The TicTag application: towards tag-based meta-search-for browsing the web

Papakonstantinou, C.; Panagiotou, I.; Verbeek, F.J.

The TicTag Application: Towards Tag-based Meta-search for Browsing the Web

Christina Papakonstantinou
Media Technology, LIACS
Niels Bohrweg 1
2333 CA Leiden, NL
+31 (0)63 41 06 553
c.papakon@tictag.cc

Ioannis Panagiotou
Media Technology, LIACS
Niels Bohrweg 1
2333 CA Leiden, NL
+31 (0)62 85 93 644
i.panagiotou@tictag.cc

Fons Verbeek
LIACS
Niels Bohrweg 1
2333 CA Leiden, NL
+31 (0)71 52 75 773
fverbeek@liacs.nl

ABSTRACT

The Web has developed as an enormous information resource. Especially in the last few years with the thriving of the so-called web2.0 services allowing user-generated content easily to be entered in -and shared by- typical web 2.0 databases. Bookmarking systems enable users to label their resources with tags, which in aggregation give rise to dynamic categorization schemes, i.e. folksonomies. Web querying through folksonomies presents an interesting potential in contrast to traditional search engines, such as the suggestion of relevant topics that may refine or even define the original search term. Both approaches are in wide use, each appreciated for their own qualities. However, it is possible for the methods to be used complementary, making use of the specific advantages of each of them. In this paper we specifically introduce the notions of tags and folksonomies and we present a method of using these notions in a tag based search technique. We discuss and elaborate results on the basis of preliminary experiments with our tag-based search engine.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Graphical user interfaces (GUI)*

General Terms

Algorithms, Design, Experimentation

Keywords

Tags, Keywords, Web Search, User Interfaces, Experimentation, Folksonomies

1. INTRODUCTION

Web-browsing as a user experience has changed a lot within the last years. To some extent, this can be attributed to the development of Social Networks and Bookmarking Services that enhance the user interaction with the web content allowing users to share files with each other (photos, articles, bookmarks etc), as well as to organize and annotate these as desired.

Therefore, users gain a richer pool of recommended items suggested by other members of the web community as well as a more sensible interaction with their own items that they now contribute both for sharing and for later personal retrieval.

Known examples of such services are Flickr [6], a photo and video sharing community, where users upload their photos, categorize them in sets, label and describe them; Technorati, an aggregator and search engine especially for blogs and blog articles and Delicious [10], a bookmarking service, in which users can save their own bookmarked content and add a description to it in the form of a few words; these descriptive words are commonly referred to as *tags*. Additionally, Delicious allows browsing through content and bookmarks or popular items of other users, i.e. often bookmarked sites by the people using Delicious.

Description of content is treated differently for different services. Flickr for example allows full description in the form of phrases, Blogging services allow a combination of a few words, whereas Delicious recognizes only word units, e.g. the description 'a war photographer shooting' will be understood by the system as the four unordered, discrete and non-structurally connected words 'a', 'war', 'photographer', and 'shooting'.

There are restrictions as to who can assign tags to a resource. In Technorati's search engine [16] however, only tags assigned by the creator of a blog-entry are used and Flickr's system works in a similar manner for photos. Delicious, on the other hand, allows for free tagging of any resource by anyone who uses the service.

The collection of the most frequently user-submitted tags is usually presented in a *tag cloud*, e.g. Delicious provides such a tag-cloud view at http://delicious.com/tag/. Tag clouds are substitute for a list of items in which tags are presented in a paragraph-like format and most popular tags are typed with a bigger or bolder font to denote importance. The order of tags is either alphabetical or by popularity and lacks semantic correlations between items, although proposals have been made to improve this by making use of tag similarity and a clustering-based layout [3].

The complete collection of tags in a tagging system is said to form a *folksonomy*. The term is a combination of the words 'folk' and 'taxonomy' and was first coined by Thomas Vander Wal in a mailing list discussion. Although the legacy of the term is under dispute we will use it for the sake of comprehensiveness. Folksonomy can be considered as a social categorization scheme. Tagging services mentioned above offer an alternative strategy of web searching/browsing to conventional search engines; they allow searching through this social scheme. In combining both of these strategies we have developed a method for searching that embraces web semantics in a different manner. This method is implemented in our

browsing application TicTag (cf. section 5.2 on implementation).

The remainder of this paper is structured as follows. We will first go through a comparison of the advantages and disadvantages of the notions introduced and their corresponding searching approaches. We will then present an application which enables searching/browsing through tags in an innovative way that offers browsing through a tag's related URL results as retrieved from four different sources, and allows for comparison between different derived tag clouds. At the end of the paper we will take a look at four cases of tag cloud comparison and the related observations. We will then conclude that folksonomic data can be employed for relevant topic retrieval but also as control data for search engines' results.

2. TAGS

A tag usually contains a one-word description of a web resource. The collection of tags of one particular user constitutes a personal categorization scheme over a personal collection of bookmarks. Since tags are primarily used to facilitate relocation of the tagged item, they may reflect both the user's interest on the (often multi-topic) content as well as the user's understanding and opinion on it.

2.1 Usage of Tags and Advantages

Tags are easy to use. Firstly they are not restricted to any vocabulary; the user is allowed to use any word – or even a random combination of letters – he finds appropriate to annotate a resource. Therefore free tagging is much more flexible, easy to use and does not require the general consensus on what should be tagged with which tag [4] in contrast to a pre-decided fixed ontology scheme. Secondly tagging process requires a low cognitive cost and "*enables loose coordination, but does not enforce the same interpretation of a concept*. […] *That would create chaos in a shared folder scheme, but works well in a social tagging system*. […] *By allowing loose coordination, tagging systems allow social exchange of conceptual information*." [1]

Tags reflect the user's vocabulary at the time and are able to fast enough catch-up with the vocabulary changes, which makes tagging suitable as a web categorization strategy. Clay Shirky [13] argues that a strict categorization system would not be successful to the web, mainly because of its large and unstable corpus, but also because users are unfamiliar with categorization systems: '*Users have a terrifically hard time guessing how something they want will have been categorized in advance*'.

The success of tagging has been attributed on feedback and asynchronous communication in process of tagging [2] [14]. A user has access to the tags of the group; comparing to his tags this gives him feedback on the meaning of the terms as perceived by the group, so the community negotiates the meaning of the tags in the system's folksonomy.

In contrast to keywords and publishers' metadata[1], tags also constitute a third party opinion. While publishers' metadata may be quite accurate they are still a one-person's contribution. It is not necessary that the publisher succeed in labeling -or even bother to label- the item so that any interested party will retrieve it. Tagging by many readers though is more likely to cover all aspects of the item, even the not quite promoting or flattering ones [18]. In this sense tags can be considered to be more accurate and more democratic [4].

2.2 Tag Limitation and Disadvantages

Tagging in social bookmarking services facilitates document retrieval; however there is a social aspect in the process, since bookmarks are shared to the community. Tags therefore serve both retrieval by the individual and the tagging community; they should therefore both comply with the user's and the society's norms. The duality of the goal is thought to bring about many categorization problems. i.e.:

In particular, tagging suffers from two known problems of human indexing, i.e. problems that often arise when humans are called to freely categorize items.

- Synonymity, that is when different indexers use same indexes in different context, which results in tags often being ambiguous [2] and inaccurate; and the opposite problem of
- Inter-indexer inconsistency that is when different indexers use different indexes for the same document [7].

Tags such as 'toread' or 'todo' are often used for personal organization of bookmarks and are not meant to contribute to the social annotation of web content. On the other hand, it is difficult to force users to always follow the community norm in their tagging strategies. This would make the system inflexible and ultimately unusable.

Moreover, tagging systems do not usually offer synonym control. 'Mac' and 'macintosh' exist in the system as different tags, although they refer to the same thing. Plurals and singulars of the same word are also treated as different tags. A search on 'blog' and 'blogs' in Delicious, for example, will return different sets of results. In the same category of 'sloppy' tagging fall tags, such as 'l.a.', 'los-angeles' and 'los_angeles', that are bi-lexical expressions whose words need to bound together to express the meaning, but there is no consensus as to how this should be done. These problems complicate the construction of a formal and well-defined categorization and classification system derived from users' tags, that being ontology or taxonomy. Instead the structure that emerges what is nowadays referred to as a folksonomy.

3. FOLKSONOMIES

Folksonomy is the categorization system that is formed by the total set of tags in a tagging system and the corresponding resources; so, it can be described as the collection of all users' personal categorization schemes[2]. It has been often compared

[1] Metadata is generally a piece of information referring to another piece of information. Metadata's goal is to describe the information they refer to. An attempt to map the web pushed many web-site owners to insert metadata information in their site's header (the publisher's metadata). Metadata would help search engines to know what the site is about so as to return better results, but not every site has employed them. A main characteristic of a web-site's metadata is that they are not visible to the user. In order to access them one should view the source code of a page.

[2] The definition of folksonomy depends on the author. In this paper we will consider folksonomies derived from a tagging system that also enables the viewers of an item to tag it, apart from its creators. The form and inherent dynamics of this tag-set immediately emerge from the fact that folksonomies are built within a process that involves human critical thought,

with ontology and taxonomy, but it has different rules and properties.

Ontologies are more often domain specific, i.e. they are used to classify concepts in a specific (usually scientific) field. They are produced by experts and presuppose the consent of the scientific community on the meaning of the included terms. In order to better map the semantics of these items ontologies use relations between concepts. One can make assertions on how a term is related to another in ontology, if it includes it, or is a property of it etc. Folksonomies conversely are flat spaces with no obvious semantic structures and loose in the semantics of the terms.

In contrast to other classification systems and taxonomies, a folksonomy has no clear edges of discrete content. In the highly hierarchical taxonomy of biological organisms an animal can be contained in only one category, which in its turn is contained in a bigger category and so on. But a folksonomy is loose and allows content to be categorized in more than one class and there is no inferred hierarchy of the tags within it. Folksonomy also inherits the advantages and problems of its construction units, tags.

Its popularity results from the ease of construction and the low barriers of entry; but it suffers from multiplicity of terms and the consequences of having non-controlled vocabulary. It is more appropriate to categorize the dynamic and frequently updated web content than a formal ontology, but it also contains much noise because of too personal or 'sloppy' tags pointing to irrelevant information.

Research, however, has shown that there exist some patterns in their usage. For example, users tend to assign broader terms for tagging their resources [8] [12]. Broad tags imply high recall but low precision, whereas narrow tags imply low recall and high precision; consequently broad tags are better for browsing tasks whereas narrow tags are more suitable for querying tasks [3].

Folksonomies have been also appreciated for their bottom-up approach of classification instead of the more inflexible top-down approach used by traditional classification methods [5]. "Even if the consensus of a common ontology can be achieved it may not be able to catch the fast pace of change of the targeted web resources or the change of user vocabularies in their applications" [4].

Much research has been done concerning the improvement of the latent clustering structure of the folksonomy and identification of tags affinity, in order to provide better defined clusters of topics to facilitate navigation of the user [3][11][12]. To that end, suggestions have been made towards guidelines for the formation of tags, especially bi-lexical expressions, so as to minimize the appearance of sloppy tags [9], as well as the possible layouts [3].

Folksonomies, also present a remarkable stability of tags' frequency proportions for a resource approximately after the first 100 tags have been assigned to the resource [8]. For example, YouTube is tagged 29.6% with 'video', 15.4% with 'you tube', 14.8% with 'videos' and 8.3% with 'web2.0' [17]. Observations have indicated that these proportions do not significantly change over time but instead remain almost fixed; they can therefore provide a description in terms of proportional tags for the resource in question in contrast to the short text description given in commercial search engines' results.

The fixed frequency of tags might result from the fact that tagging, as a naive classification process involves human cognitive processes. To this respect, from relevant investigations [20] it has been argued that tagging consistency is due to fundamental aspects of mental architecture all humans share, because it is not only observed in the popular tags –so one can attribute the fact to the compliance of the tagger to the group's norm- but also in less frequently used tags.

4. COMPARISON OF SOCIAL BOOKMARKING WITH TRADITIONAL SEARCH

In our comparison we will be focusing on two subjects: the choice of the search term and the general model used and results returned by each approach.

4.1 Keyword vs. Tag

A basic distinction of social bookmarking search and traditional search is the choice of search term. Although the concepts of keyword and tag are close, there is difference in the user's expected results using a tag in Delicious and using a keyword in Yahoo!. This expectation can be explained by the mental model of the user, i.e. the users' knowledge and familiarity with the way these systems work- especially when the user is computer literate. So, one of the differences -from user perspective- between *tag* and *keyword* is the service itself which employs them and the way they are used by that service; although in essence, both are an indication-word or cue of what is being searched for.

4.2 Folksonomy vs. Link-Based Structure and their Relevant Results

The different settings and models used are also a distinction of the two search strategies. Firstly, the amount of URLs covered by the databases of a search engine generally outnumbers that of a bookmarking service. For example, Google recently announced to have reached the amount of 1 trillion indexed unique URLs [19], whereas Delicious counts 150 million urls [23] bookmarked by its users. This is also expected if one considers the process of a URL to enter the database of these services. While search engines crawl to almost any web-address, the content of the book-marking services is more selective by nature. For a web site to be listed in a tagging system there must be at least one person that considers it useful to be in his bookmarks and in addition, that person must, at the same time, be a user of the specific tagging system.

Secondly, search engines base their results on the keyword occurrence of a document and 'authorities'[3] derived by the link-structure of the web. Tagging systems, on the other hand, base it on human categorization abilities and how heavily bookmarked a site is.

It is difficult to assess on which method is better and which results are more relevant to the user needs at the time. In a comparative research [15] however quantitative data show 'the URLs of the bookmarking system cover over-proportionally the top-results of the search engines ranking. A likely explanation is that taggers use search engines to find interesting bookmarks;

social interactions and temporal events that trigger off patterns in tagging behavior.

[3] Mostly linked web-sites are considered by Google's algorithm as 'authorities' and are higher ranked.

meaning that both methods are used in alternating fashion hoping for better results.

As shown in [22], however, a consensus of a shared ontology, thus an agreement on organization of content and topics can be more easily found amongst users within a small group of similar interests and even more amongst the members of the core of the group. It is possible therefore that retrieval can be facilitated by approximation of the user's group in terms of interest and vocabulary.

So, if one presupposes that users belonging in a group share similar interests and vocabularies, then results based on the folksonomy created from the users of that group would be more accurate for a member of the same group. A proposed ranking algorithm for folksonomies called FolkRank [24] –an altered version of PageRank- takes into account the tripartite graph of the folksonomy, so that the results depend also on *who* tagged a resource.

Another difference in searching through a folksonomy is the social element of the process: the overview and feedback on the usage of a term in the tagging community. Tagging systems already have available related-tag suggestions[4]. The user can make an estimation of the semantics of his/her search term and then refine his search by selecting one or browse and discover new related content. This interaction is not possible using a search engine alone, but is one of the key features –the inferred semantics- of ontology as stated earlier.

5. COMBINING APPROACHES IN THE TICTAG APPLICATION

Instead of juxtaposing the different search approaches, it is possible that they can operate complementarily. One can use both systems, in one that combines retrieved data from both and uses them accordingly - such a system is also called *mash-up*. In this section we develop our methodology and explain the corresponding implementation of this methodology in our TicTag application.

5.1 Tools and Methods

In order to make use of the semantic properties of a folksonomy a bookmarking service's data can be used. A search on a term in Delicious for example, returns two kinds of results:

1. The URL results tagged with the search term and

2. A set of –maximally 11[5] - related tags often assigned together with the search term.

The second kind of results is offered to the user as an additional tag suggestion in order to help navigation to relevant topics.

Additionally, in order to make use of the amplitude of keyword search, search engines can be used. A search on these search engines will return the top ranked URL results, which might be already bookmarked and tagged within a bookmarking service.

Because, as we mentioned before, of the overlapping of top results in search engines and bookmarking service, it makes sense to expect that some of the URLs will have been tagged. Additionally, it makes sense to assume that the tags assigned are also related to the search term that brings those URLs in highly-ranked positions.

Consequently, search engines can provide an additional set of Delicious' tags –in case these exist- which can then be presented to the user along with the previously mentioned set of related tags.

Moreover, based on the findings of [8] which suggest that there may be a nearly fixed set of tags popularly assigned to a URL resource, it is possible to derive some tag-based description aggregated by users' bookmarks as an overview and feedback on the content of a specific URL.

Such a combinatory search would result in an extended set of related tags to the search term, parts of which are descriptive for top-ranked results from the search engines. This would give an overview of the concept of the search term, its related topics and how it is used by the users of the bookmarking community.

We have elaborated this idea in our TicTag application. In the next sections we will explain the way this application works and use it to test the idea. The application can be downloaded at: http://tictag.cc.

5.2 Implementation

TicTag was built in Processing [21] and uses one bookmarking service, the RSS- feed of that service and two commercial search engines to retrieve results.

Delicious was used as the bookmarking-service resource. The application accesses the corresponding page of a search on a tag (for the tag 'air' then the page http://delicious.com/tag/air is accessed) and parses source code of that page for the related tags offered by the system.

For the keyword search TicTag uses the APIs of Google and Yahoo! retrieving the top eight[6] results from each engine for a custom search on the user's search term. All 16 results are tested on Delicious system (this is done by accessing the page http://delicious.com/url/ concatenating at the end the md5 hash of the URL) and if they are found to have been bookmarked, then the tags assigned are also parsed.

The tags collected using both these ways are combined to a tag cloud of related to the search term tags, which we call its *characteristic tag cloud* (cf. Fig. 1). This tag cloud consists of all suggested tags from Delicious plus the tags annotating the top results of the search engines, giving this way a broader set of related topics for the user to chose from.

Every tag in the tag cloud is a link to the set of its relevant URLs filtered with the user's search term. This set of the tag's URLs we call its *index* and consists of four subsets of URLs (cf. Fig. 2) placed in radial order around the tag. The first two sets of URLs are retrieved using the APIs of the two search engines; on which the request is put for the concatenation of the tag and the search term input by the user. The last two are retrieved accessing the Delicious' page of the most popular bookmarks tagged with the search term and its corresponding RSS feed for

[4] The Google search engine has very recently been enhanced with auto-suggestion of previous searches and features of providing feedback on the quality of results. The social aspect and the available feedback are two of the most important characteristics of a so called web2.0 service, a very promising model towards many new services turn to.

[5] Delicious system by default offers no more than 11 related tags.

[6] Google's API returns up to eight top results. For the sake of balance and symmetry we only use eight out of ten of Yahoo! API-returned results.

the most recently added.

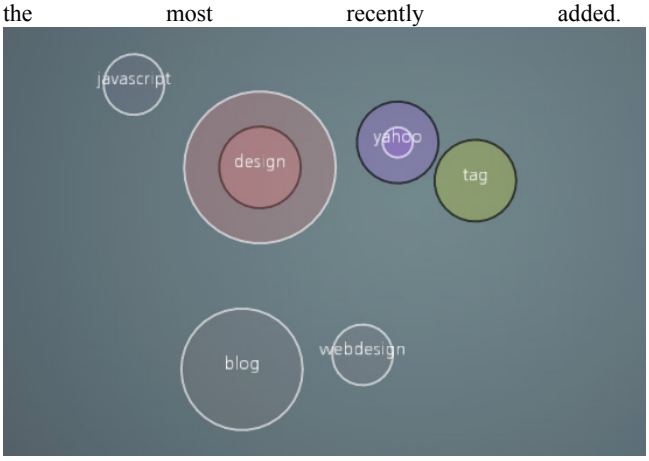


Figure 1. The characteristic tag-cloud of the word 'tags'

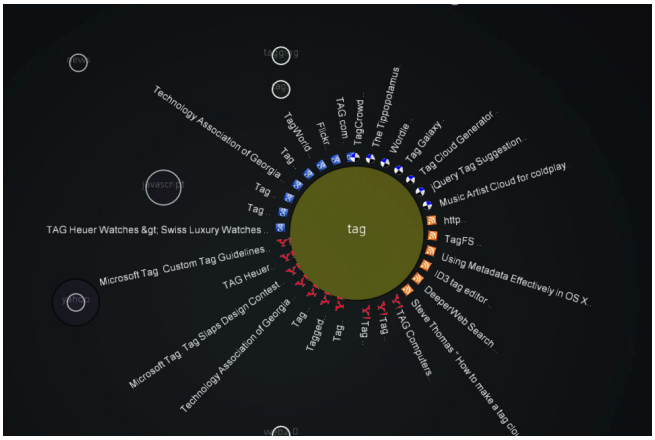


Figure 2. The radial index of a tag

In the application it is possible to see what caused each tag to appear in the tag cloud, i.e. if it was assigned to some URL and which one was that, or if it was included in the Delicious list of related tags. The user this way has the ability to judge upon a tag's relevance on the basis of its occurrence throughout the different search methods.

In addition, it is possible to compare two tag clouds derived from two different search approaches.

Comparison can be done either by juxtaposing the two tag clouds one in each side of the screen or by viewing their intersection that are the tags which appear in both characteristic tag clouds.

5.3 Measuring Tag Importance

In traditional tag clouds, relative importance of a tag is visualized with a bigger font size or bolder letters. A way to measure importance is to count the popularity of a tag, i.e. how many times users have used it to describe a resource. In tag clouds related to a specific term, co-occurrence can be used instead, i.e. how many times the tag has been assigned to a resource to which the search term has also been assigned.

Tag clouds in Delicious offer an estimation of the extent of popularity of the most popular tags in its system by means of font size (by scale of 1 to 5). Moreover, it is possible to see the amount of users that have assigned a certain tag to a certain link (arbitrarily many; there can be found links bookmarked by more than 20,000 users).

In TicTag both of these pieces of information contribute to the determination of a tag-importance. In order to normalize the two scales one should know the amount of users which corresponds to a tag of great importance –as shown in Delicious- and the amount which corresponds to a tag of minimum importance. We transformed the amount of users to a logarithmic scale as we do not have access to such information.

5.4 Combination Is Powerful

From our experiments and testing that we have conducted with the application the method seems to enhance the initial tag set. We observed that usually the second set of tags –i.e. those retrieved by the Yahoo and Google top ranked sites- gives equally accurate results in broader contexts as well.

Although in topics well covered by the Delicious users many tags of the top results of Google and Yahoo! overlap the Delicious related tags – an observation also made by Krause et al. [15], in less popular tags or tags not used at all, such as the names and street names mentioned before, Delicious' results are outnumbered by those retrieved by traditional search engines.

On the other hand results from the two big search engines are merged with the results suggested by the Delicious' users. Overlapping of URLs is also observed but less often than tags. According to findings obtained from other experiments [15], it is more likely for URL overlapping to occur in the top results - which are the ones we use- rather than the lower ranked ones.

6. PRELIMINARY EXPERIMENTS

To gain an additional overview of the method and the advantages of the combination of searching approaches in this mash-up application we will present some experiments results. We have divided the examples in two pairs of synonyms and opposites, and popular general and specific tag examples.

6.1 Synonyms

As mentioned earlier, one of the concerns about folksonomies is synonymity; that is whether it is possible when looking for 'Information Visualization' for example, to retrieve resources tagged as 'Infovis' as well, whose meaning is essentially the same- and thus contains relevant content.

It is not necessary for a searcher to know all synonyms of a term or for all taggers interested to the topic to have used the same tag; tagging systems do not have synonym control. However, constructing the characteristic tag cloud of 'Information Visualization' we observe that many of the related tags are actually synonyms of it, within which 'Infovis' is also included (cf. Table 1).

Table 1. Comparison of 'Information Visualization' and 'Infovis' tag clouds

	Information visualization	**Infovis**
exclusive tags	Wikipedia, Information_vi, Hci, ui, Journal, Infographics, Software	Mapping, Tools, Usability, Search, Database, Archive, Conference, News
common tags	Data, Research, Interface, Information design, Blog, Reference, Design, Visualization, Graphics, info vis, Information, Infoviz	

Looking closer to the results shown in Table 1, we see that 'Information Visualization' is accompanied exclusively either by synonyms, such as Information-vi or 'Infographics', or tags

representing a some-what more general field, such as HCI or UI (User Interfaces). In this context 'Wikipedia', a quite common tag, also appears.

On the other hand, tags exclusive to 'Infovis' do not contain synonyms at all, and are a bit more technical: 'Databases', 'Usability', and 'Mapping'. We also see tags such as 'Conference', 'News' and 'Archive'; this gives the impression of taggers somewhat more active in the field.

The observation could be a starting point for an extended research and quantitative analysis of the use of synonyms from group to group. Results might show that users using 'Information Visualization' instead of 'Infovis', are only recently interested in the field, possibly still collecting information (see. 'Wikipedia' tag or 'Journal' tags), whereas 'Infovis'-taggers are already familiar with the concept. Such a result would emphasize the importance of separate synonymous tags to exist as such, since each is the key to access information for a different group of people.

6.2 Opposites

Along with the synonyms example we have performed an experiment with opposites. For the pair of opposite tags we chose 'war' and 'peace'. The words although opposite in meaning, would be expected to often appear together as tags on resources and articles, possibly of journalistic or political nature.

Table 2. Comparison of 'war' and 'peace' tag clouds

	War	Peace
exclusive tags	Ngo, Children, Human-rights, Movies, Russia, World, Military, Philosophy, Bush, Georgia, News, Child soldiers, NATO, Jason-Statham, etc.	Global, Nobel, International, Projects, Organization, War_crimes, Positivity, Education, Religion, Imperialism, Capitalism, Hippie ,Progressive, Paix, Colombia, etc
common tags	reference, Iraq, us, activism, politics, school, peace, war	

Indeed, as shown in Table 2, war and peace both appeared in each other's cloud of related tags, unlike in the previous example (relation of specialization). Their intersection also embodied: {Iraq, us, activism, politics, school, reference}. 'Reference' is another common tag, similar to "Wikipedia".

Another observation is that while 'peace' related tags are in general abstract, like 'Positivity' and 'Religion', 'war' presents more specific concepts, such as 'Georgia', 'Bush' and 'Jason-Statham'[7].

Last, the majority of the more popular tags belong to the cloud tag of *war* among which we also find {'NGO' (Non-governmental organization), 'children', 'human-rights}'. The opposite concept of the *peace* tags, although more numerous, are predominantly of minimum frequency.

6.3 Specific and Non Popular

We argued earlier that tags retrieved by URL enhance the related tag-set offered by the Delicious service, especially when the search is on non popular terms, for example terms that are specific to some smaller group of users, such as a part of the non English-speaking users. Examples of such specific terms could be non English names, such as the name {'Giannis Panagiotou'} or the name of the newly developed {'TicTag'} application (cf. Table 3). We see that Delicious offers no related tags, so a user would not be able to approximate the meaning of the tag, or in this case the fields concerning the person, by relevant topics.

Table 3. Comparison of results based on source

	Giannis Panagiotou	TicTag
delicious Tags	0	5
Google URL Tags	20	0
Yahoo! URL Tags	11	10
Overlapping tags	10	0

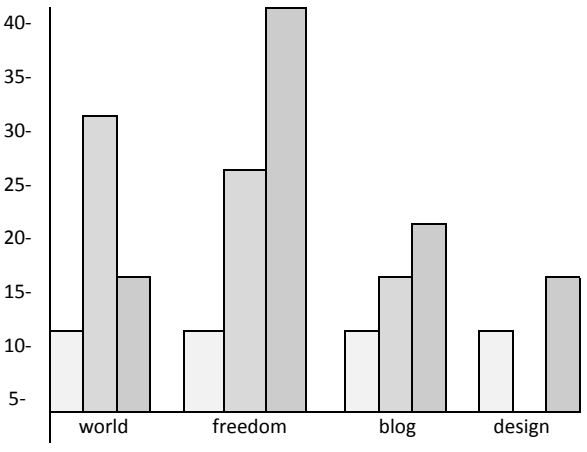
On the other hand, there were found two links amongst the top 8 results of the search engines related to the name, one is {'nisgia.com | interactive designs'} and the other as {'Exhibitions/Processing 1.0'} that were tagged with {agency, gr_designers, inspiration, weiss, webdesign, portfolioexample, greek, portfolio, design, flash} and {research, cool, java, software, graphics, art, visualization, programming, design, processing}) respectively.

In the 'TicTag' example, we see that Delicious offers 5 tags: i.e. {innovative, application, dev, keywords, creative}, Google returns none, but Yahoo! contributes 10 more tags: i.e. {generative, code, visualization, processing, programming, design, processing.org, art, blog, blogs} approximating the term's meaning somewhat more. From these results we can learn that TicTag is an application possibly built in Processing, has something to do with visualizations and there may also be a blog referring to it.

6.4 General and Popular

It is common that popular or very general tags return a lot of results in their characteristic tag clouds. Tags that are confirmed from both Delicious, and top URL results as related, are often in these cases. What is interesting to see, though, are the relative contributions of each service in the tag cloud of a search term.

Table 4. Comparison of relative contributions



[7] Famous actor playing in a movie called 'War'.

We compared the relative contributions of Delicious, Google and Yahoo! in the tag clouds of 'world', 'freedom', 'blog' and 'design' the first two of which are randomly chosen as general and the last two appeared as the two most popular tags in the delicious tag cloud (d.d. August 13,2008).

In Table 4 the results for Delicious are shown in light grey, Google's in middle grey and Yahoo!'s in dark grey. Delicious' system restricts the number of related tags to maximally 11. Yahoo!'s top URL results for 'freedom' were found actively bookmarked contributing 40 tags. Last, Google outnumbers Yahoo! only in 'world's tag cloud, while in 'design' returns no tag-results at all; an observation for which we do not have an obvious explanation.

Although the sample is rather small in order to deduce vigorous conclusions, it is interesting to further investigate the fluctuations of tag-contributions of the different engines under different searches. A study specifically designed to this research question might show that differences are topic specific; e.g. users may prefer one engine over the other for discovery of new interesting bookmarks for certain topics.

7. DISCUSSION

TicTag reveals the concept areas related to the search term of a user as these are existent in the internet world. These are the 'internet semantics' of the term that the user encounters by using a search engine anyway, but the feedback given by folksonomic data can be proven valuable for searching and browsing.

The characteristic tag cloud returns concepts of higher or lower specification, i.e. more general or less, synonyms, opposites (cf. section 6.2 'war' and 'peace' example); all according to the semantics of the folksonomy one chooses to use and filtered by the keyword-based search engines.

Results provide information about the content of search-engine results in terms of user acknowledgement of relevancy and quality. The information would, of course, have been more complete with the inclusion of more folksonomies (cf. 8 on future work) in the system and with all -or most- of the results having been tagged; the reasons why many are not, however, is still to be investigated. They may have not been discovered yet; or they may have not been considered to be sufficiently interesting to be bookmarked; or they may appear steadily on the top ranked results of the engine so that relocation is guaranteed.

Moreover, our method can be described as providing an approach for semantic surfing on the link structure of the web. Web-pages also demonstrate some topic consistency which is reflected in users' tags and picked up by our approach. The top ranking of a frequently linked-to page and its content is then either verified by the folksonomy or not, and if data is available the perceived content is shown by the means of tags.

In the example of war in section 6.2 the dynamics of a folksonomy are shown in the recently popular tags 'ossetia' and 'georgia'; Delicious catches up with the world news[8] fast! This quality is a consequence of the fact that a large number of users with diverse interests contributes to the content of the database and the folksonomy.

Last an important observation concerning synonyms is the small amount of their overlapping tags. In the example of 'infovis' and 'information visualization' (cf. section 6.1) the intersection of the respective tag clouds was consisted by quite few tags. It's both about *how* a user searched a topic in a search engine and *what words* he used to tag the results. Our results suggest that tags are pathways to knowledge but from different starting points, just like people have different backgrounds.

[8] Conflicts is Ossetia were a major topic in the world news in August 2008, when experiments were conducted.

8. FUTURE WORK

The relations between the two dominating search approaches at the moment should be studied further in an extend research project. Qualitative data can reveal further differences and the results can subsequently be used to identify which method is more appropriate for a particular kind of search terms.

In addition, for the meta-search it would be useful to include more tagging services, such as StubleUpon or Digg, as well as more search engines or meta-search engines. It would be interesting to extend our methodology with the Flickr database so that tags for related to pictures could also be employed.

Statistical and linguistic analysis of the frequency and qualities of tags can provide data for assessment on the properties of tag clouds. Refinement of a query could then be navigated solely through tags starting from the more general ones and moving on to the more specific (cf. section 3 on broad and narrow tags).

Furthermore, new directions in research can be taken on personalization of searching (cf. section 3 on specific group folksonomy). Defining and using group folksonomies might be a good step towards the perfect search engine.

9. REFERENCES

[1] Sinha, R. A Cognitive Analysis of Tagging. Rashmi Sinha Blog. Retrieved from: http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html (2005).

[2] Mathes, A. Folksonomies – Cooperative Classification and Communication Through Shared Metadata, Retrieved from http://www.adammathes.com/academic/computer-mediated-communication/ folksonomies.html. (2004)

[3] Montero, Y., and Herrero-Solana, V. Improving Tag-Clouds as Visual Information Retrieval Interfaces. In Proceedings of the International Conference on Multidisciplinary Information Sciences and Technologies. Retrieved from: http://nosolousabilidad.com/hassan/improving _tagclouds.pdf (2006)

[4] Wu, X., Zhang, L., and Yu, Y. Exploring Social Annotations for the Semantic Web. In Proc. of WWW'06. Retrieved from: http://www2006.org/programme/files/pdf/4071.pdf (2006)

[5] Merholz, P. Metadata for the Masses, Retrieved from: http://www.adaptivepath.com/publications /essays/archives/000361.php. (2004)

[6] http://www.flickr.com an image and video hosting web-site and online community platform.

[7] Olson, H.A. and Wolfram, D. Indexing Consistency and its Implications for Information Architecture: A Pilot Study. ASSIS&T Information Architecture Summit, Vancouver, British Columbia, Canada, March 23-27, 2006. Retrieved from: http://www.iasummit.org/2006/files/175_Presentation_Desc.pdf (2006)

[8] Golder, S. and Huberman, B.A. Usage Patterns of Collaborative Tagging Systems. Journal of Information Science, 32(2): 198-208. (2006)

[9] Guy, M. and Tonkin, E. Folksonomies: Tidying up Tags? D-Lib Magazine, January 2006, 12 (1). Retrieved from: http://www.dlib.org/dlib/january06/guy/01guy.html

[10] Delicious, a social bookmarking web service for storing, sharing, and discovering web bookmarks. http://delicious.com (2006)

[11] Begelman, G.; Keller, P. and Smadja, F. Automated Tag Clustering: Improving search and exploration in the tag space. WWW2006, May 22–26, 2006, Edinburgh, UK. Retrieved from: http://www.rawsugar.com/www2006/20.pdf (2006)

[12] Brooks, C.H. and Montanez, N. Improved Annotation of the Blogosphere via Auto tagging and Hierarchical Clustering. WWW 2006, May 23–26, 2006, Edinburgh, UK. Retrieved from: http://www2006.org/programme/item.php?id=583(2006)

[13] Shirky, C. Ontology is Overrated: Categories, Links, and Tags. In Economics & Culture, Media & Community. Retrieved from: http://www.shirky.com/writings/ontology_overrated.html (2005).

[14] Ning , W. Textmining and Organization in Large Corpus. MSc thesis in Computer System Engineering. Department of Informatics and Mathematical Modeling (IMM), Technical University of Denmark (DTU). Retrieved from: http://www2.imm.dtu.dk/pubdb/views/edoc_download.php /4338/pdf/ imm4338.pdf (2005).

[15] Krause, B., Hotho, A. and Stumme, G. A Comparison of Social Bookmarking with Traditional Search. Knowledge & Data Engineering Group, University of Kassel Wilhelmsh¨oher Allee 73, D-34121 Kassel, Germany Research Center L3S, Appelstr. 9a, D-30167 Hannover, Germany. C. Macdonald et al. (Eds.): ECIR 2008, LNCS 4956, pp. 101–113, Springer-Verlag Retrieved from: http://www.kde.cs.uni-kassel.de/hotho/pub/2008/ecir2008krause.pdf (2008).

[16] Technorati, an Internet search engine for searching blogs, competing with Google and Yahoo. http://www.technorati.com/

[17] Data retrieved on 6th August 2008 from delicious at: http://delicious.com/url/dfac2ac2f8102bbfbd4ef18a247d74cb More than 61,100 tags assigned at the time.

[18] http://delicious.com/tag/crap

[19] http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html

[20] Embley, D.W. , Olive, A. and Ram, S. Csaba Veres Concept Modeling by the Masses: Folksonomy structure and Interoperability. (Eds.): ER 2006, LNCS 1215, pp.325-338, 2006. Springer-Verlag Berlin Heidelberg (2006)

[21] http://processing.org

[22] Mica, P. Ontologies are us: A unified model of social networks and semantics. Web Semantics: Science, Services and Agents on the World Wide Web 5 (2007) 5–15 Retrieved from: http://www.marioux.com/Alex/Social%20Networks/Structure%20des%20r%E9seaux%20sociaux/Ontologies%20or%20us.pdf

[23] Wetzker, R., Zimmermann, C., Bauckhage, C. Analyzing Social Bookmarking Systems: A del.icio.us Cookbook. (2008) Retrieved from: http://robertwetzker.com/wp-content/uploads/2008/06/wetzker_delicious_ecai2008_final.pdf

[24] Hotho, A., Jaschke, R., Schmitz,C., Stumme, G.. FolkRank: A Ranking Algorithm for Folksonomies Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.5271&rep=rep1&type=pdf