Journal of Applied Computing Research, 1(1):48-60
January-June 2011
© 2011 by Unisinos - doi: 10.4013/jacr.2011.11.04

DJINN: Implementation and evaluation of implicit social bookmarking paradigm[1]

Bruno Merlin
Universidade Federal do Pará. Tv. Pe. Antônio Franco, 2617
68400-000, Cametá, PA, Brasil. bruno.merlin@gmail.com

Cleidson R.B. de Souza
IBM Research Brasil. Rua Tutóia, 1157, 04007-005, São Paulo, SP, Brasil.
cleidson.desouza@acm.org

Abstract: *Social bookmarking* is a popular way to share and publish bookmarks. The growth of the *social bookmarking* community is creating a parallel resource for web searches. In this paper we present an approach called *implicit social bookmarking,* that creates a bridge between classical web searches initiated in a search engine and search in social bookmarks. Our approach allows a user to implicitly contribute to a *social bookmarking* system and benefit from other user searches, therefore improving the accuracy of results of web searches. This paper details our approach, presents a prototype implementation of the approach called DJINN, and presents two evaluations of the concept. We conclude with a discussion of our results and directions for future work.

Key words: social bookmarking, collaborative filtering, tagging.

Introduction

Providing ways to retrieve content and to increase the reliability of search results have always been an important challenge of the web boom. The fast growth of available content on the Internet led to the rise of search tools. The goal of these tools is to help users to find and retrieve relevant content based on their search criteria. The first step towards this direction has been the creation of search engines. The first generation of indexation engines analyzed the semantic content of pages and created an automatic indexation of sites. However, rapidly, the exponential proliferation of contents required complementary strategies to improve page indexation. An initial response to this issue has been the creation of the tag “<meta name=”keywords”>”. This tag enabled content publishers to define the semantic information representative of the page content. However, search engines were quickly confronted with commercial abuse (Heymann *et al.*, 2007). They required additional clues to trust the published information and to evaluate the page interest. As a consequence, the first referencing system, implicitly collective, appeared: the PageRank algorithm [US 6.285.999] used by Google. This algorithm analyses the link contained in a page to increase the rank of the site being targeted by the link. Thus, by publishing its content, a site contributes to the evaluation of other sites. The aggregated contribution of all other sites determines the ranking of a particular site.

In addition to the PageRank, new collaborative referencing techniques and services appeared. Among them, *recommenders* (Chang

[1] This article is an extended version of “Implicit Social Bookmarking to Improve Web Searches”, originally published in *Simpósio Brasileiro de Sistemas Colaborativos* (SBSC), 2009, p. 36-43. DOI: 10.1109/SBSC.2009.8.

Lee and Kwon, 2008) establish a user's profile to recommend information somewhat similar to this profile. The electronic bookstore is a classical example of *recommenders* (Mooney and Roy, 2000): a user buys a book, (i) all the books previously purchased by the user enables the system to determine his profile, i.e., the user's preferred themes; (ii) the system recommends other books based on the user profile; and, (iii) the system recommends books bought by other users with similar profile. Another popular collaborative technique is the *social bookmarking* (Dourish and Chalmers, 1994), which consists in explicitly sharing tagged references. Specific bookmark sites enable users to retrieve bookmarks based on the available tags.

Previous work studied the users' motivations for *social bookmarking* (Thom-Santelli *et al.*, 2008; Ames and Naaman, 2007). Additional work discussed the reliability of the links returned by *social bookmarking* sites and concluded that *social bookmarking* is a pertinent resource to improve page indexation and web searches (Heymann *et al.,* 2008). However, some researchers indirectly questioned the future impact of the social motivations on the link reliability: they identify spamming as a main threat for *social bookmarking* (Heymann *et al.*, 2007; Krause *et al.*, 2008; Koutrika *et al.*, 2007).

Along the same lines, we argue that the impact of spamming will be reinforced by the low number of bookmark publishers in comparison with the high number of potential bookmark consumers. To preserve the neutrality of the published links, we argue that it is necessary to increase the number of publishers to balance the ratio publishers/consumers. In response to this issue, this paper proposes the concept of *implicit social bookmarking*. We call our approach *implicit social bookmarking* because it is based on an *implicit* users' contribution to *social bookmarking* and use of it, i.e., the user does not need to take any additional step for that to work (Grudin, 1988). We also present an implementation of this concept: DJINN. DJINN deduces data needed for *social bookmarking* from the users' interaction and navigation and augments a user's search results with information retrieved from *social bookmarking*. We finally report two initial evaluations of the concept.

The rest of the paper is organized as follows. In the first section, we discuss the pertinence of the data provided by *social bookmarking*. Then, we present the concept of *implicit social bookmarking*, which is followed by DJINN: a prototype that implements the concept of implicit *social bookmarking*. The fifth and sixth sections present two initial evaluations of the concept. The next session discusses future work aiming to improve the concept based on the results of our evaluations. Finally, the last session presents our final remarks.

Social bookmarking

Social bookmarking enables users to store, manage and, more importantly, share bookmarks (Dourish, 1994). Bookmarks are organized using tags. Tags are freely chosen keywords that can be assigned to bookmarks and shared among different users. Web sites, such as del.icio.us [http://delicious.com/] or BlogMarks [http://blogmarks.net], collect the bookmarks and offer tools to perform bookmarks searches in the collected tags.

(a) Impact of social motivations on social bookmarks reliability

One of the questions that drive this research is the following: can we improve web search by social bookmarking? Heymann *et al.* (2008) explain that the relevant quality of links and tags provided by a web site such as del.icio.us enables the identification of relevant links; and, in some cases, the identification of relevant links not found during a classical web search. However, they also observe that del.icio.us covers only a small part of the Web. According to them, social bookmarking has a great potential to grow. This expectation raises two important questions:

- If *social bookmarking* becomes a major way to access information, will link credibility decrease because of the publishers, evangelists, leaders (Thom-Santelli *et al.*, 2008), spammers (Heymann *et al.*, 2007), and other people with commercial interests? This question arises from the fact that *social bookmarking* is based on a recommendation relationship from *one* information publisher toward *several* consumers ("one toward N") and remains efficient because, in the context of a community, it is possible to trust the publisher. To maintain the relevance of bookmarks toward the growth of participation, Pereira (Pereira and Silva, 2008) suggests the creation of an author trust rank.

- *Social bookmarking* is a context where a publisher recommends content (in this case, a bookmark) to a large number of potential consumers. Based on that, can the concepts of social navigation (Dourish, 1994) and collaborative filtering (Goldberg *et al.*, 1992) (e.g. implicit filtering) help *social bookmarking* to change the relationship "one publisher toward several consumers" to "*several* publishers toward *several* consumers" (N toward N)?

A central problem of tagging, and other collaborative systems, is the balance between who performs extra work and who benefits from this work? (Grudin, 1988). As a matter of fact, tagging is time-consuming and requires a personal, social or economical motivation to be performed. Beside contexts where the user is the main beneficiary of the tagging (Ames and Naaman, 2007), Thom-Santelli *et al.* (2008) identified several social roles of tagging users: community-seekers, evangelists, community-builders, publishers and small team leaders. In general, Thom-Santelli's work reports that active social taggers tag information to increase their visibility and/or promote specific content. These social motivations also raise the question of spamming. In fact, many researchers identify the fight against spam in social tagging as a main stake for the social bookmarking issue (Heymann *et al.*, 2007; Krause *et al.*, 2008; Koutrika *et al.*, 2007). We discuss this issue below.

The relationship "one publisher toward several consumers" when is out of a restricted and controlled community might increase the impact of a spammer. In this context, a spammer can easily and artificially increase the visibility of the content (s)he wants to promote[2]. Because of this, social bookmarking requires a high level of moderation and its growing popularity may decrease. As a consequence, we argue that systematic use of social bookmarking to improve web searches may rapidly become problematic.

Our hypothesis is that, in a larger and non-controlled context, we must balance the number of publishers and consumers to change the relationship "one toward N" into a relationship "N toward N". A massive participation of consumers would enable a recommendation from "consumers toward consumers" instead of the traditional "publisher toward consumers" (Ruffo and Schifanella, 2009).

In fact, in order to contribute to improve web search efficiently, the user should make a systematic evaluation and tagging of the visited page. However, an explicit contribution of the user, an *active* collaborative filtering, is not very realistic (Goldberg *et al.*, 1992; Maltz and Ehrlich, 1995). Therefore, in this paper we ask whether it would be possible to achieve a *passive* collaborative filtering.

(b) Other social bookmarking enhancements

Recent work, such as the BibSonomy (Hotho *et al.*, 2006) and others (Hammond *et al.*, 2005), concentrate their effort on the bookmark publication task and the underlying model to select and share the tag vocabulary (the so-called folksonomy[3] (Sturtz, 2004; Mathes, 2005)). Other researchers focus on the controversial tag clouds (Sinclair and Cardew-Hall, 2008), which are visual representations of a folksonomy that help users to select search tags. Frequently criticized to be a fancy but useless representation, tag clouds have been more formally evaluated in recent experiments (Zeldman, 2005; Rivadeneira *et al.*, 2007) and showed promising results for general searches (Riddle, 2005).

Even if social bookmarking approaches, such as del.icio.us, are now better integrated to the user's desktop (add-ons for Firefox enabling to tag a page in a few clicks), searching in social bookmark sites remains a task performed independently from a standard web search. Furthermore, the results provided by a search engine, such as Google, do not benefit from social bookmarking information. The user must engage in two different search efforts and separately analyze the two search results. We address this limitation with our approach, which is described in the next section.

The approach: Implicit social bookmarking

We propose the concept of implicit social bookmarking whose goal is to improve web

[2] For instance, a recent problem occurred on YouTube when anti-social taggers published pornographic videos in adolescent search fields.
[3] See Wikipedia (s.d.) and Van der Wal (2007).

searches. This approach is a blend of social bookmarking and passive collaborative filtering (Maltz, 1995). When the task of social bookmarking needs specific actions from the user, the implicit social bookmarking:

- Automatically deduces information required for social bookmarking. This information is extracted from the user's interaction during other tasks (web searches, visits and re-visits to pages, "personal" bookmarking, etc); and
- Automatically extracts information from social bookmarking without requiring an explicit request from the user.

Each one of these steps is explained in details below.

(a) Deducing information from user's behavior and interaction

A simple example can be used to illustrate this part of our approach: in a context of evaluating sellers of a site such as e-bay, if a specific user buys several items from the same seller, even if this buyer does not record an explicit evaluation, the user's loyalty can be interpreted as an implicit satisfaction with that particular seller. This user behavior allows one to make inferences about the user's appreciation of the seller. The same kind of information is extracted by recommenders to improve a user's profile and suggest articles related to the user's preferences.

Similarly to a recommender system, one of the main aspects of the implicit social bookmarking approach is extracting useful information from the user's activity. In this case, it is necessary to be able to evaluate two types of information:

- The relevance of a link for the user; and
- The semantic information necessary to tag the links and make an automatic categorization of bookmarks (Staff and Bugeja, 2007).

(i) Evaluating the relevance of a link

The relevance of a link can be calculated by using information from the user's navigation. Visiting published pages on the same web site, re-visiting a page (Bilenko and White, 2008), visiting links provided by a page and, finally, explicit actions such as bookmarking a page or sharing the bookmark with someone else are different ways to express an interest for a page/site. By monitoring these actions, we can interpret the user's behavior toward a page/site to implicitly attribute a mark for the page in the context of that particular web search. The sum of the marks attributed by all users' behaviors is used to augment the ranking of the page/site.

(ii) Tagging the links

Several sources of information can help us to obtain semantic data to characterize an ongoing search. One of the more relevant is the input made by the user in the search fields and search toolbars. Even if this information must be consolidated because of orthographic errors, it constitutes an important source of information to identify the user's search themes. A second relevant source is information about the visited page itself, for instance page keywords and titles that can be automatically extracted. However, this information must be filtered to counterbalance the strategies used by publishers to increase their page visibility. An example of such strategy is to include popular words among the keywords and title to be largely referenced. Nevertheless, this practice decreased since the use of the PageRank algorithm. Finally, the last relevant source of information could be the page content itself. However, in order to use this information, we should widely analyze the vocabulary to extract pertinent semantic data.

(b) Providing information from social bookmarking

Currently, searches within social bookmarks web-sites and search engines are independently performed, i.e., a request does not simultaneously provide results from a social bookmarking web-site and from a search engine. The second principle of implicit social bookmarking is to use the social bookmarked data to make the search performed by the search engine more relevant.

As meta-search engines combine the search data from different search engines, an implementation of the implicit social bookmarking concept must combine the result from one or more search engine(s) and the results collected from social bookmarks. The implementation of this approach is described in the following section.

DJINN: An implementation of implicit social bookmarking

DJINN is an implementation of the implicit social bookmarking approach. It tackles the aspects of this concept described in the previous section. First, it automatically analyzes users' navigation and interaction to retrieve information about users' interest in a page. Second, it automatically extracts semantic information from the user's navigation to tag the interesting pages. These two pieces of information are used to create tagged bookmarks that are shared between users. Finally, it augments the search results performed in traditional search engines with information from shared bookmarks.

(a) DJINN main architecture and technical issues

DJINN is composed of three modules (cf. Figure 1):

- DJINN Proxy – The heart of DJINN is a proxy service developed in Java. This proxy collects and manages tagged bookmarks, receives and propagates HTTP requests from browsers, and reformulates the results of search queries made in search engines (currently, Google and Yahoo) based on the bookmarked pages;
- Firefox Add-On – A second module is integrated in Firefox as an add-on. The add-on, developed in XUL and JavaScript, collects information about the user's interaction with the browser: tabs opened, closed and focused, data input in search fields, pages explicitly bookmarked in the personal user bookmarks, pages retrieved using bookmarks or history, information on visited and revisited pages (such as keywords and titles), etc; and
- Context Manager – The Context Manager, developed in Java, collects the information from the previous modules and manages "search contexts" for each open tab or group of tabs. It analyses information from input data and visited pages in order to extract tags. The relevant couples tag/page or tag/site are marked and propagated to the proxy.

The Firefox Add-on and the Context Manager constitute a DJINN Client. Every user gets an instance of the DJINN client. The proxy is shared between all users and receives all the HTTP requests, all the bookmarks from every context manager, handles every search query, and reformulates every search response.

The Firefox add-on communicates with the Context Manager through the file system (Figure 2). The Context Manager communicates with the Proxy by HTTP requests. The presence of the proxy is transparent for the browser such as any other proxy.

(b) Search contexts inferred from user browsing and interaction

A search context is a set of weighted tags and a set of weighted pages. Each page visited during a search gets a mark between 0 and 10 in function of the user's behavior toward the page. A page briefly consulted and never revisited gets a mark of 0. A page visited for

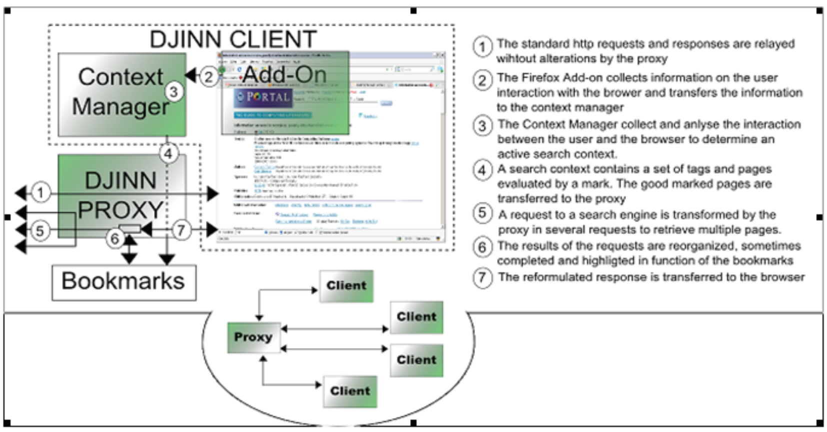

Figure 1. DJINN architecture.

a long time gains a point. The visit of pages from the same site increases the page mark by 1. The two first times, a revisited page gains 1 point more. We considered the first revisit as a confirmation of the user's interest for the page, and the second as a page status change: the page become a reference on the subject. Consequently, further revisits do not increase the page mark because they only confirm this status. A bookmarked page (bookmarked in the browser) gains 2 points and a page retrieved from bookmarks or history receives two more points the two first times.

The elements required to mark the pages are extracted by the Firefox Add-On. In addition, handlers (cf. Figure 2) retrieve the tab and window focus, the history and bookmark items pressed (and more generally how the user access to page: address input, link, back/next button etc.) and the addresses of visited pages.

We are currently extracting information from four sources: (i) the information input by the user to make search queries ("search", "q", "query", etc, fields of web pages and text input in search bars), (ii) the text input to find a word into a loaded page, (iii) an analysis of the text contained in the title of the visited pages, and (iv) the keywords (tag <meta name="keywords">) contained in the head of the pages. An algorithm combines all this information and it weights the words, analyzes their frequency, and removes words that are not relevant, in addition to other natural language processing approaches. The algorithm provides between 0 and 7 words to characterize the search. At the end of the search, the pages with a mark up to 2 are automatically tagged with the word characterizing the search. The tuple bookmark/mark/tag is transferred to the proxy.

(c) Bookmark global mark and inertia

The proxy maintains a global mark for the each bookmark/tag information. This global mark depends on the mark given by the contributors to the page and the ratio number of contributor/number of potential system users. This ratio defines the inertia of the system. The inertia enables one to moderate the influence of the first contributors: it can be low for a small homogenous group (a search team for example), medium for a larger community or high for open use.

(d) Beginning a new search and initiating new search contexts

As we could observe it in preliminary our observations, users frequently use several tabs and/or windows simultaneously during a web search. Different reasons lead one to open these new tab/windows: for instance, an explicit action of the user opening a default page, a click opening a link into a new tab/windows automatically or another explicit action of the user opening a link (or the current page) in a different tab/window. These different tabs lead to different and parallel navigation flows. These navigation flows may contribute to the same search and may feed the same Search Context. However, users may use simultaneously different tabs to manage other searches, access different information (email, social network portals, etc). In these contexts, the visited pages must not share the same search context.

As a consequence, the Context Manager implements an algorithm to determine: which tab/windows share the same search context; when a search is concluded and a new search

```
window.addEventListener (
  "focus",
  function () {
          sendToContextManager ("FOCUS", currentTabRef);
     },
     false
);

function sendToContextManager (action, tabRef) {
   var text = tabRef + "\n" + action + "\n" + time ();
   var file = Components.classes
           ["@mozilla.org/file/local;1"].createInstance
           (Components.interfaces.nsILocalFile);
   file.initWithFile (exchangeDirectory);
   file.append (_index + ".adr");
    _index ++;
    var fStream = Components.classes
            ["@mozilla.org/network/file-output-stream;1"]
        .createInstance
        (Components.interfaces.nsIFileOutputStream);
    fStream.init (file, 0x02 | 0x08 | 0x20, 0664, 0);
    fstream.write (text, text.length);
    fStream.close ();
}
```

Figure 2. JavaScript code to capture and transfer browser focus to Search Context.

begin. This algorithm is based on two factors: the user's interaction and the information provided by the user. The Add-On collects information about the interaction (tabs/windows opened, action opening the tab/window). Tabs/windows explicitly or automatically opened from a page are associated to the same search context. Pages visited using these tabs contribute to enrich the semantic context and to mark the pages and words. A tab opened independently from a page starts with a different search context (when the user opens a default page). The search context may be merged with another search context when the two search contexts get a high semantic correlation.

We consider that a new search starts in a tab when the user goes back to the home page or makes a search query without correlation with the semantic data present in the current search context. A search context is closed when every tab is closed or a new search starts.

(e) Integrating implicit social bookmarks in search query results

The DJINN Proxy propagates the HTTP requests and responses, and filters the search query made in the search engine Yahoo and Google. It adopts a specific handling to these requests that is described below.

When the proxy receives a query, it retrieves the first page of results and a maximum of the following result pages during a time interval of 1.5 seconds. Simultaneously, the DJINN Proxy retrieves the link tagged by the word used for the query in the implicit social bookmark database. The delay of 1.5 is arbitrary, it is supposed to enable the proxy to retrieve the 6 first result pages (on average) for a query and not to delay the response. The result pages retrieved are temporarily stored in a cache. If the user requests the second, third, etc., result pages, DJINN retrieve them from the cache.

The retrieved results are then reordered, highlighted and eventually completed in function of the bookmarks (cf. example figure 3). The links that exist both in the query results and in the bookmarks are raised in the top of the result order (ordered by the mark of the bookmark). The links present only in the bookmarks with a high mark follow in the result order. The other bookmarks are ignored. The other query results complete the new result ordered in function of the natural order given by the search engine.

The results with a high mark are highlighted by an index of relevance (between 1-3 stars at the beginning of the result line). For instance, Figure 3 presents the results of the search for "sistemas colaborativos 2009" (search done in April 2009 on www.google.com.br) and, it is possible to notice that the first result presented is the conference webpage for the Brazilian conference on "sistemas colaborativos", i.e., collaborative systems. The first links, consulted during previous searches, have been extracted from different result pages, brought to the head of the first result page and highlighted by the algorithm.

The first evaluation: An experiment

We conducted an initial evaluation of the concept in a controlled environment. The goal of this evaluation was to determine if the use of implicit social bookmarks could reduce the search time for different users doing successively a search on the same theme. This usage context targets different users working in the same team and frequently tackling the same search subjects, for instance, researchers from the same laboratory or students in a classroom.

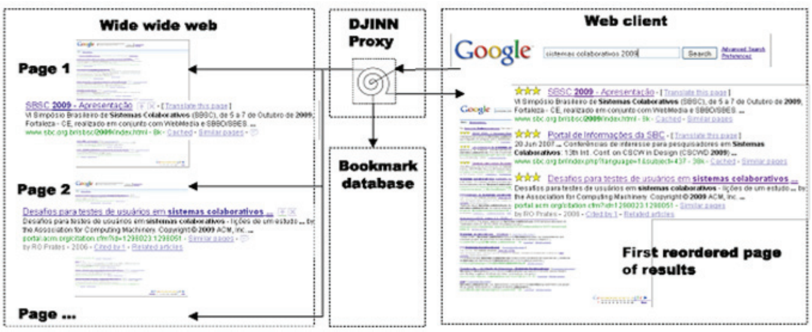

Figure 3. Integration of implicit social bookmarks in search result.

(a) The experiment

Five users from the same research team participated in the experiment. In the first session, they conducted 5 short searches (about 15 minutes each for the first user) on 5 different topics: 3 scientific themes and 2 historic themes. They had to retrieve specific information about each theme and provide a copy of three links reporting this information. The proxy was configured with a low level of inertia (see previous section).

Each user began by a different theme and moved successively to the other themes. Through the implicit social bookmarking, they benefited indirectly of the work previously done by the other users.

The same research team performed exactly the same exercise one week later in a second session. The bookmarks were cleanup between each session.

(b) Results and discussion

In this controlled context, the results were predictable. The users easily trusted the link implicitly recommended and highlighted by the proxy in the search results. The number of different visited pages and the time required to retrieve the information and select pages decreased rapidly based on the number of searches already performed on each theme (cf. Figure 4).

In addition, as expected, the time to perform the task decreased during the second session with the subjects. In the first page of results, these subjects retrieved the links that lead to the pages selected during the first session. Therefore, the results of this experiment motivated us to perform a second evaluation.

We also observed that unless a heterogeneous vocabulary is used by the different users to initiate the searches, the tags contained in the search contexts converged sufficiently to share the pertinent links between the users (see Figure 5).

(c) Threats to validity

Our first experiment, as any other experiments, has limitations. It does not evaluate whether the concept of *implicit social bookmarking* will still be efficient in the context of a high number of heterogeneous users. This means that the results of the first evaluation must be combined with the results of other experiments in different contexts to determine whether the concept of implicit social bookmarking can improve web searches in general.

In addition, even if users were not aware of the algorithmic principle increasing the rank of a page, because of the design of the

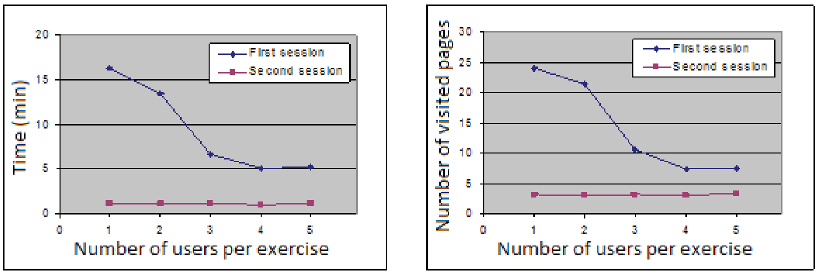

Figure 4. Average time and number of visited pages based on the number of users who had already performed the exercises.

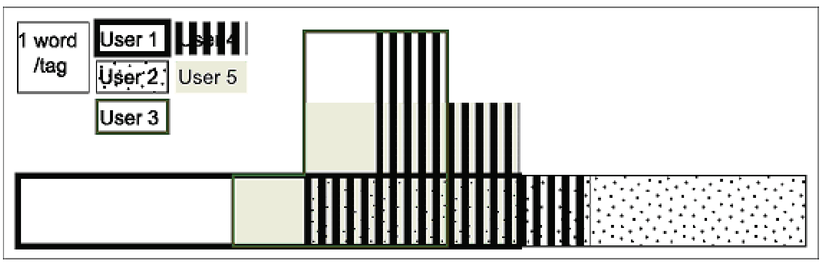

Figure 5. Intersection of the tags extracted from the users' interaction during the first exercise (first session).

task, they were still implicitly encouraged to store pages in their personal bookmarks and to revisit the links. Indeed, in order to select a set of relevant links containing three pages, the users pre-selected a larger set of links and revisited these links to establish a final selection. Our observation corroborates other usage reports (Tabard *et al.*, 2007) where the use of bookmark tools is less frequent in a non-controlled context. These actions turned the user's interaction very expressive increasing, maybe artificially, the mark of pages, and as a consequence our results.

The second evaluation: Open usage

Before designing a larger experiment aimed at evaluating DJINN in the natural context of web navigation, we performed a second evaluation to test the technical robustness of the system during a long period of use. This evaluation allowed us to assess the robustness of the marking algorithm. In order to do that, we shared the system between the research team members during a period of six weeks. Then, we analyzed the triplets tag/page/mark recorded in the bookmark database at the end of this period. We randomly selected a set o triplets and studied this set with the help of the users who used DJINN during the period. We compared the words used to tag the pages and the marks of the pages according to the users' point of view.

Overall, the words used as tags were reliable even if some generic and recurrent terms (such as "Web", "news" etc.) needed to be filtered. We observed that the system enabled relevant links to be identified, but this not happened in all cases. The number of links per tag was high and the classification between these links (the marks) was mostly relevant, but not in all cases.

In general, the results of this second evaluation were less positive as compared to the results observed during the controlled experiment. Two reasons justified this difference. First, the task proposed during the first experiment implicitly suggested the words used by the users to initiate their searches. As a consequence, the users started their searches based on the same request, as expected, since it was a controlled experiment. In addition, they visited several similar links and contributed in the same way to mark the pages. This reinforced the discrimination between the links and improved the classification. However, the task encouraged the users to visit and revisit the pages and to use bookmarks and history tools. As a consequence, the use of these tools increased significantly the page marks.

We noticed that, in a non-controlled context, users rarely accessed a page by the same way. This is consistent with previous results that suggest that bookmarks and history are rarely used (Tauscher and Greenberg, 1997; Cockburn *et al.*, 1999; Teevan, 2007). Consequently, during the six weeks of the second evaluation, the navigation was less expressive than during the controlled experiment and did not reach our expectations. In this regard, it should be noted that the evaluation of collaborative applications is more difficult than single-user applications, since it requires a minimal critical mass of people and also needs to take into account the possible interactions among the actors (Grudin, 1988).

Future work: Improving bookmarks and historic tools

In order to improve the reliability of the tags computed by DJINN in a non controlled context, we must be able to better understand the users' interest for contents. We highlighted two aspects of future research to help us to reach this goal. On the one hand, it is possible to improve the algorithm that computes the marks based on the users' navigation. For instance, we plan to evaluate the algorithm proposed by Bilenko and White (2008) in a short future. On the other hand, we identified in the first experiment that increasing the use of historic and bookmark tools may improve the navigation expressiveness. A possible strategy in this case is to explore the usage of enhanced re-visitation and re-find tools to both turn the navigation more expressive and help users to re-find contents.

Several projects have chased the goal of improving bookmarks and historic tools. In this case, the research mostly focused on the representation of pages history and bookmarks. WebMap (Doemel, 1994) and Browsing Icons (Mayer, 2001), for instance, improved the reuse of web pages history by representing each search as a graph of links pointing to visited web pages. Web Forager (Stuart *et al.*, 1996) suggested representing the bookmark as a "library" and storing the page in "books". PadPrints (Hightower *et al.*, 98) and WebView (Cockburn *et al.*, 1999) organized history as a navigation tree and associated each page with an image from the page to facilitate the remembering process of previous searches. These three representations are new ways to organize and structure search history improving the user's ability to interact with it. However, these different representations face the same problem: in order to remain efficient, they have to dedicate large space in disk to display complex re-visitation sequences. This space competes with the space dedicated to the actual displaying of the web pages. Consequently, using a traditional window-based organization, the two types of information can not be displayed at the same time hurting the tools' efficiency. In addition, our preliminary studies suggest that the identification through different images from pages is not so easy, especially if graphical charts of pages are similar (e.g., pages extracted from the same web site).

The second aspect of our research focused on data grouping. LeeTiernan *et al.* (2003) tool analyzes the pages relationship toward page semantic data (page texts and keywords) and proposes a graph connecting the links based on this relationship. PageLinker (Tabard *et al.*, 2007) groups connected links through a simple user interaction and suggests contextual bookmarks during the user navigation. If this solution does not target short-term review, it is a powerful tool to reduce the interaction cost of searches in bookmarks. Search Bar (Morris *et al.*, 2008) is another powerful tool to improve searches across the history of different application. Finally, Teevan (2007) when evaluating the Re:Search engine observed that the user frequently performs the same request into a search engine to retrieve a page previously visited (Teevan *et al.*, 2005). However, between the two searches, the results displayed by a search engine may change. To facilitate the revisitation of these pages by exploring the user's memory, Re:Search maintains the position of the visited pages into the search engine results.

The different projects cited above suggest three interesting directions for future work: (i) improve the representation of historical data to be able to retrieve information in the short and medium term; (ii) implicitly provide information back from the users' historical data to allow the retrieval of information in the medium and long term; and (iii) explore users' previous efforts performed during previous searches to be able to retrieve information in the medium term. Note that independently, these approaches do not offer a way to retrieve information at short, medium and long term at the same time. Consequently, it would be interesting to combine the three paradigms and to solve the problem of the wide space required to display historical data.

According to these observations, we proposed to develop a graphical interface, oriented by the following guidelines:

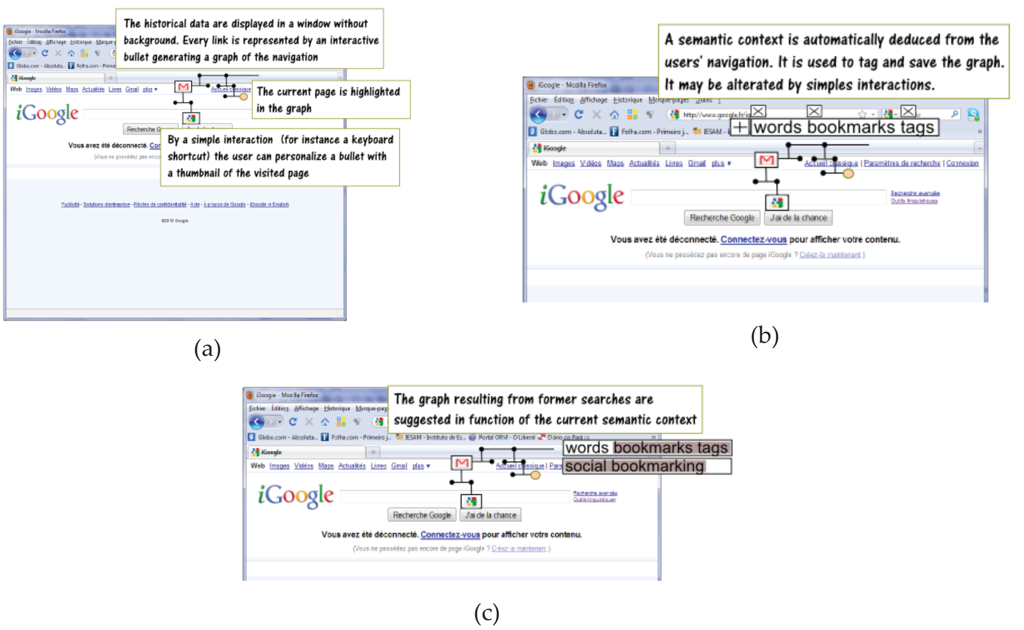

Figure 6. (a) Building a navigation historical tree, (b) associating a semantic context, (c) restoring the previous search contexts.

- The links must be directly accessible in the user environment and the design must balance the contradiction to have space for the link representation without reducing the web page surface. Consequently, we suggested to display the historical information in a windows without background (cf. Figure 6a);
- The cost of interaction to access, store and manage the links must be very low. This means that the links should be automatically tagged, saved (cf. Figure 6b) and then suggested in an appropriated context, namely when the user performs a search in a similar theme and may require them (cf. Figure 6c); and
- The design must help the user to have a better understanding of the re-visitation context. It must help the user to add personal marks into the represented links and remember the initial context of research during the re-visitation phases (cf. Figure 6c).

We are currently developing the graphical interface sketched below. Our future work will lead us to evaluate the impact of the interface as a re-visitation tool and measure the possible benefits toward the implicit social bookmarking.

Conclusion

The concept of *implicit social bookmarking* enables one to integrate *social bookmarking* information with the other user's navigation tasks to improve web searches. By interacting with his web browser, the user implicitly informs the *social bookmarking* system about his interest for specific web pages and the semantic context of this interest. In return, the user benefits from collective contributions that are displayed integrated into his search results. The neutrality of this additional information is preserved because: every system's user contributes to *social bookmarking*; every user has limited impact on a page mark (a mark between 0-10 for every user consulting the page); and the mark is not deduced from the published information, but instead from the interest of users for different contents.

This paper detailed our approach, the prototype implementation (DJINN) that we created, and two initial evaluations that we performed. Based on our results we argue that different users sharing similar interests could easily benefit from this collaborative work that is performed implicitly. However, the performance of the prototype could be improved by a better interpretation of the users' navigation. Consequently, we explored two research aspects to reach this goal, namely: additional algorithms to interpret the users' behavior during the navigation (Bilenko and White, 2008); and providing an HMI to help users to better manage their history and indirectly contributing to inform DJINN of their interest for contents. We plan to evaluate these two solutions in our future work.

Acknowledgments

The second author was supported by the Brazilian Government under grant CNPq 473220/2008-3 and by the *Fundação de Amparo à Pesquisa do Estado do Pará* (FAPESPA) through *Edital Universal N° 003/2008*. The second author conducted this work while at Federal University of Pará.

References

AMES, M.; NAAMAN, M. 2007. Why we tag: motivations for annotation in mobile and online media. *In:* ANNUAL SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS - CHI 2007, 25, San Jose, 2007. *Proceedings...* San Jose, p. 971-980.

http://dx.doi.org/10.1145/1240624.1240772

BILENKO, M.; WHITE, R.W. 2008. Mining the search trails of surfing crowds: identifying relevant websites from user activity. *In:* INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 17, Beijing, 2008. *Proceedings...* New York, ACM, p. 51-60.

http://dx.doi.org/10.1145/1367497.1367505

BLOGMARKS.NET. [s.d.]. Available at: http://blogmarks.net. Accessed on 01/10/2010.

CARD, S.K.; ROBERTSON, G.G.; YORK, W. 1996. The WebBook and the Web Forager: An information workspace for the World Wide Web. *In*: SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, Vancouver, 1996. *Proceedings...* Vancouver, ACM Press, p. 416-417. http://dx.doi.org/10.1145/257089.257407

CHANG LEE, K.; KWON, S. 2008. Online shopping recommendation mechanism and its influence on consumer decisions and behaviors: A causal map approach. *Expert Systems with Applications: An International Journal*, **35**(4):1567-1574.

http://dx.doi.org/10.1016/j.eswa.2007.08.109

COCKBURN, A.; GREENBERG, S.; MCKENZIE, B.; JASONSMITH, M.; KAASTEN, S. 1999. WebView: A graphical aid for revisiting web pages. *In:* OZCHI'99 - AUSTRALIAN CONFERENCE ON HUMAN COMPUTER INTERACTION, Wagga, 1999. *Proceedings...* Wagga.

DELICIOUS. [s.d.]. Available at: http://delicious.com/. Accessed on 01/10/2010.

DOEMEL, P. 1994. WebMap - A graphical hypertext navigation tool. *In:* INTERNATIONAL CONFERENCE ON THE WORLD WIDE WEB, 2, Chicago, 1994. *Proceedings...* Chicago, p. 785-789.

DOURISH P.; CHALMERS, M. 1994. Running out of space: Models of information navigation. *In:* HUMAN COMPUTER INTERACTION (HCI'94), 1994. *Proceedings...* Glasgow.

GOLDBERG, D.; NICHOLS, D.; OKI, B.M.; TERRY, D. 1992. Using collaborative filtering to weave an information tapestry. *Communications of the ACM,* **35**(12):61-70.
http://dx.doi.org/10.1145/138859.138867

GRUDIN, J. 1988. Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. *In:* ACM CONFERENCE ON COMPUTER-SUPPORTED COOPERATIVE WORK, 1, Portland, 1988. *Proceedings...* Portland, ACM Press, p. 65-84.

HAMMOND, T.; HANNAY, T.; LUND, B.; SCOTT, J. 2005. Social bookmarking tools (I) a general review. *D-Lib Magazine,* **11**(4).
http://dx.doi.org/10.1045/april2005-hammond

HEYMANN, P.; KOUTRIKA, G.; GARCIA-MOLINA, H. 2008. Can social bookmarking improve web search? *In:* ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING (WSDM'08), I, Stanford, 2008. *Proceedings...* Standford, p. 195-206. http://dx.doi.org/10.1145/1341531.1341558

HEYMANN, P.; KOUTRIKA, G.; GARCIA-MOLINA, H. 2007. Fighting spam on social web sites: A survey of approaches and future challenges. *IEEE Internet Computing,* **11**(6):36-45.
http://dx.doi.org/10.1109/MIC.2007.125

HIGHTOWER, R.R.; RING, L.T.; HELFMAN, J.I.; BEDERSON, B.B.; HOLLAN, J.D. 1998. Graphical multiscale web histories: a study of padprints. *In:* ACM CONFERENCE ON HYPERTEXt, 9, Pittsburgh, 1998. *Proceedings...* Pittsburgh, p. 58-65. http://dx.doi.org/10.1145/276627.276634

HOTHO, A.; JÄSCHKE, R.; SCHMITZ, C.; STUMME, G. 2006. BibSonomy: A social bookmark and publication sharing system. *In:* INTERNATIONAL CONFERENCE ON CONCEPTUAL STRUCTURES - CS-TIW '06, 14, Aalborg, 2006. *Proceedings...* Aalborg, Aalborg University Press, p. 87-102.

KOUTRIKA, G.; EFFENDI, F.A.; GYÖNGYI, Z.; HEYMANN, P.; GARCIA-MOLINA, H. 2007. Combating spam in tagging systems. *In:* INTERNATIONAL WORKSHOP ON ADVERSARIAL INFORMATION RETRIEVAL ON THE WEB - AIRWeb '07, 3, New York, 2007. *Proceedings...* New York, p. 57-64.

KRAUSE, B.; SCHMITZ, C.; HOTHO, A.; STUMME, G. 2008. The Anti-Social Tagger - Detecting Spam in Social Bookmarking Systems. *In:* INTERNATIONAL WORKSHOP ON ADVERSARIAL INFORMATION RETRIEVAL ON THE WEB - AIRWeb '08, 4, Beijing, 2008. *Proceedings...* Beijing, p. 61-68.
http://dx.doi.org/10.1145/1451983.1451998

LEETIERNAN, S.; FARNHAM, S.; CHENG, L. 2003. Two methods for auto-organizing personal web history. *In:* CHI'03 HUMAN FACTORS IN COMPUTING SYSTEMS, Fort Lauderdale, 2003. *Proceedings...* Fort Lauderdale, Fl, p. 812-813. http://dx.doi.org/10.1145/765891.766008

MALTZ, D.; EHRLICH, K. 1995. Pointing the way: active collaborative filtering. *In:* SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, Denver, 1995. *Proceedings...* p. 202-209.
http://dx.doi.org/10.1145/223904.223930

MATHES, A. 2005. Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Available at: http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html Accessed on: 01/10/2010.

MAYER, M.; BEDERSON, B. 2001. Browsing Icons: A Task-Based Approach for a Visual Web History. *In:* Technical Report CS-TR-4308, Computer Science Department, HCIL-2001-19, CS-TR-4308, UMIACS-TR-2001-85, HCI Lab. Maryland, University of Maryland.

MORRIS, D.; MORRIS, M.R.; VENOLIA, G. 2008. SearchBar: A search-centric web history for task resumption and information re-finding. *In:* ANNUAL SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 26, Florence, 2008. *Proceedings...* Florence, p. 1207-1216.
http://dx.doi.org/10.1145/1357054.1357242

PEREIRA, R.; SILVA, S.R.P. da. 2008. The use of cognitive authority for information retrieval in folksonomy-based systems. *In:* INTERNATIONAL CONFERENCE OF WEB ENGINEERING (ICWE08), 8, New York, 2008. *Proceedings...* New York, IEEE, p. 327-331.
http://dx.doi.org/10.1109/ICWE.2008.31

MOONEY, R.J.; ROY, J. 2000. Content-based book recommending using learning for text categorization. *In:* ACM CONFERENCE ON DIGITAL LIBRARIES, 5, San Antonio, 2000. *Proceedings...* San Antonio, p. 195-204.
http://dx.doi.org/10.1145/336597.336662

RIDDLE, P. 2005. Tags: What are they good for? Available at: http://prentissriddle.com/papers/riddle-2005-tags.pdf. Accessed on: 01/10/2010.

RIVADENEIRA, A.W.; GRUEN, D.M.; MULLER, M.J.; MILLEN, D.R. 2007. Getting our head in the clouds: Toward evaluation studies of tagclouds. *In:* ANNUAL SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, CHI 2007, 25, San Jose, 2007. *Proceedings...* San Jose, p. 995-998.
http://dx.doi.org/10.1145/1240624.1240775

RUFFO, G.; SCHIFANELLA, R. 2009. A peer-to-peer recommender system based on spontaneous affinities. *ACM Transactions on Internet Technology (TOIT)*, **9**(1):1-34.

http://dx.doi.org/10.1145/1462159.1462163

SINCLAIR, J.; CARDEW-HALL, M. 2008. The folksonomy tag cloud: When is it useful? *Journal of Information Science*, **34**(1):15-29.

http://dx.doi.org/10.1177/0165551506078083

STAFF, C.; BUGEJA, I. 2007. Automatic classification of web pages into bookmark categories, *In:* ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 30, Amsterdam, 2007. *Proceedings...* Amsterdam, p. 731-732.

http://dx.doi.org/10.1145/1277741.1277881

STURTZ, D.N. 2004. Communal Categorization: The Folksonomy. *INFO622: Content Representation*, **16**. Available at: http://davidsturtz.com/drexel/622/communal-categorization-the-folksonomy.html. Accessed on: 07/07/2011.

TABARD, A.; MACKAY, W.E.; ROUSSEL, N.; LETONDAL, C. 2007. Pagelinker: integrating contextual bookmarks within a browser. *In:* ANNUAL SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, CHI 2007, 25, San Jose, 2007. *Proceedings...* San Jose, p. 337-346.

http://dx.doi.org/10.1145/1240624.1240680

TAUSCHER, L.; GREENBERG, S. 1997. How people revisit web pages: Empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, **47**(1):97-137.

http://dx.doi.org/10.1006/ijhc.1997.0125

TEEVAN, J. 2007. The research engine: simultaneous support for finding and re-finding. *In:* ANNUAL ACM SYMPOSIUM ON USER INTERFACE SOFTWARE AND TECHNOLOGY, 20, Newport, 2007. *Proceedings...* New York, ACM, p. 23-32.

http://dx.doi.org/10.1145/1294211.1294217

TEEVAN, J.; ADAR, E.; JONES, R.; POTTS, M. 2005. History repeats itself: Repeat queries in Yahoo's query logs. *In:* ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, SIGIR '06, 29, Seattle, 2005. *Proceedings...* Seattle, p. 703-704.

http://dx.doi.org/10.1145/1148170.1148326

THOM-SANTELLI, J.; MULLER, M.J.; MILLEN, D.R. 2008. Social tagging roles: publishers, evangelists, leaders. *In:* ANNUAL SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 26, Florence, 2008. *Proceedings...* Florence, p. 1041-1044.

http://dx.doi.org/10.1145/1357054.1357215

WIKIPEDIA. [s.d.]. Folksonomy. Available at: http://en.wikipedia.org/wiki/Folksonomy. Accessed on July, 2011.

ZELDMAN, J. 2005. Tag clouds are the new mullets. Available at: http://www.zeldman.com/daily/0405d.shtml. Accessed on July, 2011.

VANDER WAL, T. 2007. Available at: http://vanderwal.net/folksonomy.html. Accessed on: July 2011.

US patent: US 6.285.999 B1 – Method for node ranking in a linked database.

Submitted on October 14, 2010.
Accepted on November 15, 2010.