Visualizing Search Results as Web Conversations

Shreeharsh Kelkar
Avaya Labs Research
233 Mt. Airy Rd
Basking Ridge, NJ
1-908-696-5110
skelkar@avaya.com

Ajita John
Avaya Labs Research
307 Middletown-Lincroft Rd
Lincroft NJ 07738
1-732-852-2057
ajita@avaya.com

Doree Duncan Seligmann
Avaya Labs Research
2 Penn Plaza
New York NY 10121
1-212-841-3036
doree@avaya.com

ABSTRACT
The user experience on the World Wide Web has radically changed in recent years into a participatory model. Users can now attach comments, tags, bookmarks, ratings to different kinds of web content such as news articles, blog-posts, online videos, etc. Thus conversations, both implicit and explicit, develop around web content providing valuable context for web search results. This paper advocates a conversational approach to visualizing web search results where each link resulting from a web search is considered the central artifact of a web conversation. Using websites like del.icio.us and Digg, this artifact is augmented with tags, comments, ratings along with the people who provided them. We have developed a system where search results from Google Search are visualized as a social network of web conversations. The visualization allows a user to see the rich inter-linking between tags, links and people, and the conversational relevance of each search result, thereby facilitating exploratory search. Currently, our system augments web search results from Google with meta-data (i.e. people, tags, notes) from del.icio.us.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]

General Terms
Design, Human Factors

Keywords
visualization, summarization, web search, conversations

1. INTRODUCTION
The World-Wide Web plays an important role in our lives today. It places a lot of information at our finger-tips – literally, since a click of the mouse brings a huge amount of information to our computer screens. Users are constantly searching the web to find information relevant to their tasks and needs.

There has been prior work on analyzing web searches. [5] identifies two *types* of search: *lookup* search and *exploratory* search. In lookup search, users are looking for discrete and well-structured objects of information e.g. someone's phone number or the URL of a certain biannual conference. In exploratory search, users search to *learn* about a certain topic. Instead of some fixed object of information, their objective is more diffuse and open-ended. [7] identifies two kinds of search *strategies*: teleporting and orienteering. Teleporting means that users prefer to get to their objective in just one step; orienteering implies that users try to reach their objective in a series of small steps wherein each step serves to "orient" them towards their objective. The basis of each step is the contextual knowledge that users have (e.g. that the file is probably in folder X, or that it was saved around Y months ago). The authors note that users often choose not to teleport, and even when they do, they are rarely successful in that they often end up orienteering rather than teleporting.

WWW 2009, April 20–24, 2009, Madrid, Spain.

Finally, there are search *tactics*, as distinct from *strategies* [1]. A search tactic is each step taken on the way to searching for something; a strategy is the overall path a search follows. Examples of tactics include typing in keywords and looking at the results and clicking on a folder to look at its contents. Search engines like Google afford one general tactic for searches: the user is asked to type in their search terms, upon which, Google returns an ordered linear list of URLs (links) that it believes are most relevant to the user. It follows from this that the ideal Google Search is one where the user is looking for a concrete piece of information (i.e. lookup search) and gets there in one step, by typing in a query and clicking on the first result (i.e. teleporting). However, like most interfaces, users do end up using it in different ways [7].

The aim of this paper is to propose some modifications to the standard tactic afforded by most web search engines today and create the following additional affordances for exploratory search and orienteering: (a) help in the "orienteering" process by bringing in contextual knowledge about search results from Web 2.0 portals like del.icio.us, (b) aid exploratory search by allowing serendipitous discoveries and dynamic categorization of search results [4] by visualizing search results as particles in 2-d space. In effect, our visualization makes new kinds of search tactics possible.

1.1 Our Contribution
Our modifications to the standard web search model can be divided into two components:

Using "conversation" data from the Web 2.0 to provide more context: Page Rank [2] was designed to show the relevance of search results based on how websites link to each other. Thus, the content authors determined the relevance of different web pages. But the Web has become more participatory. Publishing on the web has become easier thanks to blogging platforms like Blogger and Wordpress. Also, in addition to content authors, consumers can now leave their imprint on the web through tags, ratings, comments and bookmarks. We believe that taking these "web conversations" into consideration while searching the web would greatly facilitate exploratory search by providing more social context (tags, comments and users) for each link. In our work, we augment each link returned by Google Search with the tags, notes and people associated with that link on del.icio.us.

Visualizing query results as a network of web conversations: Most search engines present their results to the user in the form of linear lists. The linear list style of presentation has several drawbacks, the most important of which is that the list does not

show the inter-relationships between the various items. We are proposing an interactive visualization where web search results, embedded in their conversational context, are presented as particles in 2-d space. Users can introduce various conversational entities such as tags and people into the 2-d space, thereby revealing rich relationships between search results, tags and people.

The advantages of supplementing web search results with Web 2.0 data and visualizing them as conversations are:

(1) The links get supplemented with tags provided by different users. The tags may categorize a link, describe it, comment on it or suggest actions that lead us to more information. Common tags between links will allow us to see the relationships between them. The tags can also provide "summarization" of content e.g. by using tag clouds.

(2) Users can find people with common interests who may then be sources of more information. E.g. looking at other pages bookmarked by user X with tag Y may lead a user to more information on topic Y. We may also explicitly contact the person X, if such contact is possible.

(3) The comments (called "notes" in del.icio.us) help users see which links have more discussion around them, indicating their popularity or "interestingness" [3]. The details of the comments help a user get more context about the link.

(4) The particle visualization allows a user to see the rich inter-linking between tags, links and users. It allows people to see relevant web content, which would ordinarily be seen as a list of relevant links given a certain query, as a series of (often implicit) conversations between people who have tagged and commented on the links, thereby helping identify communities of interest.

In Section 2, we will explain our web conversation model in detail.

2. THE WEB CONVERSATION MODEL

A web conversation is defined as *an, often implicit, online interaction between actors on the web, who may be unaware of each other, and which results in the production of public, semi-public or private web artifacts.* Semi-public artifacts are web artifacts that can only be accessed by a specific group of people (e.g. emails, Facebook messages or links).

We can classify the *actions* (commenting, bookmarking, blogging, tagging, emailing, etc.) that actors perform as either implicit or explicit depending on whether the actors are aware that they are participating in a conversation. An example of an *explicit* action is commenting: here actors use the "comments" feature available on blogs and now on many websites. An example of an *implicit* action is bookmarking (e.g. on del.icio.us). Here the web actor is almost certainly doing this for her own benefit or perhaps that of her "network" (a del.icio.us feature). However since most contributed bookmarks, tags and notes remain public, they are often used by other users of del.icio.us but generally no explicit conversation occurs between users.

Web conversations start from a *central artifact*. Examples of a central artifact are a blog post, a news article, a YouTube video, an Amazon product, Flickr photo. Often, there is an *author* associated with an artifact and possibly a *time-stamp* that tells us when it was last modified.

Around the central artifact are the other web artifacts that are produced as a result of user actions – both implicit and explicit (see Figure 1). We call these *peripheral artifacts*. Some common peripheral artifacts are listed below: (a) *Bookmarks*: Bookmarks can be public (e.g. del.icio.us) or semi-public (e.g. Facebook). (b)

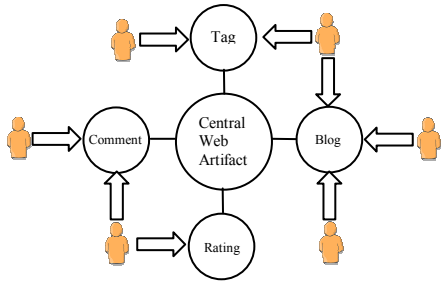

Figure 1: The web conversation model. The central artifact can be a blog-post or a news-article. Web actors (consumers), by performing actions such as bookmarking or commenting, create peripheral artifacts such as tags, comments, ratings, votes. Note that these peripheral artifacts can become central artifacts and spin off their own conversations.

Comments: Users post public comments on blogs and articles. (c) *Blog-posts*: Users can link to other articles by publishing on their own (public) blog. (d) *Emails*: Users can email interesting URLs to each other. Emails are semi-public artifacts i.e. they are only accessible to a few people to whom they are sent. (e) *Tags*: Tags are keywords associated with the central artifact. Tags can be created by the author (of the central artifact) or by other users who access it. Tags may be produced during the act of bookmarking, e.g. on del.icio.us. (f) *Ratings/Votes*: Users can rate or vote for the central web-artifact on websites like Digg or on the central artifact itself (e.g. a blog-post). Note that this list is not comprehensive. New ways of conversing may arise as the web evolves, giving rise to other kinds of peripheral artifacts.

A peripheral artifact has the following characteristics: (1) a *time-stamp* associated with it that indicates when it was created or last modified. (2) A user ID indicating who created it (i.e. its author). Note that a peripheral artifact may itself become a central artifact and spin off a different conversation around itself.

In our current implementation, the central artifact is the web search result returned by Google Search. The peripheral artifacts are the tags and notes created by del.icio.us users around that web search result. In Section 3, we describe our visualization to represent web conversations that occur around web search results as central artifacts.

3. VISUALIZING SEARCH RESULTS AS WEB CONVERSATIONS

Our key challenge is to design a visualization that presents to the user the web search results as web conversations in a way that facilitates both exploratory and lookup search. We have created a "particle" visualization of web search results that represents them (i.e. the central artifacts) as free floating particles in 2-d space. Users can introduce peripheral artifacts (in this case del.icio.us tags and the authors of these tags) into this space, upon which the web search results will cluster around the tags and users they are associated with. This allows a user to see the rich interlinking between web search results, users, tags and comments.

What follows is a description of our current implementation. The visualization was originally developed for Echoes, a system to store, tag, and visualize audio conversations in the enterprise [6].

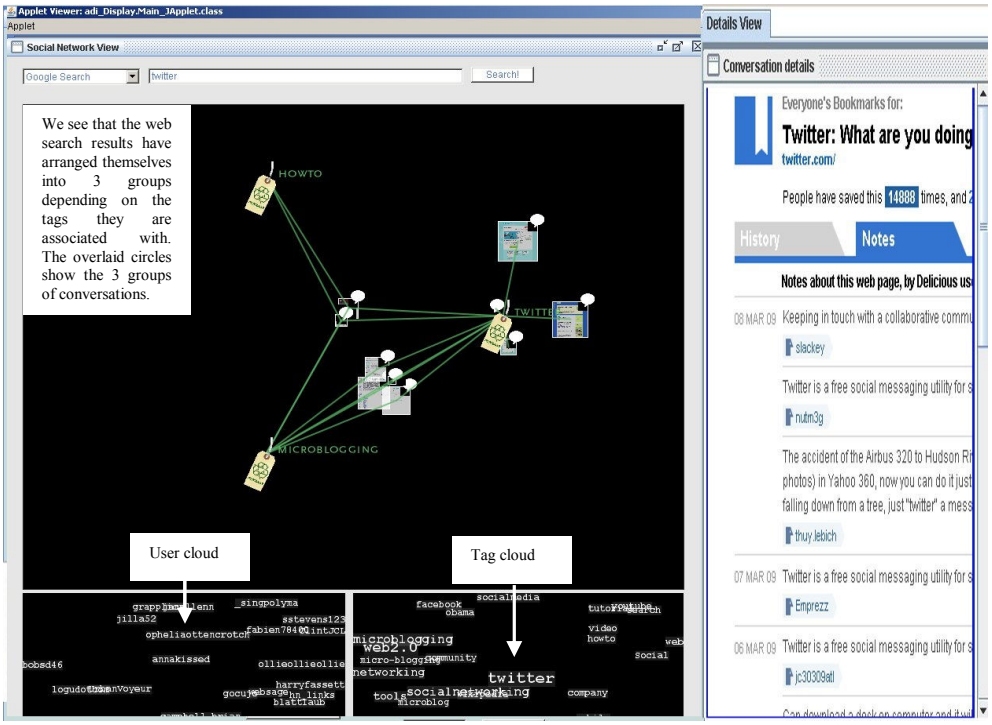

Figure 2: Our interface is divided into two views: the Social Network View (SNV) and the Details View (DV). In the SNV, the top 10 results of the web search "twitter" are shown as floating thumbnail images. Three tags -- twitter, microblogging and howto – have been introduced into the social network. We see that the queries sort themselves out into 3 groups: (1) direct links to Twitter itself, (2) links that are explanations of twitter on Wikipedia and TechCrunch and (3) videos that give a hands-on introduction to getting started with twitter.

The visual interface, shown in Figure 2, is divided into two views. On the left-hand side, occupying most of the screen space is the Social Network View. On the right-hand side is the Details View. The Social Network View contains a search box where the user can type in his query. The top links from Google Search (along with the relevant details about those links culled from del.icio.us) are returned in response to the query.

These web search results are represented as "particles" freely moving within the space. These particles are the *central artifacts* of web conversations. Each particle has two components: (a) the thumbnail of the web page the search result points to: The size of the thumbnail is proportional to the Google relevance of the search result; thus the top link returned by Google Search has the largest thumbnail, the second link is smaller and so on. (b) A "callout" icon whose size is proportional to the "conversational relevance" of the search result. In our current implementation, the conversational relevance is the number of users who have bookmarked that search result on del.icio.us. The conversational relevance can be enhanced based on other factors such as number of comments, number of tags, bursts and periodicity of web conversation activity etc.

When a user moves his mouse over the particle, the size of the thumbnail becomes larger, so the user gets an idea of the content of the webpage. Using CTRL+click will open the URL directly in a web-browser.

Along with the particles, the Social Network View displays users and tags at the bottom as clouds: these are the users and tags associated with the search results. The size of a tag or user is proportional to *its* conversational relevance, in this case, the number of web search results (or central artifacts) it is associated with. Thus, tags and users that are associated with more web search results appear prominent. (In this case too, the conversational relevance of users and tags can be enhanced based on other factors.)

When one of these users or tags is selected, a "query particle" is introduced into the search space. Search results that are associated with this query get attracted to it. A search result is associated with a user if she has bookmarked it on del.icio.us. A search result is associated with a tag if a user has bookmarked the web-page with that tag. Lines are drawn between inserted queries and the associated web search result. When multiple queries are introduced, the web search results re-arrange themselves so that they are closest to the queries they are related to. This aspect of our visualization can be thought of as a rich "categorization" of search results. In Figure 2, we see that 3 tags have been introduced into the Social Network View. The search results have sorted themselves out into 3 groups, based on their relationships with one or more of these tags.

The Details View refreshes when a particle is clicked, displaying the "details" about the search result. The details – which include the tags applied to the URL, the users who are associated with that tag and the notes that they used – are taken directly from the del.icio.us page about that link. Clicking on the user or tag opens the del.icio.us page of that tag or user in the Details View.

In subsequent implementations, we plan to introduce additional features to our visualization. In order for our visualization to scale (i.e. when the number of peripheral artifacts is high), we plan to incorporate techniques for extracting the most "interesting" [3] peripheral artifacts. We also plan to explore the temporal flow of conversation activity, look for patterns like periodicity and burstiness, and render them visually to offer more insights to the user.

We illustrate the current implementation of our visualization with an example. Let us consider a user X who is interested in querying the web to learn more about "Twitter". Now user X may

have different types of information needs: he may know what Twitter is and just want to get to the link so that he can start using it. Or he may be a new user who has just got on to Twitter and is interested in exploring its possibilities. Or he may just be someone who has heard of Twitter and is interested in finding out what it is exactly and how it works.

X types in "twitter" in the search box. The top links obtained from Google Search are displayed as free floating particles. The top link returned by Google is twitter.com. In our visualization too, this link is shown prominently and users can CTRL+click on the thumbnail to go straight to the Twitter page. Thus, our visualization facilitates lookup search.

Clouds of associated users and tags are shown at the bottom of the Social Network View in our visualization. Some of the tags that get rendered prominently include "Web 2.0", "microblogging", "networking" and "community" which may help user X learn more about Twitter.

Suppose that user X wants to sort through the search results. He clicks on the tags "twitter", "microblogging" and "howto" which introduces them into the Social Network View. In Figure 2, we see the result of these actions. The web search results sort themselves into 3 groups. The first group is a set of links that gets attracted only to the tag "twitter": we see that these are all official Twitter links, they take the user to the Twitter login page, the Twitter registration page etc. The second group of links is attracted to both "twitter" and "microblogging": these are links *about* Twitter e.g. the Wikipedia page on Twitter, the CrunchBase page on Twitter, etc. If user X is interested in knowing what Twitter is, he will find them useful. Finally, a third group of links that is attracted to "howto" "twitter" and "microblogging" points the user to videos that give a hands-on introduction to using Twitter. We see that the web search results have been richly categorized and can serve a variety of information needs. Moreover the categorization is dynamic and will evolve with time as more peripheral artifacts get associated to the search result.

Now suppose X clicks on the largest thumbnail he sees the tags that have been applied to the page as well as the "notes" written about that page by users in the Details View. One of those notes by del.icio.us user Lacey514 reads: "*Twitter is a great tool. It is fast and easy to implement into the classroom. Students must think carefully to communicate effectively in 140 characters. Accounts can be setup so that participants are invited into one's network which helps reduce the chance of inappropriate content. Some of the weaknesses of Twitter are teachers are unable to manage inappropriate content and sometimes the 140 word limit may limit self-expression of students. Students can build personal learning networks and teachers can network based on similar interests. There are great resources for unlimited resources globally. Some ideas for the classroom are students can create an ongoing story built upon by each new participant, teacher's can pose questions for students to solve and students can have a discussion on it, This allows the teacher to see who is participating and responding to the class discussion.*" Clearly Lacey514 has bookmarked this link as part of an overall plan to look into the uses of twitter as a tool for teaching and tagged it with "teacherresources" (among other tags). By clicking on the tag, user X can see other bookmarks by Lacey514 tagged with "teacherresources". He may bookmark Lacey514 as someone to keep track of.

To summarize, our particle visualization of web search results as web conversations helps both lookup and exploratory search in the following ways: (1) By rendering our "particles" in sizes that are proportional to the Google relevance of the search result, we ensure that a user can skip directly to the top web search result. By also visually showing the "conversational relevance" of a web search result, we provide the user with another facet into the relevance of web search results. (2) A user can interact with the search results by introducing queries (i.e. any combination of users and tags) into the Social Network View. Because the web search results will move closer to the tags and users they are associated with, they get richly categorized into groups. This has two advantages over traditional "hard" categorization schemes [4] to group web search results: a web search result can belong to more than 1 category and the categorization is malleable, depending only on the peripheral artifacts that gather around the web search result. Rich categorization facilitates better exploratory search. (3) By incorporating del.icio.us data into the web search results, a user gets access to a rich community of users thereby increasing the chances of finding new and interesting things serendipitously.

4. FUTURE WORK

Our work is still in its initial stages. We plan on adding features like bookmarking to our interface. We also plan to conduct user studies to compare our web conversation-based visualization to the standard linear list presentation of search results, and then perhaps go beyond del.icio.us and integrate other portals of web conversations e.g. Digg and My Yahoo.

5. REFERENCES

[1] Bates, M. (1979). *Information Search Tactics.* JASIS **30**: 205-214.

[2] Brin, S. and Page, L. (1998). *The anatomy of a large-scale hypertextual {Web} search engine.* Computer Networks and ISDN Systems **30**: 107--117.

[3] Choudhury, M. D., Sundaram, H., John, A. and Seligmann, D. D. (2009). *What Makes Conversations Interesting? Themes, Participants and Consequences of Conversations in Online Social Media.*, International World Wide Web Conference, Madrid, Spain, April 20-24.

[4] Hearst, M. (2006). *Clustering versus faceted categories for information exploration.* Commun. ACM **49**(4): 59-61.

[5] Marchionini, G. (2006). *Exploratory Search: From finding to understanding.* Communications of the ACM **49**(4): 41-46.

[6] Renduchintala, A., Kelkar, S., John, A. and Seligmann, D. D. (2007). *Designing for Persistent Audio Conversations in the Enterprise*, DUX '07, Chicago, IL, USA, November 5–7.

[7] Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. R. (2004). *The Perfect Search Engine is Not Enough*, CHI, Vienna, Austria, April 24-29.