.).

It is nevertheless clear that the specific factors likely to influence the efficacy of social tagging for resource discovery or knowledge organisation have been ignored and further theoretical analyses are required to facilitate – and to provide focus to – valid and testable hypotheses, and future applied research or enquiry. In particular, the literature to date has focused on the ideological merits (or otherwise) of social tagging, with little attempt being made to understand the theoretical practicalities. This lack of conceptual progress has consequently manifested itself in a lack of testable conceptual models and empirical studies. Librarians and information science researchers - with knowledge of the

issues and practicalities surrounding information retrieval with controlled and uncontrolled vocabularies - should therefore being taking a lead role in conducting meaningful research to assess the true value of collaborative tagging in relation to information storage, organisation, and retrieval, and to influence the future development of collaborative tagging systems.

Acknowledgements

References

Abbott, R. (2004), "Subjectivity as a concern for information science: a Popperian perspective", Journal of Information Science, Vol. 30, No. 1, pp.95-106.

Anderson, J, D. & Perez-Carball, J. (2001), "The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing", Information Processing & Management, Vol. 37 No. 2, pp.231-254.

Broughton, V. (2004) Essential Classification, Facet Publishing, London.

Chamis, A, Y. (1991), Vocabulary Control and Search Strategies in Online Searching, Greenwood Press, Connecticut.

Cruz, B. (2004), "Corporate taxonomies can open up the big picture", Handbook of Business Strategy, Vol. 5 No. 1, pp.247-251.

Dawson, A. (2004), "Creating metadata that works for digital libraries and Google", Library Review, Vol. 53 No. 7, pp.347-350.

Davis, I. (2005), "Why Tagging is Expensive", Silkworm Blog, available at: http://silkworm.talis.com/blog/archives/2005/09/why_tagging_is.html (accessed 20 February 2006)

Delphi Group. (2004), Information Intelligence: Content Classification and the Enterprise Taxonomy Practice, Delphi Group, Boston, USA, available at: http://www.delphigroup.com/research/whitepapers/20040601-taxonomy-WP.pdf (accessed 20 February 2006)

Duval, E., Hodgins, W., Sutton, S. & Weibel, S, L. (2002), "Metadata Principles and Practicalities", D-Lib Magazine, Vol. 8 No. 4, available at: http://www.dlib.org/dlib/april02/weibel/04weibel.html (accessed 20 February 2006)

Foucault, M. (1977), The Order of Things: An Archaeology of the Human Sciences, Tavistock Publications, London.

Garshol, L, M. (2004), "Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all", Journal of Information Science, Vol. 30 No. 4, pp.378-391.

Golder, S, A. & Huberman, B, A. (2005), The Structure of Collaborative Tagging Systems, Information Dynamics Lab: HP Labs, Palo Alto, USA, available at: http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf (accessed 20 February 2006)

Greenberg, J. (2004) "Metadata extraction and harvesting: a comparison of two automatic metadata generation applications", Journal of Internet Cataloging, Vol. 6 No. 4, pp.59-82.

Greenberg, J., Spurgin, K. & Crystal, A. (2006), "Functionalities for automatic metadata generation applications: a survey of metadata experts' opinions", International Journal of Metadata, Semantics and Ontologies, Vol. 1 No. 1, pp.3-20.

Guy, M. & Tonkin, E. (2006), "Folksonomies: Tidying up Tags?" D-Lib Magazine, Vol. 12 No. 1, available at: http://www.dlib.org/dlib/january06/guy/01guy/html (accessed 20 February 2006)

Kremer, S., Kolbe, L, M. & Brenner, W. (2005), "Towards a procedure model in terminology management", Journal of Documentation, Vol. 61 No. 2, pp.281-295.

Lancaster, F, W. (1979), Information Retrieval Systems: Characteristics, Testing and Evaluation (2nd Ed.), John Wiley & Sons, Chichester, UK.

Lancaster, F, W. (2003), Indexing and Abstracting in Theory and Practice (3rd Ed.), Thomson-Shore Inc., Michigan USA.

Mai, J.-E. (2004) "Classification in Context: Relativity, Reality, and Representation", Knowledge Organization, Vol. 31 No. 1, pp.39-48.

Maltby, A. (1975), Sayers' Manual of Classification for Librarians (5th Ed.), Andre Deutsch, London.

Mathes, A. (2004), "Folksonomies – Cooperative Classification and Communication Through Shared Metadata", Adam Mathes.com, USA, available at: http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf (accessed 20 February 2006)

McCulloch, E. (2004), "Multiple terminologies: an obstacle to information retrieval", Library Review, Vol. 53 No. 6, pp.297-300.

Merholz, P. (2005), "Clay Shirky's Viewpoints are Overrated", Peterme.com: links, thoughts, and essays from Peter Merholz, available at: http://www.peterme.com/archives/000558.html (accessed 20 February 2006)

Nicholson, D., Neill, S., Currier, S., Will, L. Gilchrist, A., Russell, R. & Day, M. (2001), HILT: High Level Thesaurus Project – Final Report to RSLP & JISC, Centre for Digital Library Research, Glasgow, UK, available at: **http://hilt.cdlr.strath.ac.uk/Reports/Documents/HILTfinalreport.doc** (accessed 20 February 2006)

Quintarelli, E. (2005), "Folksonomies: power to the people", Proceedings of the 1st International Society for Knowledge Organization (Italy) (ISKOI), UniMIB Meeting, June 24, Milan, Italy, ISKOI, Italy, available at: http://www.iskoi.org/doc/folksonomies.htm (accessed 20 February 2006)

Ranganathan, S, R. (1967), Prolegomena to Library Classification (3rd Ed.), Asia Publishing House, London.

Shirky, C. (2005a), "Ontology is Overrated: Categories, Links and Tags", Shirky.com, New York, USA, available: http://shirky.com/writings/ontology_overrated.html (accessed 20 February 2006)

Shirky, C. (2005b), "Folksonomies are a forced move: a response to Liz", Many2Many: A group Weblog of social software, available at: http://www.corante.com/many/archives/2005/01/22/folksonomies_are_a_forced_move_a_response_to_liz.php (accessed 20 February 2006)

Shirky, C. (2005c), "Semi-Structured Meta-data has a Posse: A response to Gene Smith", You're It! A Blog on Tagging, available: http://tagsonomy.com/index.php/semi-structured-meta-data-has-a-posse-a-response-to-gene-smith/ (accessed 20 February 2006)

Sinha, R. (2005), "A cognitive analysis of tagging", Rashmi Sinha's weblog, available: http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html (accessed 20 February 2006)

Sterling, B. (2005), "Order Out of Chaos", Wired Magazine, Vol. 13 No. 4, available at: http://www.wired.com/wired/archive/13.04/view.html?pg=4 (accessed 20 February 2006)

Vander Wal, T. (2005), "Explaining and Showing Broad and Narrow Folksonomies", vanderwal.net, available at: **http://www.vanderwal.net/random/entrysel.php?blog=1635** (accessed 20 February 2006)

Vickery, B, C. (1971), Techniques of Information Retrieval, Butterworths, London.

Zeng, M. L., & Chan, L. M. (2004), "Trends and issues in establishing interoperability among knowledge organization systems", Journal of the American Society for Information Science and Technology, Vol. 55 No. 5, pp.377-395.