

Popularity over Relevance in Collaborative Tagging Systems for General Resource Discovery

George Macgregor
Centre for Digital Library Research
Department of Computer & Information Sciences
University of Strathclyde
26 Richmond St., Glasgow
UK

george.macgregor@strath.ac.uk

Emma McCulloch
Centre for Digital Library Research
Department of Computer & Information Sciences
University of Strathclyde
26 Richmond St., Glasgow
UK

e.mcculloch@strath.ac.uk

ABSTRACT

Although services offering collaborative tagging provide rich interactive experiences that exemplify a sense of community among disparate user groups, their *raison d'être* ultimately remains similar to traditional information retrieval systems. In this paper we will explore potential conflict between user requirements for PIM and CIM, and the theme of popularity over relevance by revisiting basic information retrieval principles within the context of collaborative tagging. The long-term implications of these factors for the efficacy of collaborative tagging in general resource discovery will also be examined. The theoretical analysis presented suggests that PIM requirements conflict with the wider requirements of resource discovery by invoking a series of extra relevance variables and that emphasis on popularly used tags and the subsequent effect on 'collective behavior' potentially compromises the effectiveness of system performance. Hypotheses are proposed for testing and further research areas are suggested.

Categories and Subject Descriptors

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing – *Abstracting methods, Dictionaries, Indexing methods, Linguistic processing, Thesauruses*

General Terms

Performance, Standardization, Theory

Keywords

Collaborative tagging, Resource discovery, Information retrieval, Knowledge organization, Controlled vocabularies, PIM noise

1. INTRODUCTION

In the majority of information need scenarios, users require information pertaining to a particular subject [2, 3]. This requirement is traditionally facilitated by the assignment of controlled subject index terms to resources within a metadata structure. Since the corpus of information on the Web that requires describing or classifying will remain beyond the scope of any one authority, collaborative tagging is deemed by some as a useful way in which to supplant the subject indexing role of the information professional [11, 12].

Services offering collaborative tagging provide rich interactive experiences that exemplify a sense of community among disparate user groups, yet their *raison d'être* remains similar to that of traditional retrieval systems based on controlled indexing languages: to enhance information retrieval effectiveness (and therefore resource discovery) and to facilitate knowledge organization. However, the long-term ability of collaborative tagging to fulfill the basic premise of an information retrieval system, to provide relevant results – and thereby enhance resource discovery – is currently compromised. Such a compromise is imposed by system emphasis on tag popularity (as opposed to relevance) and the inherent conflict between Personal Information Management (PIM) and Collaborative Information Management (CIM).

In this paper we will explore the theme of PIM noise and 'popularity vs. relevance' by revisiting some basic information retrieval principles within the context of collaborative tagging and consider whether these can be reconciled with current tagging functionality to provide a suitable architecture for general resource discovery and knowledge organization.

2. RELEVANCE IN RETRIEVAL

2.1 Metadata and Information Need

Metadata can improve the resource discovery process by imparting sufficient information about an object to enable users or intelligent agents to differentiate between what is relevant to their needs and what is not. For electronic information services not based on automatic indexing, the use of metadata to underpin resource discovery remains essential, particularly within specific communities of practice (e.g. digital libraries, digital repositories, corporate intranets, government information services, etc.). Descriptive metadata fields such as Author, Title, Subject, etc. permit searching or browsing in a manner that is unattainable using existing post-coordinate systems [2, 5]. It also facilitates 'known item retrieval' user search strategies, since characteristics of the resource (e.g. title, author, publisher, date, format, etc.) will be explicitly recorded within the object surrogate.

Belkin [3] has noted that when a user feels compelled to use an information retrieval system it is because they are experiencing a gap in their knowledge. More often than not, this gap in their knowledge can not be satisfied by known item retrieval strategies. Instead, the user relies upon *unknown* item retrieval strategies that involve searching or browsing for relevant subject information, irrespective of who may have authored the information or published it.

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 a *controlled indexing language*. An indexing language constitutes a defined set of terms (or classes) utilizing established conventions for ordering and combining subject terminology. This degree of control assists in managing synonyms (or near synonyms), homonyms and homographs, and lexical anomalies such as superfluous vocabulary or grammatical variations. These aforementioned features, combined with a semantic hierarchical structure, ensure that index terms are assigned consistently and similar or related resources are collocated for ease of discovery by the user.

The order and arrangement of these terms affect the specificity and exhaustivity of the indexing language, and directly affect the ability of the information retrieval system to return relevant results. 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. The purpose of the indexer is therefore to ensure a balance is struck between specificity and exhaustivity so that the retrieval of relevant documents (or document surrogates) is optimized. (Note - ‘Documents’ is used throughout this paper to refer to all information resources or resource surrogates e.g. metadata).

2.2 Collaborative Tagging and Popularity

As acknowledged in the literature [8, 11, 12], ‘collaborative tagging’ has recently emerged as a popular tool for information retrieval, resource discovery, knowledge organization, and PIM, and contrasts with systems based on controlled indexing. Collaborative tagging describes a process whereby users assign uncontrolled keywords (tags) to resources. These tags are used to assist in the organization of information within a personal workspace, but are also made communal and are shared among other service users. This facilitates the browsing or – in some services – the searching of tags attached to information resources by others. It also allows users to tag their information resources with those tags that demonstrate popularity within a particular tagging community.

A guide to the popularity of tags is indicated by the way in which users have applied them to date and is often depicted as a ‘tag cloud’ (see Figure 1). Tags are frequently required to be single terms, although multiple tags can be assigned to a single resource by omitting spaces or punctuation and by using symbols to combine terms (e.g. collaborative+tagging). Terms are treated ‘equally’; that is, there is no hierarchical structure indicating the precise nesting of sub-terms or semantic relationships, nor is there any measure of the relevance of tags to resources (e.g. weighting). Essential properties typically characterizing controlled indexing languages are also absent (i.e. no synonym or near synonym control, no homonym or homograph control, no control over lexical anomalies, no semantic structure or notation, no treatment for compound subjects).

Vander Wal [14] has noted that tagging provides a powerful tool for PIM, empowering users to index their personal information resources with their preferred vocabulary. The use of tagging for PIM, however, has stimulated less debate since the benefits within such user contexts – although not yet empirically tested – are easily justifiable and understandable, due to parallels with file naming conventions, email filtering, and so forth. Discussion has therefore focused on the use of collaborative tagging for general

resource discovery and knowledge organization on the Web; the latter manifesting itself as a form of CIM.



Figure 1. Portion of ‘tag cloud’ as displayed by del.icio.us

3. BASIS FOR EFFECTIVE RETRIEVAL

3.1 Basic Retrieval Requirements

Little empirical research has been conducted to assess the value of collaborative tagging in relation to information storage, organization, and retrieval, meaning that their potential to supersede more traditional methods of information management remains unclear. Nevertheless, the ability of a system based on collaborative tagging to fulfill the basic architectural requirements of an information retrieval system is limited and many of the fundamentals relating to successful information retrieval have largely been ignored by tagging systems.

Lancaster [9] proposes various criteria for evaluating the effectiveness of retrieval systems and these remain valid today. Lancaster notes that *any* information retrieval operation can be evaluated in terms of (1) quality, (2) effort, and (3) response time.

(1) *Quality* entails four factors: (a) the coverage of the database; (b) the capacity of the system to retrieve relevant documents in response to a subject based query (i.e. recall); (c) the capacity of the system to simultaneously suppress the retrieval of non-relevant documents (i.e. precision), and; (d) the degree to which the form of results output is expedient for use. (2) *Effort* denotes the degree of effort required by the user to use the system. (3) *Response time* is simply the length of time taken for the system to respond to a user query.

Lancaster acknowledges that some of the above criteria are more important than others, but stipulates that all users expect the system to contribute to the satisfaction of their information need by meeting quality criteria (b) and (c). Meeting such criteria essentially dictates that the system is capable of retrieving as many relevant documents as possible.

It is worth revisiting the essential attributes of relevance here, since their applicability to current tagging systems is pertinent and such an exposition will later assist when we attempt to understand the implications of popularity over relevance and PIM on effective retrieval.

3.2 Relevance Revisited

The attainment of relevance in information retrieval demands that the system can achieve sufficient recall and precision. Recall is the capacity of the system to retrieve a large proportion of the total number of relevant documents in response to a user search query and precision the capacity of the system to simultaneously suppress the retrieval of non-relevant documents. The former can normally be expressed as the classic recall ratio (i.e. $a \div b \times 100$, in which a is the number of relevant documents retrieved and b is the total number of relevant documents stored in the system). Maximum recall can theoretically always be attained if the user employs broad search strategies and manually sifts through large result sets in order to ascertain which documents are relevant and which are not. Such strategies, however, are obviously not practicable in the majority of cases and therefore a second measure of retrieval effectiveness is required.

The latter measure (i.e. precision) represents the capacity of the system to simultaneously *suppress* the retrieval of non-relevant documents and normally can be expressed as the classic precision ratio (i.e. $a \div c \times 100$, in which a is the number of relevant documents stored in the system and b is the total number of documents retrieved). In effect then, precision measures the efficacy with which the system can attain a specific level of recall.

One branch of information retrieval research is concerned with assessing the relevance of results to queries. More rigorous systems – whether manual or automatic - typically employ a measure of the extent to which a particular search query matches documents (or document surrogates) within a given collection. Such measures may include term frequency, in order to provide the user with appropriately ranked results.

4. POPULARITY AND PIM NOISE IN COLLABORATIVE TAGGING

4.1 Tag Types

Golder and Huberman [8] have identified seven functions that tags perform within collaborative tagging systems:

- *Identifying what (or who) it is about* (i.e. tags that denote the subject of items)
- *Identifying what it is* (i.e. using a tag to describe the manifestation of the information entity. For example, ‘blog’, ‘e-book’, etc.)
- *Identifying who owns it* (i.e. instances where tags impart information about the creator of the tagged item)
- *Refining categories* (i.e. where tags are used to qualify existing categories)
- *Identifying qualities or characteristics* (i.e. using tags to denote user experience with adjectives. For example, ‘funny’, ‘stupid’, etc.)
- *Self reference* (i.e. the use of ownership tags. For example, ‘myarticles’, ‘myphotos’, etc.)
- *Task organization* (i.e. tags assigned to information entities to assist in the completion of some task. For example, ‘toread’, ‘jobsearch’, etc.)

If it is assumed – as noted in section 2.1 – that most users’ information needs are satisfied by assuming an unknown item retrieval strategy based upon subject characteristics, then the retrieval system in question has to be sympathetic to such search strategies. Collaborative tagging systems are not. The reasons for this are twofold:

- The inherent conflict between the requirements of PIM and CIM.
- Popularity preceding relevance in tagging behavior.

Lack of controlled vocabulary properties (i.e. no synonym or near synonym control, no homonym or homograph control, no control over lexical anomalies, no semantic structure or notation, no established treatment of compound subjects) has numerous implications also, but for the purposes of this paper we have chosen not to discuss them because – although their effects have been insufficiently researched within the context of collaborative tagging – they have at least been acknowledged by some in the literature as an impediment to retrieval [8, 11].

4.2 PIM Noise

As can be discerned from the seven tag functions identified by Golder and Huberman, only the first tag function could be considered to provide the basic subject access needed to facilitate an unknown item user search strategy and the satisfaction of most information needs. The six remaining tag types could be considered extraneous since their function is *primarily* to enable PIM and not *necessarily* CIM.

Such PIM requirements conflict with the wider requirements of resource discovery by invoking a series of extra relevance variables hitherto absent in conventional information retrieval systems and which collaborative tagging systems are unable to decode (e.g. resources might be indexed by ‘identifying what it is’ or by ‘qualities or characteristics’, rather than by the intellectual content of the resource).

For the majority of users adopting unknown item search strategies, such extra variables are likely to create noise in users’ result sets. Since many documents will be indexed (tagged) within PIM contexts and because the system is unable to decode the variety of tag types in use, the precision of the system to suppress non-relevant documents for general resource discovery will decline and the ability of the system to retrieve relevant documents will decrease (resulting in misses). This sub-optimal performance is partly attributable to the inherent conflict that arises when the needs of PIM (i.e. the individual) are merged with the need for general resource discovery within CIM (i.e. the community).

Although the needs of CIM theoretically conflict with PIM also, this conflict is less likely to influence the management of personal information because the needs of CIM merely pertain to one tag type (i.e. those tags that characterize the intellectual and subject content of a resource). Additionally, the deployment and context in which information is consumed or manipulated within a personal information space directly influences the way in which it is organized within that space [7]. Contextual factors and information re-use or re-access factors also influence the management of information within personal information spaces and consequently distances user practices found in PIM with those found within CIM and general resource discovery [6].

4.3 Popularity Over Relevance

The unsympathetic state of collaborative tagging systems for general resource discovery and unknown item search strategies is only partly attributable to PIM noise (and misses). Sub-optimal performance can be attributed to the way in which popularity precedes relevance in users' tagging behavior and the way in which collaborative tagging systems place reputation on popularly used tags.

As a tag becomes more popular it is more likely to be used again by other users [8]. Firstly, popular tags will feature within the collaborative 'tag cloud' more prominently than less popular tags. For example, the more popular a tag becomes, the larger the font size in which the tag is displayed to the user (e.g. see Figure 1). Popular tags can often appear in a 'popular tags' list, also displayed to the user. Secondly, in some services, such as del.icio.us, information gathered about the document unique identifier (i.e. the URL) is supplemented with information about the most popular tags used for that document by other users. Both these popularity measures are designed to perpetuate the continued use of that tag by other users. The former suggests a degree of reputation on popularly used tags. The latter allows del.icio.us to suggest popular tags to new users when they are tagging the same document.

Users can often find the assignment of indexes or the organization of information to be cognitively onerous [4]. Tools to assist in the application of appropriate tags therefore have to be made available. For some commentators [11], the use of popularity is considered to have positive implications since popularity demonstrates reputation and therefore agreement within a tagging community (i.e. 'wisdom of crowds').

Scholars in the discipline of sociology and social psychology [13] have noted various collective behavior traits in which susceptible individuals feel subject to collective pressure and group consensus (e.g. polarization). In such scenarios, violation of the consensus (by expressing divergent opinion or action) is suppressed and the 'contagion' is permitted to spread among the remaining collective. Individuals become confused and doubt their own judgments. This doubt fosters a latent submissive behavioral trait and results in that individual converging with the collective in order to avert a perceived negative consequence.

We hypothesize that a similar phenomenon can theoretically be said to exist within systems based on collaborative tagging. Emphasis on popularity leaves individuals open to 'system manipulation' (via tag clouds and tag prompts), whereby users are influenced by the collective and the system to index documents with tags that they would not otherwise use, or with tags that they would - in most circumstances - consider to be erroneous or inappropriate. This precludes the assumption that tags are subject to the laws of random probability [8] because the variables involved are not entirely free from bias.

The use of popularity as a tool therefore does not ensure that documents are tagged as objectively as possible. By 'objectively' we mean based on observable characteristics and without bias. With users exposed to bias and convergent behavioral practices, the probability that documents will be sufficiently tagged to expedite general resource discovery decreases. In this sense, popularity *precedes* relevance as tags may be assigned to resources irrespective of whether the user believes the tag to be a true representation of the resource in question or not.

If the aforementioned is true, then the corollary dictates that precision and recall will decline and increase respectively for those users experiencing 'classic' information needs. Documents will not be represented accurately or at sufficient levels of exhaustivity and specificity for the majority of users (Note - by 'accurately', we mean in such a way as to facilitate retrieval by the largest number of users with a classic information need. This concept of 'accuracy' encompasses the concept of 'objectivity'). Users' confidence in the system will decrease as they fail to experience the *quality* stipulated by Lancaster [9]. Additionally, users may also expend extra *effort* and experience unnecessary cognitive load in order to compensate for system failings (e.g. assuming a maximum recall approach entailing broad search strategies and the manual identification of relevant documents within large result sets).

This requirement for general resource discovery signifies the point at which the inherent conflict between PIM and CIM reaches its apex. Merging these two opposing communities of practice potentially compromises the efficacy of the system for both parties. Within PIM contexts users should be free to organize their personal information space as they see fit; if using popular tags assists them in managing this space then such an approach should not be questioned. However, the combination of PIM noise and popularity over relevance simultaneously undermines the ability of a retrieval system (as characterized by Lancaster [9]) to function effectively.

5. FUTURE RESEARCH AND CONCLUSION

A collaborative tagging system allows for highly interactive user experiences and demonstrates the prospective role users could assume in knowledge organization and the creation of controlled vocabularies for general resource discovery. It is worth noting that recent literature within the discipline of library and information sciences (LIS) has re-emphasized the need to engage users in collaboratively developed controlled vocabularies [1, 10].

In this paper we have attempted to theoretically explore some of the potential issues intrinsic to (a) the inherent conflict between the requirements of PIM and CIM (i.e. PIM vs. general resource discovery), and (b) the use of popularity strategies (i.e. popularity vs. relevance). This has been attempted by revisiting some fundamental information retrieval principals.

As suggested, PIM requirements conflict with the wider requirements of resource discovery by invoking a series of extra relevance variables (i.e. tag types as identified in the literature [8]). These variables have hitherto been absent in conventional information retrieval systems and are therefore currently incapable of being decoded. This factor compromises the effectiveness of the retrieval system for general resource discovery (as stipulated by Lancaster [9]). For the majority of users adopting 'classic' unknown item search strategies, such extra variables are likely to create noise or misses in users' result sets. In addition, the behavior of users within personal information spaces does not necessarily mirror behavior for general resource discovery, thereby giving rise to a conflict of information organization styles.

The way in which collaborative tagging systems place reputation on popularly used tags and the effect 'collective behavior' has on tagging behavior potentially compromises the effectiveness of system performance also. With users exposed to bias and

convergent behavioral practices, the probability that documents will be sufficiently tagged to expedite general resource discovery decreases, resulting in lower measures of relevance. Quintarelli [11] considers precision to be unimportant, but it is difficult to envisage such systems – as agents of general resource discovery - sustaining user confidence unless they can demonstrate otherwise.

We propose the aforementioned conflict and the emphasis on popularity over relevance as hypotheses for testing and experimental research.

Based on our theoretical exposition, further research should seek to establish whether the requirements of PIM and CIM can co-exist within collaborative tagging systems. Merging these two opposing communities of practice potentially compromises the efficacy of the system for both parties and it appears expedient to explore the growth of services that cater for each community separately. This would enable the creation of vocabularies more conducive to how information is searched, organized and deployed by users within each context. Moreover, since a number of tag types appear to be in operation, there exists an opportunity to offer profiled searching of particular tag types. It would be incumbent upon users of such systems to indicate what type of tag was being assigned during the tagging process and therefore research should establish whether this increased cognitive load is acceptable to users.

Given users' proclivity towards tag imitation, further work should also seek to explore the use and effects of popularity as a tool, and perhaps propose alternative models or tools that minimize bias within a collaborative tagging system for general resource discovery.

6. ACKNOWLEDGMENTS

Thanks are extended to del.icio.us for granting permission to use screenshots in this paper.

7. REFERENCES

- [1] Abbott, R. Subjectivity as a concern for information science: a Popperian perspective. *Journal of Information Science*, 30, 2, 95-106, 2004.
- [2] Anderson, J.D., and Perez-Carballo, J. 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 and Management*, 37, 2, 231-254, 2001.
- [3] Belkin, N.J., Oddy, R.N., and Brooks, H.M. ASK for information retrieval: Part I. background and theory. *Journal of Documentation*, 38, 2, 61-71, 1982.
- [4] Boardman, R. and Sasse, M.A. "Stuff goes into the computer and doesn't come out" A Cross-tool Study of Personal Information Management... in Proceedings of the SIGCHI conference on Human factors in computing systems (Vienna, April 2004), ACM Press, New York, 583-590.
- [5] Dawson, A. Creating metadata that works for digital libraries and Google. *Library Review*, 53, 7, 347-350, 2004.
- [6] Dumais, S., Cutrell, E., Cadiz, J.J., Jancke, G., Sarin, R., Robbins, D.C. Stuff I've seen: a system for personal information retrieval and re-use... in Proceedings of SIGIR . (Toronto, July-August 2003), ACM Press, New York, 72-79.
- [7] Elswailer, D., Ruthven, I., and Jones, C. Dealing with fragmented recollection of context in information management... in Proceedings of CIR-05, Workshop in Fifth International and Interdisciplinary Conference on Modeling and Using Context (Paris, July 2005), CEUR Workshop Proceedings, Germany. http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-151/CIR-05_4.pdf
- [8] Golder, S.A. and Huberman, B.A. The Structure of Collaborative Tagging Systems. Information Dynamics Lab, HP, 2005. <http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf>
- [9] Lancaster, F.W. *Information Retrieval Systems: Characteristics, Testing and Evaluation* (2nd Ed.). John Wiley & Sons, Chichester, UK, 1979.
- [10] Mai, J.E. Classification in Context: Relativity, Reality, and Representation. *Knowledge Organization*, 31, 1, 39-48, 2004.
- [11] Quintarelli, E. Folksonomies: power to the people... in Proceedings of the 1st International Society for Knowledge Organization (Italy) (ISKOI), UniMIB Meeting (Milan, June 2005), ISKOI. <http://www.iskoi.org/doc/folksonomies.htm>
- [12] Shirky, C. *Ontology is Overrated: Categories, Links and Tags*. Shirky.com, New York, 2005. http://shirky.com/writings/ontology_overrated.html
- [13] Turner, R.H., and Killian, L.M. *Collective Behavior* (2nd Ed.). Prentice-Hall, Inc., New Jersey, 1972.
- [14] Vander Wal, T. *Explaining and Showing Broad and Narrow Folksonomies*. vanderwal.net, 2005. <http://www.vanderwal.net/random/entrysel.php?blog=1635>