

# Tag Based Meta-Search for Browsing the Web: The TicTag Application

**Christina Papakonstantinou**

Leiden Institute of Advanced  
Computer Science  
Niels Bohrweg 1  
2333 CA Leiden, NL  
+31 (0)63 41 06 553

c.papakonstantinou@umail.leidenuniv.nl

**Ioannis Panagiotou**

Leiden Institute of Advanced  
Computer Science  
Niels Bohrweg 1  
2333 CA Leiden, NL  
+31 (0)62 85 93 644

i.panagiotou@umail.leidenuniv.nl

**Fons Verbeek**

Leiden Institute of Advanced  
Computer Science  
Niels Bohrweg 1  
2333 CA Leiden, NL  
+31 (0)71 52 75 773

fverbeek@liacs.nl

## ABSTRACT

Social networks, such as bookmarking services, offer new ways of storing, categorizing (tagging) and accessing information, competing the conventional ways of search-engines. But in what ways are these social structures changing the search-results in the benefit of the searcher and how can we visualize these differences to improve user awareness on the differentiations of the tactics? The set of tags used collectively in such a social network to describe and categorize web-content is called folksonomy and reflects a common understanding of what content falls under what tags. In this paper we present a novel algorithm that constructs and compares two sets of tags for each of the search-tactics to tackle the problem. We furthermore, have built an application that combines the results and presents them to the user in a comparative layout very different from the usual.

## Keywords

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

## INTRODUCTION

Browsing and searching activity in the Web has gone through big changes -and facilitations- in the last years. Under the need of more accurate and user related search-results some projects have turned to social structures of sharing information. Blogs and bookmarking services fall under this category. A usual way to present the underlying information under each of these services is the usage of tag clouds; which function as a means to form a general impression of the content at a glance. The importance of tags and tag-clouds, as well as that of the emergent structure they form called Folksonomy, is discussed in paragraph *Tags, Tag-Clouds and Folksonomies*.

But in what ways are these social structures changing the search-results in the benefit of the searcher and how can we visualize these differences to improve user awareness on the differentiations of the tactics?

In order to have a first idea we would like to first ‘scan’ the differences of results derived by the two search

tactics: one represented by a conventional search-engine and one represented by a bookmarking service. We developed an algorithm – described in paragraph *TicTag-Algorithm*- which performs both search-strategies when given a search term in order to compare the results; we present the results in paragraph *Experiment Results*. The comparison juxtaposes the results returned for a set of given words and their derived tag-clouds of related words.

Based on the above algorithm we have built an application which allows users to search through a combination of the methods –discussed in paragraph *TicTag Application*- and browse the web through the results’ derived tags and tag-clouds. The application can be found online at the address [www.tictag.cc](http://www.tictag.cc).

The following of the paper concludes with a discussion on the findings and the directions we believe future work should focus towards.

## TAGS, TAG-CLOUDS AND FOLKSONOMIES

A *tag* is a one-unit word that describes a piece of information. In the internet-practice tags have been put in wide use in blogs and bookmarking services for a visual summary of the site’s content.

Tags in blogs are submitted by the author of the blog entry. In bookmarking services, however, tags are submitted by any user who has bookmarked the content; and so, they present a particular interest as they are collectively used and agreed upon within the context of social sharing of information. The collective organization of content under tags -also called a *folksonomy* -has challenged pre-existing categorization schemes due to its many advantages for the purpose of categorizing web-content, but has its limitations as well.

## Folksonomy

Folksonomy is the categorization system that is formed by the total set of tags in a tagging system and their corresponding resources; it can be therefore thought as the collection of all users’ personal categorization schemes<sup>7</sup> in a bookmarking system.

<sup>7</sup> 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

It has been often compared with ontology and taxonomy, but it has different rules and properties. 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 they are loose in the semantics of the terms, thus disambiguation often occurs. Because folksonomy is loose and allows content to be categorized in more than one class, there is no inferred hierarchy of the tags within it.

### Limitations

The main critique on tags and folksonomies is that they often allow for disambiguation. In the social context of a bookmarking community tags serve both retrieval by the individual and the tagging community; they should therefore comply with both the user's and the society's norms.

The duality of the goal is thought to bring about many categorization problems, such as *synonymity*, that is when different persons use same tags 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 persons use different tags 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. Moreover, because tagging systems do not usually offer synonym control, the tags 'Mac' and 'macintosh' may exist in the system as different tags, although they refer to the same thing. Plurals and singulars of the same word may also be 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. On the other hand, it is difficult to force users to always follow strictly the community norm in their tagging strategies. This would make the system inflexible and ultimately unusable. It seems that it this looseness that makes folksonomies powerful.

### 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.

---

inherent dynamics of this tag-set immediately emerge from the fact that folksonomies are built within a process that involves human critical thought, social interactions and temporal events that trigger off patterns in tagging behavior.

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 they are able to fast enough catch-up with the vocabulary changes, which makes tagging suitable as a dynamic 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’.

In contrast to keywords and publishers' metadata<sup>8</sup>, 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].

Research on folksonomic tags has shown that there exist some patterns in their usage. For example, users tend to assign broader terms for tagging their resources [8] [12]. Both types are necessary: broad tags are better for browsing tasks whereas narrow tags are more suitable for querying tasks [3]. Broader tags have also been found to precede the more narrow tags when a user labels a resource.

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] and so, 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 and facilitate navigation of the user [3][11][12] through tags. 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].

---

<sup>8</sup> 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.

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 with 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.

### **Tag Clouds**

Tag clouds are the most predominant way of representing a tag set. Tag clouds are weighted lists of tags often presented in alphabetical or by frequency order.

A common problem of tag clouds is that a few tags and their related tags tend to dominate the whole of the tag cloud, making this way less often tags to vanish. However this can be remedied by clustering. Research on the latent structure of folksonomies has also resulted in suggestions of advanced tag-cloud layouts that will improve navigation through them.

However this is obviously a problem when the size of the tag cloud is constrained, whereas in a tag cloud of unconstrained size all tags can be included, and readability can still be achieved with appropriate use of the tag's weights.

As indicated by their extensive usage, tag clouds are a good tool for giving a quick visual impression on some content. As such we will use the set of tags which accompany a set of URLs, so as to compare the ones coming from a search in a traditional search-engine and those that are retrieved by a search in a bookmarking service.

Bookmarking services allow for overview of the tag label of a bookmarked URL, whereas search engines do not embed this option, and so we will have to construct the tag set of its URL results from the available resources.

### **USER INTERFACE AND ALGORITHM OF THE TICTAG**

Based on the above research on the ways to interact and navigate through web content using tags and folksonomies we have built the TicTag application whose UI design and functions facilitate a user's tag based search.

#### **User Interface**

The basic layout of the application is very close to a search engine's layout structure. There is a text field on the top of the screen where users can enter their search term and the main space below the text field is empty for the results to be returned. This facilitates the user to get

accustomed to the application as a search engine -using the common perception of search engines- and prevents confusion.

As soon as the search term is input to the system it is treated both as a tag itself and as a keyword. The algorithm as we will see in the following paragraph, gathers both the URL results from the traditional search engines and the related tags/words/concepts from the bookmarking service.

The results returned by TicTag however are not directly the retrieved URLs but the related tags of the search term instead, thus its related concepts. After the search is finished the user is returned with an interactive tag cloud derived from his term. In the center of the screen lies the input term and around it the related tags organized in clusters based on their URL source, i.e. to which web site they refer to.

Because it is usual that web sites have themselves a somewhat specific content, tags referring to the same site inherit this relation and reflect it to their cluster.

Tags are moreover interactive. They appear in circular form and can be moved around, interact with other tags in the space and be organized at the will of the user. This gives the user the joyful feeling of the application being a game and helps engagement to tag based search.

The overview of the derived tag cloud gives the users an immediate idea of towards which direction he/she will further refine his/her search. The URLs, which are the actual request and final source of information the user seeks for, are organized under these concepts/tags of the tag cloud and so, it is less likely that the user will follow the wrong link. This is because synonymity in the application occurs in the form of synonymous tags and so, it is easy to distinguish between relevant and irrelevant tags/concepts without first being driven to the wrong websites. Irrelevant tags can also be deleted easily by the press of a button.

Once the user has found the related tag of the tag cloud that is closest to his interpretation of the search term he/she is able to select it by click. Once a tag is selected, the URLs organized under it appear in the screen in a radial form around the clicked tag. This increases the perceived relation of the tag to its results and moreover, it avoids the false impression of hierarchy that leads to believing that top results are better or necessarily more relevant, as it often happens in results presented in lists (traditional search engines). In support to that we will later see in the Usability Evaluation paragraph, that there exist search terms for which some traditional search engines return no related tags at all; which means that the search engine's top URL results have either never been tagged, thus bookmarked, or have only been attributed the search term as a tag; the latter case being less probable due to the inter-indexer inconsistency within a tagging community. The URLs of the selected tag are interactive and can be clicked in order to follow the link to the corresponding website.

Furthermore each of the related tags in the tag cloud is itself a refinement of the initial search of the user. For

example, if a user inserts the term “design” and the tag cloud returned by the application contains the tag “graphics”, then a click on the tag “graphics” will present the user with the URLs retrieved by a search on both the terms “design” and “graphics”, in accordance with the user’s expectation of refinement through selection of tags.

The TicTag application offers in addition ways for comparative searches between more than one search terms whose tag clouds merge in the results space.

A sidebar is implemented on the right side of the screen where advanced options are offered, such as the color code of each of the searches, options to hide/show some of the derived tag clouds, an option to see the indication of the source of each tag and last an option to separate the different tag clouds in the screen and an option to only show the overlapping tags of the different tag clouds.

So, a user can in many ways find his way navigating through and interacting with the tag clouds. For the best guidance of the user to all these features, ‘text bubbles’ have been implemented throughout all the search steps to help the user to better exploit the capabilities of the application.

**TicTag Algorithm**

We will now describe in a formal manner the way the Tic Tag retrieves results.

Suppose a search term X is input for search by the user.

X is submitted to both traditional and social search engines and returns two sets of URL results: {S} is the set of URL results returned by the social service and {T} is the set of URL results returned by the traditional search engine.

$$\{S\} = \{S_1, S_2, S_3, \dots\} \text{ and}$$

$$\{T\} = \{T_1, T_2, T_3, \dots\}, \text{ where } S_i \text{ and } T_i \text{ are URL results.}$$

It is important to note that the engines not only work in different ways but on different data sets as well. Search engines use robots that scan most of the web and store their findings, whereas bookmarking services accept user input data; and therefore have a smaller but more selective content. However these differences are of no particular interest to the user who is in seek of information using solely either one of the services or both in combination. What we are interested in here, are the tags accompanying these results.

Each of the  $S_i$  are more than likely to have been tagged by users that bookmarked it when they added it into the service’s database (note that it is possible to bookmark a URL without attributing a tag to it, but this is unusual especially to the mostly bookmarked -thus highly rated-results).

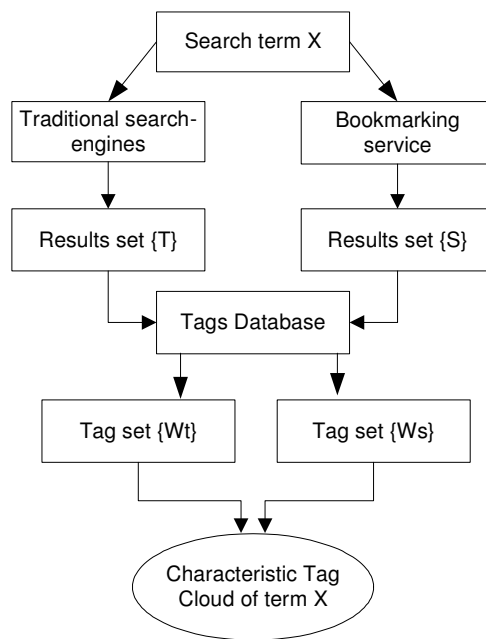
Each  $S_i$  can therefore point to a set of tag-words  $\{W_{S_i}\}$  that describe it. The union of tag sets of all URLs in  $\{S\}$  forms the set  $\{C\}_S$ , which is part of the tag cloud to be derived of the word X.

We now need to find the corresponding tag cloud  $\{C\}_T$  derived by  $\{T\}$ , i.e. the tags that describe the set of URL-results of the traditional search engine. Since tagging is not supported by conventional search engines we go through the bookmarks’ database.

It is has been found that the top results of search-engines are more often included in the database of bookmarking services as well [15]. Some of them have therefore been tagged, as well, and so we can collect the tags assigned to them to form the tag set  $\{C\}_T$  of the word X.

The union of  $\{C\}_S$  and  $\{C\}_T$  forms the *characteristic tag cloud* of the initial search term X (cf. Diagram 1). The comparison of the tag sets  $\{W_t\}$  and  $\{W_s\}$  as to their contribution to the tag cloud can then be measured in terms of size and enhancement of topics covered.

**Diagram 1**



Note that the characteristic tag cloud is dependent on tags database from which it retrieves the tags assigned to the corresponding URLs. This can be a single bookmarking service’s database or more than one’s. The bigger the resource of tags we use, the more accurate the comparison between a traditional search engine’s results and a bookmarking service’s results. However, the lack of tags for a URL can also give a hint either on users’ appreciation of the URL’s content or their retrieval habits, since these URLs are either not chosen for bookmarking by any user of the tagging community, or are thought easily retraceable as they may steadily appear in top results of a search.

**USABILITY EVALUATION**

Four small experiments were conducted to scan the contribution of tags by search engines to the tag set of a search term.

**Materials and Methods**

The experiments were carried through in August 2008 using two commercial search engines: Yahoo! and Google. One bookmarking service was used and that was Delicious. The tags chosen for the experiments were chosen by criteria in respect to the experiment’s goal.

At a first stage, a pair of synonym tags (Information visualization and Infovis) and one of opposite words (War and Peace) were chosen to check on their respective clouds overlapping.

At a second stage, a pair of non-often tag words (Giannis Panagiotou and TicTag) and a set of popular words (world, freedom, blog and design) were chosen and the respective contribution of the search engines was counted. The words blog and design were chosen as two of the most popular amongst the Delicious tags.

### Results

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 tag clouds of synonyms

	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.

### 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.

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”.

**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	

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’<sup>9</sup>.

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.

### 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 the name of one of the authors {‘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.

<sup>9</sup> Famous actor playing in a movie called ‘War’.

**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 {'nisia.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.

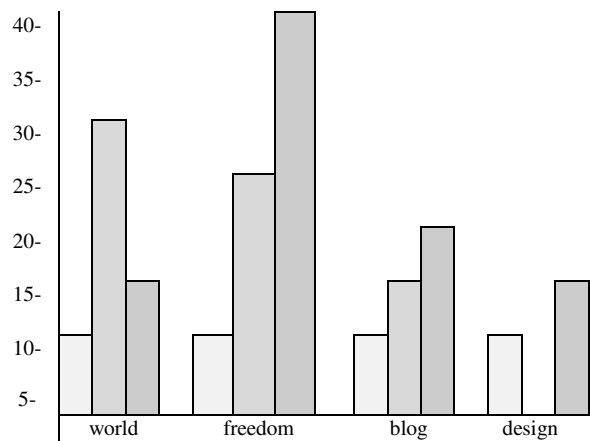
In the 'TicTag' example, we see that Delicious offers 5 tags: i.e. {innovative, application, dev, keywords, creative}, Google returns none, but Yahoo! contributes 10 more tags: i.e. {generative, code, visualization, processing, programming, design, processing.org, art, blog, blogs} approximating the 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.

#### 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.

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).

**Table 4:** Comparison of relative contributions



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 he 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.

#### TIC TAG APPLICATION

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<sup>10</sup> 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 called its *characteristic tag cloud* (cf. Fig. 1). This tag cloud consists of all suggested tags from Delicious plus

<sup>10</sup> 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.



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<sup>11</sup> 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. Therefore the described application's results are more likely to also cover the searcher's interest/interpretation of the search term, thus providing him with the most relevant results.

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.

## CONCLUSIONS

The implementation of an application supporting a tag based mash-up search of the web in an interactive and engaging way has given us the opportunity to experiment on users' expected behavior of such a system as well as the reflections of their mental model.

Users can get quickly familiar with the idea of searching and refining the search by navigating through tag clouds due to their familiarity with bookmarking services that have become very popular. Moreover the layout of the application gives a quick overview of the topics covered by the results – a feature that has made popular the tag clouds which are now used for an impression of the

content of a site, either that be a blog, a portal or a bookmarking service.

However, the enhancement of the tag set by injection of alternatively retrieved tags has given space to more concepts to appear in the results, which leads to a more accurate tag based description of what the user would find by using either of the known search tactics alone.

Last the interactivity of the resulting tag cloud and the physical motion of the tags in the screen can facilitate the engagement of the user and help his browsing and refining toward his desired results, or even the browsing towards new concepts he was unaware they related to his/her search term but they are used as such by a tagging community, allowing also for some unintended results being discovered by serendipity.

We therefore conclude that a user engaging tag-based mash up search engine can not only provide a good way to browse the web by searching through concepts, but also to provide an overview of how the search term is used in the internet world and what connotations and related concepts they encompass.

## 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 StumbleUpon 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.

## ACKNOWLEDGEMENTS

We would like to thank the Media Technology Department of the Leiden Institute of Advanced Computer Science for the support to our project.

## REFERENCES

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

---

<sup>11</sup> Conflicts in Ossetia were a major topic in the world news in August 2008, when experiments were conducted.

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](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](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\(2006\)](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\\_ouerrated.html](http://www.shirky.com/writings/ontology_ouerrated.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](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 Wilhelmshöher 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., Bauchhage, C. Analyzing Social Bookmarking Systems: A del.icio.us Cookbook. Retrieved from: [http://robertwetzker.com/wp-content/uploads/2008/06/wetzker\\_delicious\\_ecai2008\\_final.pdf](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>