"Power tags" in information retrieval

Isabella Peters and Wolfgang G. Stock
Department for Information Science, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

Received 20 October 2008
Revised 10 January 2009
Accepted 11 February 2009

Abstract

Purpose – Many Web 2.0 services (including Library 2.0 catalogs) make use of folksonomies. The purpose of this paper is to cut off all tags in the long tail of a document-specific tag distribution. The remaining tags at the beginning of a tag distribution are considered power tags and form a new, additional search option in information retrieval systems.

Design/methodology/approach – In a theoretical approach the paper discusses document-specific tag distributions (power law and inverse-logistic shape), the development of such distributions (Yule-Simon process and shuffling theory) and introduces search tags (besides the well-known index tags) as a possibility for generating tag distributions.

Findings – Search tags are compatible with broad and narrow folksonomies and with all knowledge organization systems (e.g. classification systems and thesauri), while index tags are only applicable in broad folksonomies. Based on these findings, the paper presents a sketch of an algorithm for mining and processing power tags in information retrieval systems.

Research limitations/implications – This conceptual approach is in need of empirical evaluation in a concrete retrieval system.

Practical implications – Power tags are a new search option for retrieval systems to limit the amount of hits.

Originality/value – The paper introduces power tags as a means for enhancing the precision of search results in information retrieval systems that apply folksonomies, e.g. catalogs in Library 2.0 environments.

Keywords Information management, Tagging, Information retrieval

Paper type Conceptual paper

1. Introduction

Folksonomies have become an important user-driven approach to information indexing and retrieval (Dye, 2006; Furnas *et al.* 2006; Golder and Huberman, 2006; Gordon-Murnane, 2006; Guy and Tonkin, 2006; Kroski, 2008; Mathes, 2004; Noruzi, 2006; Peterson, 2006; Smith, 2008; Spiteri, 2006, 2007). They find application in:

- collaborative web services such as Del.icio.us, Flickr, Last.fm and YouTube (Peters and Stock, 2007, 2008);
- library catalogs, e.g., PennTags (Sweda, 2006; Allen and Winkler, 2007), or "cataloguing 2.0" (Lyons and Tappeiner, 2008; Fifarek, 2008; Coyle, 2007; Weaver, 2007);
- corporate intranets (Fichter, 2006);
- museum catalogs (Trant, 2006); and
- commercial online information suppliers, e.g., Elsevier's Engineering Village and GENIOS' WISO (Stock, 2007b).

Library Hi Tech
Vol. 28 No. 1, 2010
pp. 81-93

0737-8831
DOI 10.1108/07378831011026706

The strength of this approach is collaborative indexing; its weakness lies in information retrieval, as in most cases no relevance ranking exists (for a remarkable exception see Hotho *et al.* (2006), and their system "BibSonomy"). Recall seems to be quite good, but all systems lack precision. This article deals with one simple idea: to enhance precision by using so-called "power tags". Our research is located in the intersection of information retrieval research and informetrics, for we try to deploy empirical and theoretical informetric results on the architecture of information retrieval systems. Our paper is structured as follows: the next section describes document-specific tag-distributions and determines power tags. In the third section, we will discuss the development of tag distributions, because only "matured" distributions are able to form stable power tags. Our fourth section introduces index tags (tags which users apply to index a document) and search tags (tags which users apply to search for and to successfully find a document). The last section presents a way of processing power tags in an information retrieval system, e.g. a catalog of a Library 2.0 service.

2. Document-Specific tag distributions

"Power tags only" presents an additional, new search option. The determination of power tags depends on the distribution of tags regarding the frequency of their assignment to a digital resource. The basic assumption is that different document-specific distributions of tags may appear in folksonomies:

- an inverse power law distribution, a Lotka-like curve (for empirical evidence see Huang, 2006; Munk and Mork, 2007);
- an inverse logistic distribution (Stock, 2006); and
- other distributions.

A Lotka-like power law (Egghe, 2005) has the mathematical expression:

$$f(x) = C/x^a,$$

where C is a constant, x is the rank of the tag relative to the resource, and a is a value normally ranging from about 1 to about 2. If this assumption is valid, we see a curve with only few tags at the top of the distribution, and a "long tail" of many tags with lower ranks on the right-hand side of the curve (see Figure 1) (Tonkin, 2006). The discussions about "collective intelligence" are mainly based on this observation: the first n tags of the left hand side of the power law reflect the collective intelligence in giving meaning to the annotated resource.

The second kind of distribution, called "inverse logistic," is sketched in Figure 2 (for an example from Del.icio.us see Figure 3). The inverse logistic distribution shows a lot of relevant tags at the beginning of the curve (the "long trunk") and the known "long tail". This distribution follows the formula:

$$f(x) = e^{-C'(x-1)b},$$

where e is the Euler number, x is the rank of the tag, C' is a constant and the exponent b is approximately 3. In most cases the "long trunk" will be shorter than the "long tail." In comparison with the power law the inverse logistic distribution reflects the collective intelligence differently. The curve shows a long trunk on the left and a long

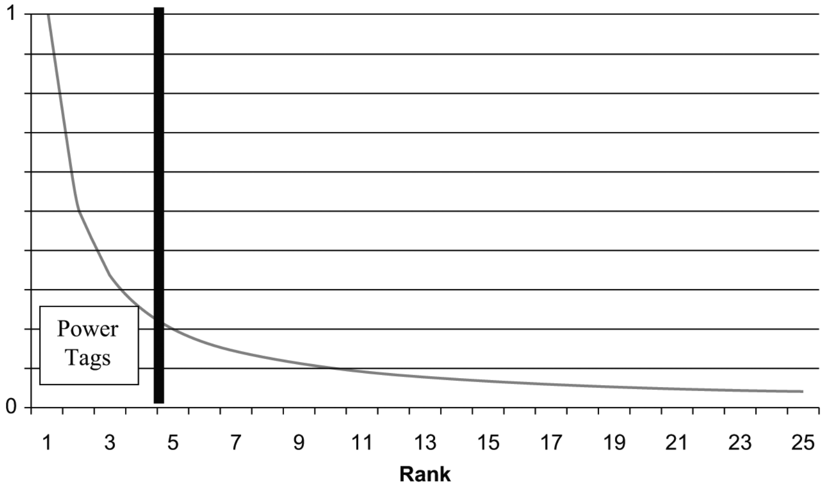

Figure 1.
An inverse power law of document-specific tags

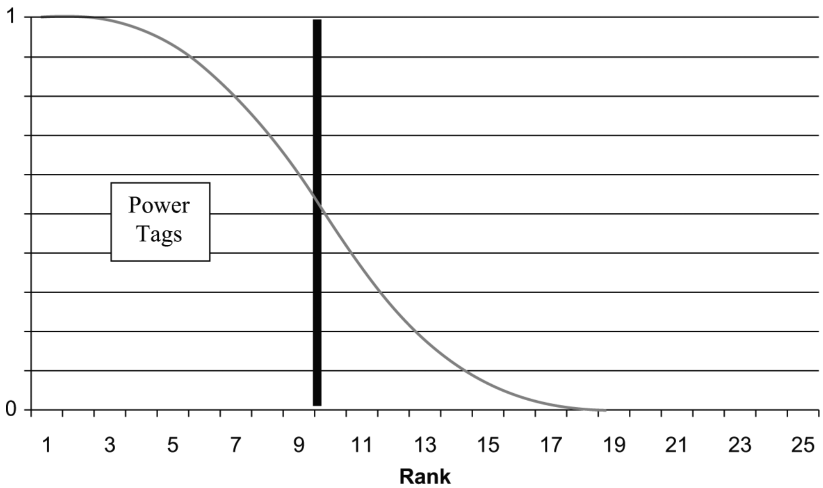

Figure 2.
An inverse logistic distribution of document-specific tags

tail of tags on the right. All left-hand tags up to the turning point of the curve should be regarded as a reflection of collective intelligence as they have been indexed with almost the same frequency. In our example (Figure 3), the first three tags – the "long trunk"-tags – have nearly the same values (rank 1: 256; rank 2: 239; rank 3: 237). The comparison of power law distributions and inverse logistic distributions shows that both share one characteristic element – the long tail – but differ fundamentally in the beginning of the curve. Research mainly draws attention to the long tail of both distributions – stating misleadingly that both are power laws. According to Newman (2005):

> [. . .] the scientist confronted with a new set of data having a broad dynamic range and a highly skewed distribution should certainly bear in mind that a power-law is only one of several possibilities for fitting it.

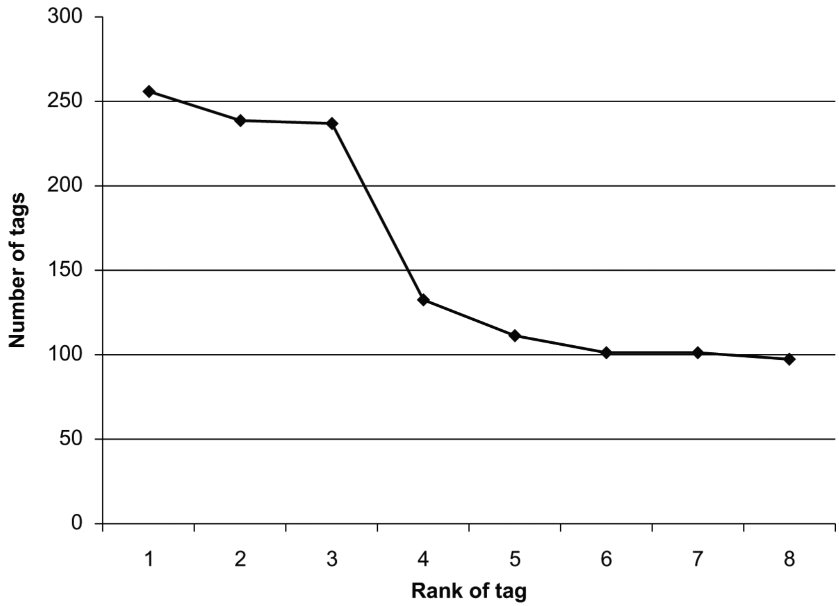

Figure 3.
Distribution of tags annotated to the web page www.startupping.com

By neglecting the main difference between the two kinds of distributions, their value for information retrieval is ignored at the same time.

Since not much research has been concerned with detecting document-specific relevance distributions within folksonomies up to now, we cannot report on other tag distributions which possibly exist.

For the determination of power tags we have to keep in mind both known tag distributions. If the document-specific distribution of tags follows the inverse power law, the first n tags are regarded as "power tags". The value of n is dependent on the exponent a: if a has a high value (say, 2 and more), n has to be very small (around 1 or 2); if a is smaller, n will be moderately higher (with $a = 1$ the value of n could be 3 or even higher). The determination of the best-fitting value of n depends on the characteristics of the concrete retrieval system, because there are wide differences between, say, a Web 2.0 service like Del.icio.us and a library 2.0 catalog like PennTags. So our proposed values are only a rough estimation that have to be empirically refined in every concrete application. If the document-specific tag distribution forms an inverse-logistic distribution, we propose to mark all tags on the left-hand side of the curve (above the turning point) as "power tags" (but here a customization on concrete systems may be necessary as well). In the end the retrieval system consists of two sets of document-specific tags, power tags and tail tags. For Thomas Vander Wal (2005) it is clear, "the power terms and the long tail both work." power tags refer to document-specific terms which are shared by a broad spectrum of people; tail tags refer to document views of minorities and are valuable for those minorities. To emphasize on tail tags, Weller and Peters (2008) suggest constructing an "inverse tag cloud" which provides an additional access point to the document collection and to single documents as well.

3. The development of tag distributions

How does a tag distribution evolve? To answer this question it is important to explain the "history" of tags of a given document, because only "matured" distributions should be used for identifying power tags. Lux *et al.* (2007) could show for the macro level (the whole folksonomy of a collaborative web service) that about 80 percent of the co-tags of analyzed tags form a power law shape. That means that about 20 percent of those tags form other shapes, including the inverse logistic distribution. Our task is to explain the development of the distributions on the micro level (the document level). So we have to transform the results from the macro level into the micro level.

It is quite evident that the development of a power law distribution follows the known "rich gets richer" or "success breeds success" assumptions (Egghe and Rousseau, 1995, 1996; Huber, 1998; Tague, 1981). Again for the macro level, Cattuto (2006) and Cattuto *et al.* (2007) apply "semiotics dynamics" to the creation of new tags and to the use of old ones when a prosumer (a conflation of "consumer" and "producer" according to Toffler, 1980) indexes a resource. They argue that the selection of tags in folksonomies is a Yule-Simon process (named after George Udny Yule and Herbert A. Simon) (Chen, 1989; Chen *et al.*, 1994; Oluić-Vuković, 1997). This approach describes a Yule-Simon process with memory (Cattuto *et al.* 2006). There is a probability p for every tag that this tag is new, and a probability of $1 - p$ that the tag is a copy of a known tag. Cattuto (2006) refines the Yule-Simon approach by including a time component:

> It seems more realistic to assume that users tend to apply recently added tags more frequently than old ones, according to a skewed memory kernel.

So the probability of $1 - p$ depends on the moment the tag was applied the last time. The shorter this time span, the higher the probability is of using the given tag again.

Another theoretical model for the development of tag distributions is the "shuffling theory" or the theory of "preferential attachment" explained by Halpin *et al.* (2007). They calculate the probability that a user chooses a tag in analogy to his behavior when drawing cards from a hat. We want to explain this theory with an example: There may be a document about music, piano, etc., which is not tagged. The first user applies piano to index the resource, while the second works with music. At this stage, p(piano) is one-half and p(music) is one-half as well. Now a user chooses piano a second time, so p(piano) is two-thirds and p(music) is one-third. When a user adds a new tag, say, digital, to the resource, the probability of the tags will change: p(piano) = one-half, p(music) = one-quarter and p(digital) = one-quarter. If users continue to add piano and always some new tags, "this process produces a power law distribution" (Halpin *et al.*, 2007). But if users tend to apply more than one tag frequently (say, piano and music in combination), the process of preferential attachment may produce an inverse logistic distribution.

If indexing users see and follow suggestions of document-specific tags, sorted by occurrence, the development of a power-law or an inverse-logistic distribution would be a kind of self-fulfilling prophecy. The user notices the suggested tag on the top of the list, agrees with this term and indexes consequently the document with the given tag, i.e. imitates this term – without thinking very much by himself. But if users do not follow recommendations (or if there is no recommendation), it is the users' background knowledge and their "active vocabulary" (Dellschaft and Staab, 2008) which determines the concrete tagging process. If tag recommendation is given (as for example, in the tagging interface of Del.icio.us), Dellschaft and Staab (2008) found that the imitation rate during tag assignment would be around 60 percent and 90 percent.

At some point in time, it is possible that the shape of the distribution of document-specific tags (not the absolute number of tags) will remain stable. If the same holds true for the concrete power tags, we can use them like controlled terms – "controlled" only by collective intelligence of the users.

But we have to be very careful at this point. Due to insufficient empirical data of the development of document-specific tags and due to just as little theories and models of the maturing of tag-distributions in time our results are preliminary. We are in need of much more empirical and theoretical research on the development of tag distributions.

4. Index tags and search tags

The properties of a folksonomy may cause some problems during the generation of power tags. The distributions of tags may only develop within broad folksonomies (Vander Wal, 2005), in which every user is allowed to add tags to every resource and as often as needed, e.g. in the social bookmarking service Del.icio.us or in the music service Last.fm. In narrow folksonomies, tags for a single resource are generally recorded just once. Thus, only new tags can be applied and there is no possibility of counting tag frequencies, e.g. in the photo sharing service Flickr or in the video service YouTube. Usually the resources' author (or content creator) provides the tags; occasionally other users are also allowed to add tags. This approach reminds one of the known procedure in professional indexing with knowledge organization systems, e.g. nomenclatures, thesauri, classification systems or ontologies (Stock and Stock, 2008) – but, the terms are not controlled in folksonomies. Accordingly, narrow folksonomies cannot depict a special distribution of tags, because all tags per resource are ranked equally (with a frequency of 1).

Because no specific distribution of tags may arise in narrow folksonomies, the determination of power tags has to be processed with another approach using collective intelligence. Here, the system is able to count the query terms with which a specific resource has been retrieved successfully (Peters and Stock, 2007; Jäschke *et al.*, 2008; Krause *et al.*, 2008). Given a specific search term (or a set of search terms, if there is more than one), the system presents a list of results. Every time the user accesses a resource via the results list, we consider this search "successful" for this resource. This approach could have been used already in every search scenario. It is a known procedure for full-text search terms in Web retrieval systems (see, e.g., Agichtein *et al.*, 2006; Baeza-Yates, 2005; Baeza-Yates *et al.*, 2007; Baeza-Yates and Tiberi, 2008); it is to our knowledge not used in connection with controlled vocabularies in library settings. It is new to web 2.0 retrieval systems as well. The only thing to do is to connect the retrieval information with the concrete resource. The system stores the information with which query terms A, B, or C a user successfully retrieves and accesses the resource X. As a result, query terms are able to form a distribution of terms or tags as well.

It is possible to apply power search tags to all kinds of web 2.0 services and digital libraries (see Table I). Narrow folksonomies, broad folksonomies and even information services which work with knowledge organization systems (KOS) may profit from power search tags. The user is taken directly into account since the collective intelligence of collaborative information retrieval is now reflected by the user's concrete search and click behavior which determines the allocation of tags to resources.

Of course, this approach will only be useful if many people contribute tags or perform searches. Should no index tags be assigned to a resource, substitute tags can be derived from additional resource annotations like titles, comments or descriptions. Here, a TF*IDF calculation may provide good results which can be used as a starting point (Brooks and Montanez, 2006). In case of non-textual resources, like photos or videos, without any additional textual description, this approach is not applicable. Thus, these kinds of resources rely on textual metadata, at least generated by query terms.

5. Processing of power tags for information retrieval

Using power search tags, power index tags or both it is possible to create a search option for the users. Accordingly, the search engine would scan only the power tags and would disregard all other tags for the comparison of query terms and assigned tags. Thus, precision of search results will be enhanced as this procedure decreases the recall of search results. As Newman (2005) puts it, "this would be simply to throw out the data in the tail of the curve". This search option can only be used as an additional feature, because it is possible that there are valuable tags in the long tail. Additional to this search option it is possible to use power tags as an element of relevance ranking (Peters and Stock, 2007).

For the processing of power tags in retrieval tasks we work with a second inverted file of tags, in which only document-specific power tags are considered. The concrete process may be described as follows using the flow chart sketched in Figure 4.

The chart presents the processing of power tags generated from search terms. The step "generating index tags from query terms" has been skipped in the chart since the processing of power tags for information retrieval relies on tagged resources and has to be solved in advance. The starting point of the process is a tagged resource showing a particular allocation and frequency of indexed tags. The next step is to decide whether there are sufficient users who successfully retrieved the tagged resource *d*. Then the system calculates TF*IDF for all tags of the resource in order to rearrange their allocation. The term frequency TF of a tag *t* is the number of searches in which the users successfully retrieved *d* by using *t* as a search argument or as a part of a search argument. It is possible to work with the absolute number of the document-specific tags or, more elaborated, with the within-document frequency weight (WDF) of tag *t* in document *d*:

$$\mathrm{WDF}(t,d) = \left[ld\left(\mathrm{freq}(t,d,)+1\right)\right]/ldL,$$

where freq(t,d) is the number of the tag t in document d and L is the number of all tags, which are applied to document d (Stock, 2007a, p. 323). The inverse document frequency weight (IDF) of a tag t is calculated by the following formula:

	Index tags	Search tags
Narrow folksonomy	No power tags	Power tags applicable
Broad folksonomy	Power tags applicable	Power tags applicable
KOS	No power tags	Power tags applicable

Note: KOS: knowledge organization system, e.g. nomenclature, classification, thesaurus

Table I. Application of index and search tags in narrow and broad folksonomies

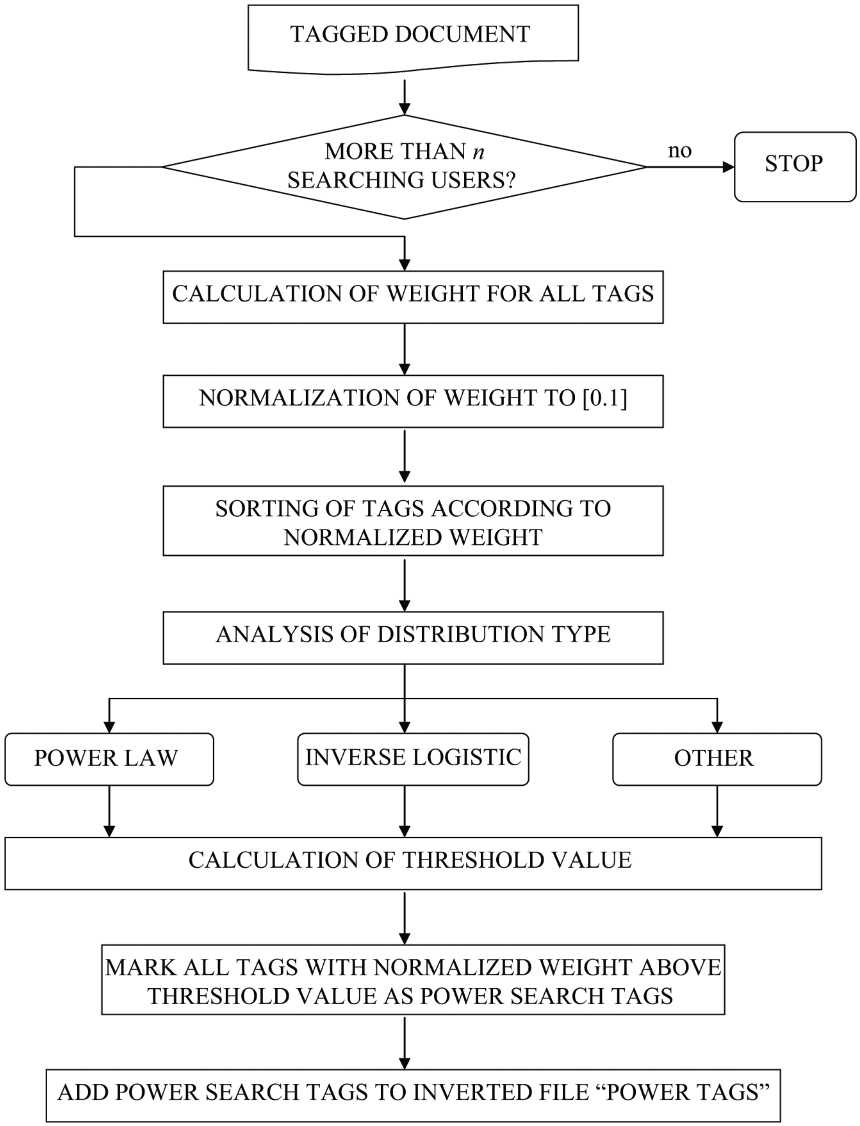

Figure 4.
Mining of power tags

$$\mathrm{IDF}(t) = ld(N/n),$$

where N is the number of all documents in the database and n the number of those documents, which are tagged with the term t (Stock, 2007a, p. 325). The document-specific weight of a tag t in a document d is the product of WDF(t,d) and IDF(t):

$$\mathrm{WEIGHT}(t, d) = \mathrm{WDF}(t, d) * \mathrm{IDF}(t).$$

It is important to note that the system has to re-calculate WEIGHT values for every document from time to time, because the WDF values change with every new document-specific tag and the IDF values depend on the size of the whole database, which is changing over time in "living" systems.

In the next step, the system normalizes the values of WEIGHT to the interval [0,1] for all document-specific tags (the document-specific maximum of WEIGHT(t,d) is considered "1") in order to give rarely used tags a better value. The next step comprises the sorting of tags according to their normalized WEIGHT values. The next important step is the analysis of the distribution type. What follows is an idea for a possible algorithm: if the second ranked tag has a value of about 0.5 or smaller and if the third ranked tag has a value of about 0.3 or smaller, it is evident that the distribution forms a power law. If this is not the case and if there is a turning point in the curve, there may be evidence for an inverse logistic distribution. Now the system has to calculate threshold values: if there is a power law, the threshold value depends on the exponent a; if there is an inverse logistic distribution, the threshold value equals the turning point. All tags with a normalized WEIGHT value higher than the threshold values are marked as "power tags". If there is no power law and no inverse logistic distribution appearing according to this calculation then it is not possible to create power tags.

The processing of "power index tags" will work in a similar way to the procedure shown in Figure 4. The only differences are found in the first two steps. Step one considers the number of users who tag the document d. If there are sufficient numbers of users, the system calculates WDF*IDF for all index tags, where freq(t,d) of tag t is the number of different users who tagged the resource d with tag t. Both power index tags and power search tags should be stored in a second inverted file (Figure 5).

The information retrieval procedure employing power tags is sketched in the following: If a user performs a search on "all tags," the system works with the "normal" inverted file "all tags," if the user wants to restrict the search argument on power tags (Figure 6), the system works with the additional inverted file "power tags." In Figure 5, some data from a single document (with the ID = 11) are presented. The system has calculated normalized WEIGHT value for, say, search tags. The sorted tags form an inverse power law with $a = 1$. Thus, the system regards the first three keywords as power tags. If a user looks for documents with "keyword 1" (or 2 or 3) and marks "power tags only", he will retrieve document 11. If another user looks for documents with "keyword 4" (or 5 or 6) and marks "power tags only" as well, he will not retrieve document 11.

Since in these steps only the equivalence of query terms and resources is determined in order to create a list of search results, the next step is to process a relevance ranking within this list of search results. The relevance ranking will be performed by applying several folksonomy-specific ranking factors for determining the retrieval status value of the resources such as tags (WDF*IDF, cosine, super posters), collaboration (click rates, number of tagging users, rates of comments, authorities and hubs) and prosumers (performatives, relevance feedback, recommendations) (Peters and Stock, 2007, 2008).

6. Conclusions and future work

Folksonomies offer a new approach to information indexing and information retrieval mainly within Web 2.0 environments. Still, little research has been concerned with

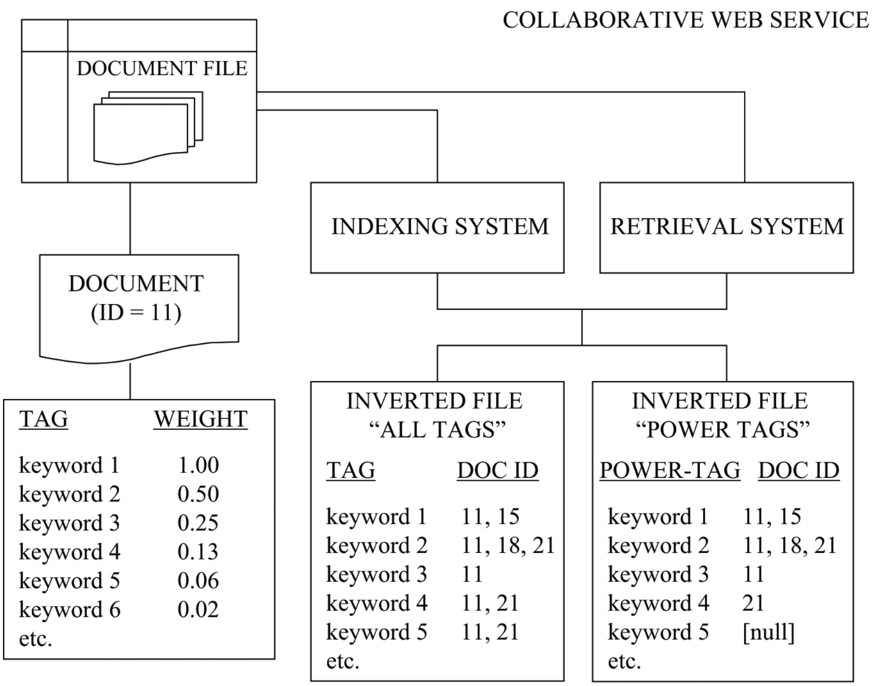

Figure 5.
Collaborative web service applying tags and power tags for retrieval tasks

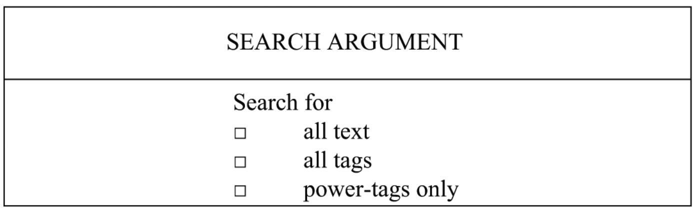

Figure 6.
Search options in collaborative web services

relevance distributions, their development and their application in information retrieval tasks. This paper has introduced power tags for enhancing the quality of retrieval results. For the implementation of power tags particularly two types of tag distributions have to be discussed and should be kept in mind: the power law distribution and the inverse logistic distribution. Both distributions reflect the collective intelligence of the searching and indexing users that can be used for retrieval tasks. Derived power tags (from index tags or from search terms) can be utilized for limiting the amount of searchable tags, in order to simultaneously limit the recall of search results but enhance precision. Thus, the concrete user behavior – the often praised collective intelligence – is taken into account for information retrieval, since index tags are an expression of collaborative information indexing and search tags result from collaborative information retrieval. Future work will have to deal with the transfer of these theoretical and technical ideas to practical search solutions that will also include the evaluation of retrieval results based on power tags.

References

Agichtein, E., Brill, E. and Dumais, S. (2006), "Improving web search ranking by incorporating user behavior information", *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, New York, NY, pp. 19-26.

Allen, L. and Winkler, M. (2007), "PennTags: creating and using an academic social bookmarking tool", *Proceedings of the ACRL 13th National Conference*, American Library Association, Baltimore, MD.

Baeza-Yates, R. (2005), "Applications of Web query mining", *Lecture Notes in Computer Science*, No. 3408, pp. 7-22.

Baeza-Yates, R. and Tiberi, A. (2008), "The anatomy of a large query log", *Journal of Physics A – Mathematical and Theoretical*, Vol. 41 No. 22, Art. No. 224002.

Baeza-Yates, R., Hurtado, C. and Mendoza, M. (2007), "Improving search engines by query clustering", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 12, pp. 1793-804.

Brooks, C.H. and Montanez, N. (2006), "Improved annotation of the blogosphere via autotagging and hierarchical clustering", *Proceedings of the 15th International World Wide Web Conference*, ACM, New York, NY, pp. 625-32.

Cattuto, C. (2006), "Semiotic dynamics in online social communities", *European Physical Journal C*, Vol. 46, Suppl. 2, pp. 33-7.

Cattuto, C., Loreto, V. and Pietronero, L. (2007), "Semiotic dynamics and collaborative tagging", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104 No. 5, pp. 1461-4.

Cattuto, C., Loreto, V. and Servedio, V.D.P. (2006), "A Yule-Simon process with memory", *Europhysics Letters*, Vol. 76 No. 2, pp. 208-14.

Chen, Y.S. (1989), "Analysis of Lotka's law: the Simon-Yule approach", *Information Processing & Management*, Vol. 25 No. 5, pp. 527-44.

Chen, Y.S., Chong, P.P. and Tong, M.Y. (1994), "The Simon-Yule approach to bibliometric modeling", *Information Processing & Management*, Vol. 30 No. 4, pp. 535-56.

Coyle, K. (2007), "The library catalog in a 2.0 world", *Journal of Academic Librarianship*, Vol. 33 No. 2, pp. 289-91.

Dellschaft, K. and Staab, S. (2008), "An epistemic dynamic model for tagging systems", *Proceedings of the 19th Conference on Hypertext and Hyermedia, Pittsburgh, PA*, pp. 71-80.

Dye, J. (2006), "Folksonomy: a game of high-tech (and high-stakes) tag", *E-content*, Vol. 29 No. 3, pp. 38-43.

Egghe, L. (2005), *Power Laws in the Information Production Process: Lotkaian Informetrics*, Elsevier Academic Press, Amsterdam.

Egghe, L. and Rousseau, R. (1995), "Generalized success-breeds-success principle leading to time-dependent informetric distributions", *Journal of the American Society for Information Science*, Vol. 46 No. 6, pp. 426-45.

Egghe, L. and Rousseau, R. (1996), "Stochastic processes determined by a general success-breeds-success principle", *Mathematical and Computer Modelling*, Vol. 23 No. 4, pp. 93-104.

Fichter, D. (2006), "Intranet applications for tagging and folksonomies", *Online*, Vol. 30 No. 3, pp. 43-5.

Fifarek, A. (2008), "The birth of catalog 2.0: innovated interfaces encore discovery platform", *Library Hi Tech News*, Vol. 24 No. 5, pp. 13-15.

Furnas, G.W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C and Naaman, M. (2006), "Why do tagging systems work?", *CHI 06 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, pp. 36-9.

Golder, S.A. and Huberman, B.A. (2006), "Usage patterns of collaborative tagging systems", *Journal of Information Science*, Vol. 32 No. 2, pp. 198-208.

Gordon-Murnane, L. (2006), "Social bookmarking, folksonomies, and Web 2.0 tools", *Searcher – The Magazine for Database Professionals*, Vol. 14 No. 6, pp. 26-38.

Guy, M. and Tonkin, E. (2006), "Folksonomies: tidying up tags?", *D-Lib Magazine*, Vol. 12 No. 1.

Halpin, H., Robu, V. and Shepherd, H. (2007), "The complex dynamics of collaborative tagging", *Proceedings of the 16th Conference on World Wide Web*, ACM, New York, NY, pp. 211-20.

Hotho, A., Jäschke, R., Schmitz, C. and Stumme, G. (2006), "Information retrieval in folksonomies: dearch and ranking", *Lecture Notes in Computer Science*, No. 4011, pp. 411-26.

Huang, H. (2006), "Tag distribution analysis using the power law to evaluate social tagging systems: a case study in the Flickr database", abstracts of posters from the 17th ASIS&T SIG/CR Classification Research Workshop, Austin, TX.

Huber, J.C. (1998), "Cumulative advantage and success-breeds-success: the value of time pattern analysis", *Journal of the American Society for Information Science*, Vol. 49 No. 5, pp. 471-6.

Jäschke, R., Krause, B., Hotho, A. and Stumme, G. (2008), "Logsonomy – a search engine folksonomy", *Proceedings of the 2nd International Conference on Weblogs and Social Media, Seattle, WA*, AAAI Press, Menlo Park, CA.

Krause, B., Jäschke, R., Hotho, A. and Stumme, G. (2008), "Logsonomy – social information retrieval with logdata", *Proceedings of the 19th ACM Conference on Hypertext and Hypermedia*, ACM, New York, NY, pp. 157-66.

Kroski, E. (2008), *Web 2.0 for Librarians and Information Professionals*, Neal-Schuman, New York, NY.

Lux, M., Granitzer, M. and Kern, R. (2007), "Aspects of broad folksonomies", *Proceedings of the 18th International Conference on Database and Expert Systems Applications (DEXA 2007)*, IEEE Computer Society, Washington, DC.

Lyons, C. and Tappeiner, E. (2008), "Cataloging 2.0: metadata research and initiatives at a community college library", *Journal of Library Metadata*, Vol. 8 No. 2, pp. 155-7.

Mathes, A. (2004), *Folksonomies – Cooperative Classification and Communication through Shared Metadata*, University of Illinois Urbana-Campaign/Graduate School of Library and Information Science, Urbana, IL.

Munk, T.B. and Mork, K. (2007), "Folksonomy, the power law and the significance of the least effort", *Knowledge Organization*, Vol. 34 No. 1, pp. 16-33.

Newman, M.E.J. (2005), "Power laws, Pareto distributions and Zipf's law", *Contemporary Physics*, Vol. 46 No. 5, pp. 323-51.

Noruzi, A. (2006), "Folksonomies: (un)controlled vocabulary?", *Knowledge Organization*, Vol. 33 No. 4, pp. 199-203.

Oluić-Vuković, V. (1997), "Bradford's distribution: from the classical bibliometric law to the more general stochastic models", *Journal of the American Society for Information Science*, Vol. 48 No. 9, pp. 833-42.

Peters, I. and Stock, W.G. (2007), "Folksonomies and information retrieval", *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology*, Vol. 45, pp. 1510-42.

Peters, I. and Stock, W.G. (2008), "Folksonomies in Wissensrepräsentation und Information Retrieval" ("Folksonomies in knowledge representation and information retrieval)", *Information – Wissenschaft und Praxis*, Vol. 59 No. 2, pp. 77-90.

Peterson, E. (2006), "Beneath the metadata: some philosophical problems with folksonomies", *D-Lib Magazine*, Vol. 12 No. 11.

Smith, G. (2008), *Tagging. People-powered Metadata for the Social Web*, New Riders, Berkeley, CA.

Spiteri, L.F. (2006), "The use of folksonomies in public library catalogues", *The Serials Librarian*, Vol. 51 No. 2, pp. 75-89.

Spiteri, L.F. (2007), "Structure and form of folksonomy tags: the road to the public library catalogue", *Webology*, Vol. 4 No. 2, p. 41.

Stock, W.G. (2006), "On relevance distributions", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 8, pp. 1126-9.

Stock, W.G. (2007a), *Information Retrieval. Informationen suchen und finden* (*Information Retrieval. Searching and Finding Information*), Oldenbourg, Munich.

Stock, W.G. (2007b), "Folksonomies and science communication. A mash-up of professional science databases and web 2.0 services", *Information Services & Use*, Vol. 27 No. 3, pp. 97-103.

Stock, W.G. and Stock, M. (2008), *Wissensrepräsentation* (*Knowledge Representation*), Oldenbourg, Munich.

Sweda, J. (2006), "Using social bookmarks in an academic setting: PennTags", paper presented at the 17th Annual ASIS&T SIG/CR Classification Research Workshop, Austin, TX.

Tague, J. (1981), "The success-breeds-success phenomena and bibliometric processes", *Journal of the American Society for Information Science*, Vol. 32 No. 4, pp. 280-6.

Toffler, A. (1980), *The Third Wave*, Morrow, New York, NY.

Tonkin, E. (2006), "Searching the long tail: hidden structure in social tagging", paper presented at the 17th ASIS&T SIG/CR Classification Research Workshop, Austin, TX.

Trant, J. (2006), "Exploring the potential for social tagging and folksonomy in art museums: proof of concept", *New Review of Hypermedia and Multimedia*, Vol. 12 No. 1, pp. 83-105.

Vander Wal, T. (2005), "Explaining and showing broad and narrow folksonomies", available at: www.vanderwal.net/random/entrysel.php?blog=1635 (accessed April 23, 2008).

Weaver, M. (2007), "Contextual metadata: faceted schemas in virtual library communities", *Library Hi Tech*, Vol. 25 No. 4, pp. 579-94.

Weller, K. and Peters, I. (2008), "Seeding, weeding, fertilizing – different tag gardening activities for folksonomy maintenance and enrichment", *Proceedings of I-SEMANTICS 08, Graz, September 3-5, 2008*, pp. 110-7.

About the authors

Isabella Peters is Researcher at the Information Science Department of the Heinrich-Heine-University Düsseldorf (Germany). Her main research topics are semantic relations and folksonomies.

Wolfgang G. Stock is the Head of the Information Science Department of the Heinrich-Heine-University Düsseldorf (Germany). He has written more than 200 articles and about ten books, including text books on information retrieval and knowledge representation. Wolfgang G. Stock is the corresponding author and can be contacted at: Stock@phil-fak.uni-duesseldorf.de