

An Empirical Evaluation of User Interfaces for Topic Management of Web Sites

Brian Amento
AT&T Labs - Research
180 Park Avenue, P.O. Box 971
Florham Park, NJ 07932 USA
brian@research.att.com

ABSTRACT

Topic management is the task of gathering, evaluating, organizing, and sharing a set of web sites for a specific topic. Current web tools do not provide adequate support for this task. We created and continue to develop the *TopicShop* system to address this need. TopicShop includes (1) a webcrawler that discovers relevant web sites and builds site profiles, and (2) user interfaces for exploring and organizing sites. We conducted an empirical pilot study comparing user performance with TopicShop vs. Yahoo™. TopicShop subjects found over 80% more high-quality sites (where quality was determined by independent expert judgements) while browsing only 81% as many sites and completing their task in 89% of the time. The site profile data that TopicShop provide – in particular, the number of pages on a site and the number of other sites that link to it – were the key to these results, as users exploited them to identify the most promising sites quickly and easily.

INTRODUCTION

Web search and navigation are difficult problems that have received much attention, with search engines like AltaVista and directories like Yahoo being the most widespread solution attempts. However, users have information needs and interests that are larger in scope and longer in duration than can be satisfied by AltaVista and Yahoo. In particular users want to manage their persistent interests in broad topics and to comprehend collections of web documents relating to topics.

MOTIVATION

Typical search solutions are content-based where a user query is filled by matching keywords to the text of web pages. While this approach works in many situations, it fails when users want to find quality information on a topic and manage the resulting information over a period of time. By utilizing the inherent structure found on the World Wide Web, we may gain more insight into the perceived quality of a web site. By viewing links to web pages as endorsements, we can use the concepts of social filtering (utilizing user preference for prediction) to create better collections of topically coherent web sites. Social filtering is a method of filtering objects (documents, videos, web pages, etc.) that concentrates on the characteristics of people and their preferences in addition to the objects' content. The focus of social filtering is shifted from strictly assessing the content of objects to evaluating the personal and organizational relationships of the community of users accessing those objects.

An important task that many web users perform is gathering, evaluating, and organizing relevant information resources for a given topic; we call this *topic management*. Sometimes users investigate topics of professional interest, at other times topics of personal interest. Users may create collections of web information resources for their own use or for sharing with coworkers or friends. For example, one might gather a collection of web sites on wireless telephony as part of a report for work or a collection on the X-Files as a service for fellow fans. Librarians might prepare topical collections for their clients, and teachers for their students.

Topic management is a difficult task that is not supported well by current web tools. A common way to find an initial set of (potentially) relevant resources is to use a search engine like AltaVista or an index like Yahoo. At this point, however, a user's work has just begun: the initial set usually is quite large, consisting of dozens to hundreds of sites of varying quality and relevance, covering assorted aspects of the topic. Users typically want to select a manageable number – say 10 to 20 – of high-quality sites that cover the topic. With existing tools, users simply have to browse and view resources one after another until they are satisfied they have a good set, or, more likely, they get tired and give up. Browsing a web site is an expensive operation, both in time and cognitive effort. And bookmarks, probably the most common form of keeping track of web sites, are a fairly primitive organizational technique.

While many web search utilities provide answers to specific queries, they do not provide convenient, efficient methods for exploring the body of knowledge available about a topic. Some search resources allow users to find a category that closely matches the topic they are interested in, but the end result is simply an alphabetical list of web sites that contain information on the given topic. New techniques that provide additional functionality need to be available on the web to support broader types of information gathering.

WEB CRAWLING

There are many different sites on the web for any given topic. An alphabetized list of all known sites is rarely the best method for finding useful information. The inherent hierarchical structure of the web can be used to gain further information about web sites. By following all hypertext links on a web site, a topic crawl can be generated for all sites linked to by a particular site. Continuing the crawl deeper past these sites will eventually provide a large body of topically related sites that can be analyzed and presented to a user. This is based on the assumption that quality sites point to other relevant quality sites. Since site designers have theoretically already put effort into filtering out poor quality sites and only linking to quality sites, a crawl can simply follow links to build a better representation of the scope of sites for a given topic.

Our crawl uses a clan graph as the primary information structure. A clan graph is a directed graph where nodes represent documents and edges represent a reference to the node pointed to. A local clan graph is the subgraph whose nodes are closely connected to the user-specified set of seed sites. Building on concepts from social network analysis, co-citation analysis, and social filtering we have developed the notion of an NK local clan graph.

- The NK local clan graph for a seed set S is $\{(v,e) \mid v \text{ is in an } N\text{-clan with at least } K \text{ members of } S\}$. An N-clan is a graph where every node is connected to every other node by a path of length N or less, and all of the connecting paths go through only nodes in the clan. Our crawler uses a 2-clan (the 2K local clan graph) because it represents a useful substructure extracted from the large structure of the web. By requiring that sites relate to a certain number of seeds (K), we ensure that we find not just dense graphs, but graphs in which a certain number of the seeds participate.

There are three types of inter-document relationships where a relationship between two of the documents can be inferred based on a known relationship between the other two. Co-citation analysis says that two documents B and C are related if a third document, A, sites them both. Social filtering says that if documents B and C both refer to a third document, A, then B and C may be likely to link to similar sorts of items in general. Transitivity says that if document A refers to B and B to C, then A implicitly refers to C. These three relationships are the minimal 2-clans which are in our case necessary because no smaller structure allows us to make inferences about document relatedness, and sufficient because no larger structure enables other simple inferences

During a crawl, a number of parameters describing sites are gathered. Number of images, audio files, and movie files are recorded as well as the number of in-links and out-links. The number of links pointing to a site by other outside sites is called the in-links. This parameter can be used to determine if the site is a popular site by finding the number of site designers that think it is good enough to be linked to. This is a form of social filtering. By considering each in-link to a site to be an endorsement to that site we can generate a list of the most linked-to or most endorsed sites. An out-link is where a site links to another site. The site with the most out-links can be considered a good index site with many links about the desired topic. Combining these two parameters can provide further information. If a site is pointed to by many sites, but does not point to any other sites, it may be an official site (perhaps a corporate site) on the topic since many sites think its important, but the site itself does not point to any other sites. If, on the other hand, a site is not pointed to by many other sites but itself points to a large number of other resources, it may be a newer site that other site designers have not noticed yet. Most likely, it is a link collection site if it has a high number of out-links.

While a crawl is being performed, two metrics are used to ensure that highly relevant sites are visited in the early stages. First, a weighted sum of the number of in-links of all sites that point to a page is used to rank the page on its potential for not only being a quality site but for recommending other quality sites. As a crawl progresses, this ranking is improved because more data about visited sites are collected. If a site is pointed to by many other sites with a high number of in-links (and hence are considered good sites because they are endorsed by others), then this site can also be considered a good site. Because of the immense size of the web, a crawl can take a very long time but by using this metric, more relevant sites are found by the crawler near the beginning of a crawl and a crawl can be stopped after some user-defined threshold number of sites is found. In addition, anchor text is searched for keywords related to the crawl.

Anchor text is the text description, written by the site designer, that is displayed for each link and is what the user clicks on to visit the site linked to. This text is usually highly related to what the site contains. So during a crawl, all occurrences of anchor text are saved for each site and can be searched to gain relevance feedback. If a match is found, then the ranking for the site is improved; if no match is found, nothing is done, because that does not necessarily mean a site is off-topic.

TOPICSHOP

TopicShop Explorer is a visualization for viewing and managing collections. It is a customized version of the normal Windows file Explorer. The TopicShop Explorer is a very small Windows executable that knows how to read and process site profile files.

Users can view their collections in two different ways: details or icons. The main feature of the details view is that it shows site profile information, and the main feature of the icons view is that users can arrange icons spatially. We had three main design goals for TopicShop Explorer:

Make relevant but invisible information visible. We hypothesize that making site profile information visible will significantly inform users in evaluating a collection of sites. No longer must they decide to visit sites — a time-consuming process — based solely on titles and (sometimes) brief textual annotations. (A chief complaint of subjects in the Abrams et al, study was that titles were inadequate descriptors of site content — and that was for sites that users already had browsed and decided to bookmark.) Instead, users can choose to visit only sites that have been endorsed (linked to) by many other sites or sites that are rich in a particular type of content (e.g., images or audio files). In addition to site profile data, the thumbnail images also are quite useful; most notably, for sites a user has visited, thumbnail images are an effective visual identifier for sites.

Make it simple for users to explore and organize resources. In the details view, users can sort resources by any of the properties (e.g., columns showing number of in-links, out-links, images, etc.) simply by clicking on the label at the top of the column. In either view, right-clicking on a site brings up a window that shows profile data from which the numbers in the columns are derived (e.g., lists of all sites that link to the selected site and all internal pages of the site). Double-clicking on a site will send the user's default web browser to that site.

Users can organize resources both spatially (in the icons view) and by creating subfolders and moving resources into the subfolders. Nardi & Barreau found that users of graphical file systems preferred spatial location as a technique for organizing their files. We believe spatial organization is particularly useful early in the exploration process while users are still discovering important distinctions among resources and user-defined categories have not yet explicitly emerged. As categories do become explicit, users can create folders to contain sites in each of the categories.

Integrate topic management into a user's normal computing and communications environment. The TopicShop Explorer may not look like a novel interface at all; interestingly enough, this was an explicit goal. We wanted it to be as similar to the normal Windows Explorer as possible so Windows users could apply all their existing knowledge, meaning there would be little or no learning time and similar ease of use. Further, this decision makes it very easy for collections of resources to be shared. Since a collection is just a normal Windows folder containing files (of the special type that we designed), they can be shared in all the normal ways. As we already have explained, a collection can be compressed and downloaded. It can also be emailed. And if users share a common network, collections simply can be read directly from any machine on the network.

The TopicShop Explorer interface allows users to organize their web site collection from any view. In the details view, users can change the order of the collection of web sites to represent their personal choice of best quality sites. This ordering becomes an additional column in the interface that can be sorted like any other column. In the icons view, spatial organization is allowed and web site icons can be arranged into groups before being moved to a new folder.

EMPIRICAL STUDY

We wanted a suitable baseline topic management tool for comparison to TopicShop. Yahoo is the most popular tool for locating collections of web sites (according to the Media Metrix PC metering company, results available at <http://searchwnginewatch.com/reports/mediamatrix.html>). Bookmark lists are a common means of organizing collections of web resources. Therefore, we decided that subjects would use either TopicShop or Yahoo/bookmarks.

We chose two topics for the study: home brewing (of beer) and the TV program "Buffy the Vampire Slayer" — each contained about 60 sites in their corresponding Yahoo category. Our choice of

these topics was influenced by the fact that pursuing special interests, including hobbies and media fandom, is one of the main ways people use the web. To quantify this, we studied a set of approximately 770K queries issued to the Magellan search engine between March 1997 and August 1998. We determined that 42% of the queries had to do with entertainment topics, including media fandom (categorization was done on the 515 most popular query strings, which cumulatively accounted for 96K query instances; categorization was done by two independent raters, inter-rater reliability of 87%.)

DESIGN

The experiment was a 2x2, between subjects design, with topic (home brewing or Buffy) and user interface (TopicShop or Yahoo) as factors. Sixteen members of our lab volunteered to participate, giving four subjects per each of four conditions. None of the subjects had seen TopicShop before, although some were familiar with the general concepts.

The two main metrics we wanted to measure were the quality of resources users gathered and the amount of effort (time and total number of sites browsed) required. To give a quality baseline, four experts for each topic were presented a list of the sites (in random order) on that topic; only titles were presented, no Yahoo annotations or TopicShop profile data. This meant that the experts had to browse each site and evaluate it based on its content and layout. Each expert collected the 20 “best” sites. For this study, we defined “best” as a set of sites that collectively provided a useful and comprehensive overview for someone wanting to learn about the topic. During analysis, we used the “expert intersection”, the set of resources that all experts for a given topic selected, as the yardstick for measuring the quality of resources selected by the subjects.

Subjects for a given topic, whether they used TopicShop or Yahoo, were presented with the same set of approximately 60 sites (obtained from the Yahoo category) to evaluate. Yahoo subjects saw (as usual) site titles and, for about half the sites, a brief textual annotation. For the TopicShop condition, we applied our webcrawler to the Yahoo sites to produce site profiles; TopicShop subjects thus had access to site tiles, thumbnail images, and profile data.

METHODOLOGY

Subjects were assigned randomly to one of the four conditions. To begin the experiment, subjects received 15 minutes of instruction and training in the task and user interface. TopicShop subjects were shown the basic interface features and taught how to collect sites by dragging and dropping icons into folders. Yahoo subjects were shown a sample list of sites and taught how to collect sites by bookmarking. After training, subjects performed a short task to ensure that they were comfortable with collecting and organizing sites.

For the main task, subjects investigated the sites for their assigned topic by using the interface (TopicShop or Yahoo) and browsing to sites. Subjects within a single topic were presented with the same collection of sites in both interface conditions. They were asked to choose the 15 “best” (as defined previously) sites and rank them by quality. Subjects were asked to complete the task in 45 minutes and were kept informed of the elapsed time. Clearly, there is a relationship between time on task and quality of results: the more time spent, the better results one can expect. By limiting the amount of time, we hoped to focus on any differences in the quality of results (i.e., the sites users selected) between the two interfaces. And people don’t spend unlimited amounts of time browsing, so we wanted to see whether users could find high-quality sites in a limited amount of time.

The task ended when subjects were satisfied with their collections of sites. Subjects then completed a short questionnaire. Finally, an informal interview was conducted to reveal strategies subjects used to perform the task, their reactions to the interface, and what could help them to complete the task more effectively.

RESULTS

We first compared the set of resources chosen by each subject to the expert intersection. For each topic, the expert intersection contained 12 resources. For the Buffy topic, Yahoo subjects selected an average of 5.0 sites that were in the expert intersection, while TopicShop subjects selected 7.5 expert-endorsed sites. For home brewing, Yahoo subjects matched 4.3 sites and TopicShop subjects matched 9.3. Overall, Yahoo subjects selected 4.6 sites from the expert intersection, while TopicShop subjects selected over 80% more, or 8.4 ($p < 0.05$). These results are summarized in Table 1. Notice that choosing sites at random would result in obtaining 3 sites in the expert intersection. (Users selected 15 out of 60 sites, or 25%; 25% of the 12 sites in the expert intersection is 3 sites.) The Yahoo score of 4.6 is not that much

better than random selection. This probably is due to task time limit of 45 minutes. If Yahoo subjects had had unlimited time, undoubtedly they would have been able to find more high quality sites. To sum up, we see that TopicShop users found significantly better resources in the time given to complete the task.

Mean Number of High-Quality sites Identified

<i>Topic</i>	<i>Interface Type</i>	
	Yahoo	TopicShop
Buffy	5.0	7.5
Home brewing	4.3	9.3
Average over Topic	4.6	8.4

Expert intersection analysis

It also is revealing to examine the amount of work subjects performed to complete their tasks. A study of data from the search engine Excite (51,473 queries, 18,113 users) showed that 86% of all users looked at no more than 30 pages returned in response to their query

In our study, Yahoo users browsed an average of 44 sites, while TopicShop subjects visited about 36, or about 19% less. Further, the task of constructing a high-quality collection of resources is more difficult than doing a simple search; the task is global, since one is trying to develop a comprehensive overview of a topic, so more sites must be considered. By providing additional dynamic data up front, TopicShop enables users to make better decisions about which sites to immediately rule out and which to investigate further. Yahoo users can rely only on textual annotations, which are provided by site maintainers. While these annotations are sometimes helpful, they can be out-of-date or self-promotional, so are not necessarily good indications of the perceived quality of a site.

We also analyzed time on task. We did not expect a large difference since we gave users a (soft) limit of 45 minutes to complete the task and kept them aware of elapsed time during the experiment. Still, TopicShop subjects took about 11% less time (41.5 minutes vs. 46.6 minutes for Yahoo).

While the differences in time and effort were not statistically significant (although we hope they will be in a larger study we are preparing to conduct), they do show that TopicShop subjects did not obtain better quality results at the cost of more work.

The questionnaire gave us data on what information subjects found most useful in evaluating a site. TopicShop subjects were asked to rank the utility of the site profile attributes, including the title and the number of in-links, out-links, images, audio files, and pages on the site. Subjects ranked these properties on a scale of 1 (most useful) to 7 (least useful). Three of these properties — in-links (2.00), title (2.75), and number of pages (3.00) — were ranked most highly. The other four properties had an average score greater than 5. Even though many subjects noted that title is not a very good indication of quality, it still was perceived as one of the most useful site properties. In interviews, subjects explained that titles were useful mainly as memory aids for sites. Thus, subjects considered the number of endorsements (in links) and the size of a site (in pages) to be the most useful indicators of quality.

The questionnaire also asked subjects what additional information would have helped them in evaluating sites. Six of the eight Yahoo subjects said that the number of links between sites would be very useful. One subject even made it a point to go to the links page of every site visited to see not only what sites were linked to, but also to read any annotations or recommendations made by the site author. Thus, link information was rated as highly useful by those subjects who saw it and as desirable by those subjects who did not.

SUMMARY

As the amount of information on the web continues to grow, tools that support users in finding and managing collections of topical resources will become increasingly significant. The focus must move from compiling collections to helping users comprehend and manage them. Our goal is to reduce the time users must spend sifting through “relevant” – but poor quality – sites and increase the amount of time they can devote to exploring high-quality information. By mining the rich data that already exist in the structure of web sites and content of their pages, we expect to show that TopicShop helps users quickly identify small, manageable, high-quality subsets of web sites.