Friends, Romans, Countrymen: Lend me your URLs. Using Social Chatter to Personalize Web Search

Abhinay Nagpal **Sudheendra Hangal** **Rifat Reza Joyee** **Monica S. Lam**

Computer Science Department
Stanford University
{abhinay, hangal, joyee, lam}@cs.stanford.edu

ABSTRACT
People often find useful content on the web via social media. However, it is difficult to manually aggregate the information and recommendations embedded in a torrent of social feeds like email and Twitter. At the same time, the ever-growing size of the web and attempts to spam commercial search engines make it a challenge for users to get search results relevant to their unique background and interests. To address these problems, we propose to let users mine their own social chatter, and thereby extract people, pages and sites of potential interest. This information can be used to effectively personalize web search results. Additionally, our approach leverages social curation, eliminates web spam and improves user privacy.

We have built a system called SLANT to automatically mine a user's email and Twitter feeds and populate four personalized search indices that are used to augment regular web search. We evaluated these indices and found that the small slice of the web indexed using social chatter can produce results that are equally or better liked by users compared to personalized search by a commercial search engine. We find that user satisfaction with search results can be improved by combining the best results from multiple indices.

Author Keywords
Web search, human curation, social networks, Twitter, email

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: Miscellaneous

INTRODUCTION
Web search is one of the most commonly used applications on the Internet. While it is a marvel that search engines like Google and Bing can quickly look up and rank hundreds of billions of pages in response to a query, we hypothesize that users can often get their answers from a relatively narrow and uniquely personal slice of the web. In this paper, we explore ways to automatically generate this slice, and thereby slant web search results towards sites of the user's interest. We show that a surprisingly small slice goes a long way towards satisfying the user's information needs.

In the early days of the Internet, users often created web pages with pointers to other pages that they liked. The original Yahoo directory is a prominent example. Algorithms like Pagerank were invented to take advantage of such curation and use the link structure of the web to rank search results [5]. However, the commercial importance of search engines has led to artificial manipulation of result rankings. A well-publicized example was J. C. Penney's success in influencing Google search to place its sites at the top for queries about a wide range of products [25].

In this context, there appears to be value in bringing back some elements of human curation to the search problem. Social recommendations are of particular value because friends often share common affiliation and interest. This fact is recognized by today's search engines, which are increasingly attempting to factor social information into search results.

Social Chatter as a Source of Curated Content
Our work is based on the observation that personalized and implicitly curated social recommendations of web content already exist in the user's personal communications (such as email) and social feeds (such as Twitter). A key idea in this paper is to personalize search by *indexing only those domains mentioned in a user's online chatter.* Many Internet users already use social media for exchanging pointers to web pages that are of interest and relevance to them. For example, Twitter is extensively used for sharing links; as of 2010, 25% of the approximately 90 million tweets per day contained links [24]. Similarly, users frequently recommend links to each other over email. We aim to let the user mine this ambient social chatter to infer sites that are likely to be relevant to her, and to use them while performing web search. Our approach of tapping into existing social chatter has the advantage of being useful for millions of online users who participate in social media (including email) and thereby allowing realistic studies.

Note that our approach is different from just searching the textual content in Twitter, email or the pages referenced

CSCW 2012, February 11–15, 2012, Seattle, Washington, USA.

therein because we use the *domains* referenced as the basis for creating a customized search index.

Four Types of Personalized Indices

In this paper, we describe a system called SLANT that incorporates four types of search indices that can *augment* (not replace) the results of a regular web search, which itself may be personalized by the search engine. The first three search indices in SLANT were designed based on links extracted from three sources of information: a user's personal email, his Twitter feed, and the topmost tweets on Twitter globally. The fourth index is based on pages that contain the names of the user's friends. We layer these indices on top of a regular search engine like Google, in order to enable rapid experimentation.

We chose these indices for study as each one has different characteristics. Emails are personal, private and generally related directly to the user's online activities. On Twitter, messages are usually public and users tend to follow others they trust, whether or not they know them personally. The topmost tweets across all of Twitter capture socially curated, but non-personalized URLs. Results with friends names point to expertise or opinions of the user's friends on the query topic. As we describe later, a slightly intangible benefit of SLANT is that by adding color from their social network, search results can be made more entertaining than the uniform and somewhat monotonous results of a traditional search engine.

Implicit consumption of information

As suggested by the Shakespearean quote in this paper's title, there are multiple levels of social chatter around a user. The chatter may involve the user's inner social circles, outer social circles, friends of friends, people the user trusts but does not know personally, or members of a wider community. Each of these sources provides information valuable in different settings, though it is difficult for the user to process each nugget of information manually. Following too many sources takes too much time, while following too few represents a potential loss of opportunity.

Active Twitter users report being inundated with too much information, and are only able to "dip a toe" in the torrent of tweets [2] meaning that they rarely read everything in their stream. Similarly, many people may want to be on mailing lists related to their broader interests, but have no time to read even more email than they already get, and hence limit their subscriptions to lists of direct and current interest. SLANT's approach of funneling social chatter into a personal search engine benefits users by enabling them to indirectly consume streams of information they may otherwise have missed – not by reading them explicitly, but by capturing recommendations embedded in them. Even for messages they have already read, it is difficult for people to remember all the recommendations and links they contain, and it is therefore valuable to pipe them automatically into a personalized search engine. We envisage that such implicit consumption (and extensive collection) of social information will become commonplace in the future, resulting in users being able to follow more people on Twitter or subscribing to more mailing lists.

Privacy Considerations

Today's search engines collect detailed user information such as search history, location, profile, and social data in order to personalize web search. SLANT tackles some of the privacy issues inherent in this model by allowing the user to analyze her communications under her own control. Only the results of the analysis, such as preferred sites or terms of interest need to be part of the user's search index; even if this index is then given to a search engine, users have better privacy since the search engine is not listening in on their detailed communications. A further benefit is that since only the user has access to all of his or her social data, the personalization can be relatively accurate and complete. In contrast, commercial search engines tap into social information only from specific sources of social media, based on commercial agreements between them and the source. For example, as of July 2011, Google has ended the commercial arrangement with Twitter that let a user's friends tweets appear in search results [27]. Similarly, Google search has no visibility into a user's Facebook information.

Spam Elimination

An important benefit of using social media is the elimination of many kinds of spam and artificial gaming of search results. Users choose who to follow in Twitter; and while one can receive email spam, SLANT uses only messages in email folders that the user has chosen. The problem of spam filtering in email is also independent of link-indexing and search, and is addressed by an independent body of work.

Contributions

We now summarize the main contributions of this paper.

- *Socially curated search.* We propose improving the quality of web search by extracting information from users emails and Twitter feeds to create personalized search indices. We designed four types of search indices that can be useful: links in email, links in tweets, friends' names, and links from top tweets.
- *Evaluation of social and curated search.* Using an experimental system that we built, we perform an exploratory study to compare the efficacy of different kinds of curated search by asking users to evaluate the performance of each search index on queries from their organic search history.
- *Empirical results.* We report quantitative results from our study, and discuss examples that provide qualitative insight. Briefly, we discovered that the quality of results from the email and Twitter-based indices were rated slightly higher than personalized Google search. We provide insight into the types of queries that do well with different search indices.

Interested readers can try out the SLANT research prototype which is publicly available at the URL: `http://mobisocial.stanford.edu/slant`.

RELATED WORK

Human curation

While automatic search algorithms currently play a large role in searching for information, there is much evidence that humans are also an important source of information. The success of social bookmarking services like del.icio.us and social news services like Digg is evidence of this fact. Social recommender systems also take advantage of human curation. For example, the zerozero88 system analyzes a user's tweets to recommend websites that may be of interest to her; however, it does not automatically use these sites for web search [8].

Search personalization

The closest related work to ours is by Teevan et al. who studied the modeling of user interests, built from both search-related information and documents and email the user has read and created [28]. They attempt to use this information to create a profile of the user and re-rank the top 50 results returned by a search engine for each query. While their approach has better privacy guarantees than ours because it works purely on the client side, it cannot uncover results that may be buried deeper down in the search engine's ranking. It also does not utilize information embedded in social streams like Twitter. Another client-side solution that attempts to diversify results is due to Radlinski et al [23]; it does not utilize social data either. Luxenburger et al. studied personalization techniques for queries that are expected to benefit from prior history information and selectively applying personalization techniques [16]. They correctly point out that personalization does not always improve search quality.

Prior work indicates that users frequently have a need to re-find information that they have encountered previously [1]. Therefore it appears plausible that pages and sites encountered in social data in some form have a high chance of satisfying a user's information needs.

Commercial search engines

Commercial search engines such as Google, Bing and Blekko have attempted to incorporate social elements into their search results. While the specific ranking algorithms are not publicly available, we summarize some of their features. Until recently, signed-in Google users who have connected their Google and Twitter accounts could see which of their Twitter followees have tweeted about a specific web page [7]. However, Google promoted the result only when the specific page in a search result had been tweeted (and not other pages on the site containing the page.) Bing has similar features, but is integrated with Facebook instead of Twitter, and therefore processes page "likes" on Facebook [17]. Blekko integrates with Facebook and allows users to specify a */likes* directive with search queries, which reports results from sites (not just pages) "liked" by their Facebook friends [4]. However, in our experience, Blekko's results tend to be noisy because when friends like any page on a very popular site, then all pages from that site are considered as endorsed by the friend, and tend to dominate search results.

All these search engines are limited to the information that the user allows them to access, and tend to be fragmented because different services access different parts of the user's social network that they can pick up. They also require full access to the user's social data, and none of them take advantage of links and other information embedded in email. In contrast, the SLANT approach not only gives users better control of their own data and privacy, but can also take advantage of multiple sources of chatter surrounding the user.

Social Q&A

There are several efforts to query social networks directly in response to an information need. Evans et al. analyze strategies employed by users when looking up information from their social network [11]. They asked users to learn about topics that were hard to directly search for, and yet found that social searching, even when including Q&A sites such as Yahoo Answers and the ability to ask questions of their social network, did not perform as well as with traditional web search. Morris et al. study scenarios of when users directly pose a query to their social networks and compare them to traditional web search [20, 21]. Smith et al. propose using social context i.e., a user's friends and the communities to which the user belongs to improve search for social media [26].

Other systems like Aardvark [14] and Collabio [3] attempt to steer user questions towards a person who is qualified to address their information need. In contrast to these systems, SLANT combines the best of both worlds: automatic algorithms to harness the large pool of web content, and social curation to select high quality information sources.

Collaborative searching

Search Together is a system that lets users collaborate in order to complete a specific task [18]. While links are exchanged between users, those links are used only for the duration of the task. Similarly, there has been much work in designing techniques to let groups work together on search tasks in a collaborative manner, whether or not they are co-located or working synchronously (e.g. [19]). In contrast, SLANT indirectly captures implicit and long-term collaboration, not focused on any specific task. Moreover, users are not even aware that they are "collaborating" with others; a link shared by a friend today may help improve a search result a year from now.

Personal data mining

Systems like Phlat [9] and Stuff I've seen [10] allow a user to browse and mine their personal data on a local computer including emails, documents and files, but have not been integrated with web search. Such mining can be exploited to improve the results of web search queries, without revealing all of the user's personal information to a service provider.

SOCIAL SEARCH INDICES

In this section, we describe four personalized search indices that we built in SLANT to exploit various kinds of social chatter. We developed these four indices by observing the results on our own search queries over time, and tuned them by deploying early test versions with ourselves and a few lead users. We performed informal user studies along the way which helped us refine the algorithms and improve precision and recall. We describe below how we extract links and other information from social data; in the next section, we will explain how these are weighted and used to create a personalized search index for the user.

Email links index

Email archives are very personal and capture a lot of information directly relevant and meaningful to the user. Email is widely used, with over 1.8 billion users worldwide [22]. Moreover, users routinely use email to send each other useful links, which constitutes an excellent form of curation. An additional advantage of email is that it is common for users to retain access to their email archives, spanning at least a few years, for their records. To create the email-based index, we ask users to download a modified version of the MUSE program to their computer [13], login to one or more email accounts and specify email folders from which links are to be extracted. Typically, we recommend that users select their sent mail folders and any other folders which they think has high-quality information. We parse the body of email messages in the specified folders and extract links from each message.

Friends' names index

People are generally interested in their friends' opinions and experiences. For example, when selecting a movie to watch, people will take a close friend's positive recommendation over an expert movie critic's negative review. MUSE extracts names of email correspondents from email messages (after performing entity resolution to account for multiple email addresses or name spellings for the same person) and ranks them using communication volume as an approximate proxy for tie strength with that person. We use these names to perform query expansion – for each user query, we fork off multiple queries to the Google search engine by simply adding the name to the text of the query. For example, when a user with a friend named *John Riedl* searches for the term *Minneapolis*, a query of the form *Minneapolis +"John Riedl"* is issued. We currently limit query expansion to the top 150 friend names in order to keep response times reasonable (in practice, this works out to a latency of about 1-5 seconds per query, which is tolerated by most users for experimental purposes). We score each link in the returned search result for each expanded query by simply accumulating the score for the user whose name was issued with the query. The results are assembled and ranked for the user, and presented using the same interface as normal Google search, including snippets. As we will describe later, users enjoy serendipitously discovering search results that involve their friends.

Users sometimes have friends with names like *Mike Smith* that are extremely common, and therefore lead to noisy search results. Our interface allows users to mark off search results with such noisy names so that future results with that friend's name are ranked below other results. We typically find that users encounter two or three friends with such names, but once these names are demoted, the noise level is considerably reduced.

Twitter links index

To form the Twitter link index, users provide SLANT access via Oauth to their Twitter account. We extract all the links in tweets from accounts followed by the user. For each link, we check whether it belongs to one of about 90 URL shortening services like bit.ly, tinyurl, etc. If it does, the URL is replaced with the original URL pointed to by the shortened one.

As mentioned earlier, piping the links in a Twitter feed to a search index greatly expands the user's ability to follow more Twitter users, and implicitly consume an expanded tweet stream. We envisage that a user may have two categories of followees: one whose tweets he actually reads, and one whose tweets are implicitly consumed by feeding URLs into a search index. Our approach also works when the user is interested in a subset of topics about which a followee tweets. For example, if Alice is interested in restaurants, and follows Bob who is interested in restaurants as well as in religion, Alice may import religious links tweeted by Bob in her search index. However, we expect that Alice will tend to perform queries relevant to her areas of interest, and therefore irrelevant links effectively get ignored.

TopTweets index

Twitter publishes a listing of top tweets that have caught the attention of many users based on retweeting and sharing patterns [29]. We create a search index based on these top tweets to understand the performance of content that, while being human-curated, is not personalized for individual users. These links also allow those users who do not have a Twitter account or do not use it actively to experience the effect of human-curated search.

SEARCHING WITH SOCIAL INDICES

We implement the backend to the search indices in SLANT as a layer on top of the Google search engine, though other search engines with similar features could conceivably be used as well. When the user types in a search query, we perform the search using each of the four personalized search indices described above, along with regular Google search which is itself personalized if the user is logged into Google.

For searching sites with links extracted from Twitter and email, SLANT uses the Google custom search engine [12]. The custom search engine restricts search results to specified sites. Weights can be assigned for each site and up to 5,000 URL patterns can be specified, which turns out to be adequate for the users in all our experiments. The Google custom search engine is popular with publishers and is frequently used by them to offer a customized search service restricted to their own sites, using Google's search technology.

SLANT instantiates a new custom search engine and populates it with the domains containing these links. For example, a page like *http://example.com/123* would cause the

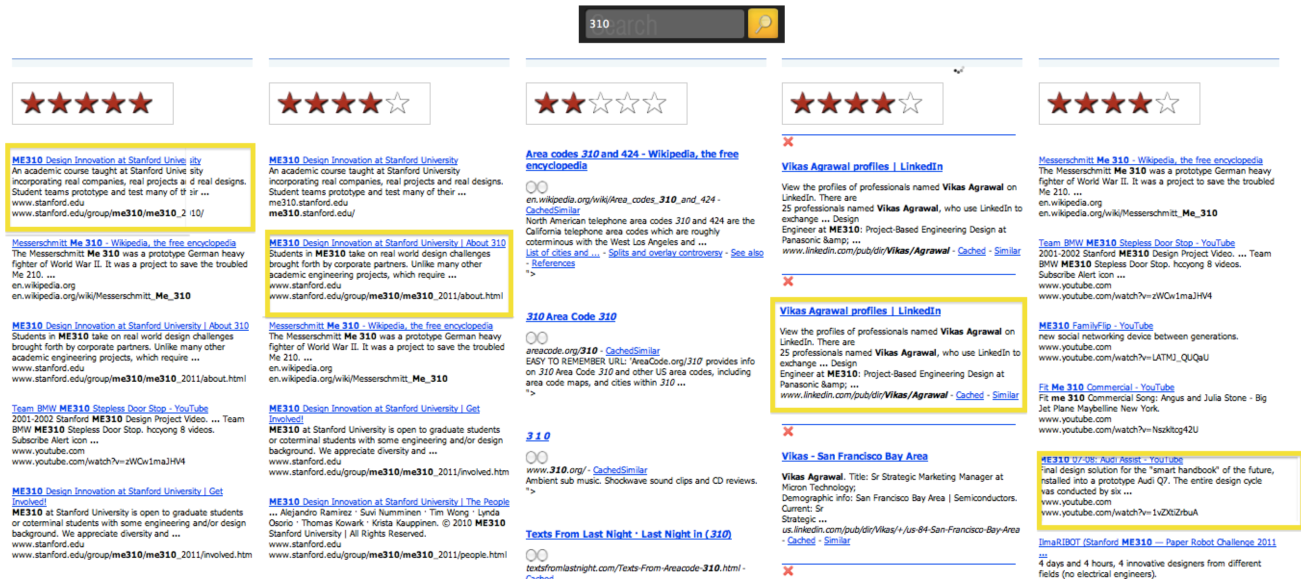

Figure 1. A screenshot of the interface used in our study. Results from the five different searches are shown side-by-side. From left to right: email-based, Twitter-based, personalized Google, friends names, and TopTweets. The user rates each set of results on a scale of 1 to 5. This screenshot shows the results when a user has issued the search query *310*. The results liked by the user for this query are highlighted in yellow.

domain *example.com* to be included in the search engine. We bias the weighting of a domain based on the frequency with which it appears in the extracted links. Weights for the Google custom search engines can be specified on a scale of -1 to 1. We assign weights to domains based on two factors:

- A positive bias to reflect the popularity of the domain in the corpus. The bias scales linearly from 0.1 to 1 depending on the domain's number of occurrences in the corpus.
- A negative bias to prevent major websites such as aol.com and amazon.com from overshadowing the more personalized results. We implemented this negative bias after we found them dominating search results, defeating the purpose of personalization. This bias is applied to a domain if its rank is one of the top N sites on the Internet (in our experiments, N=100). It scales linearly from -0.1 for the N-th ranked site to -1 for the topmost site.

In summary, each user's index is represented by a weighted vector of domain names that appear in the corpus. The weight for domain i is

$$w_i = \begin{cases} 0.1 + 0.9 \times c_i / c\text{max} & \text{if } r_i > 100 \\ 0.9 \times c_i / c\text{max} - 0.9 \times (100 - r_i)/100 & \text{otherwise} \end{cases}$$

where

r_i is the rank of domain i in the Alexa rating,
c_i is the number of times domain i occurs in the user's corpus, and
$c\text{max} = \max_i c_i$.

User Interface
As an experimental interface, SLANT presents the user with a single query box, and when the user types in a query, it fetches results from all five search indices. The results are presented side by side (see Fig. 1). This interface is designed only for prototyping and evaluation purposes; in a real deployment, we envision other interfaces such as a browser sidebar that combines results from all the social search indices.

An Example
To give readers a flavor of how these search indices differ, Figure 1 shows the results of a query actually performed by one of our users: the term *310*. Personalized Google search returned results about telephonic area code 310 as the top link. However, the user has in mind the course ME310, colloquially referred to among his friends as 310. The results from the indices seeded from email and Twitter links capture the user's context better and return results related to the course. The people search index returns a page with the name of a friend who has attended this course in the past, and has it on his resume. The TopTweets index has Youtube as one of the domains, and videos from the ME310 course appear in its results. In this example, each of the four social search engines in SLANT was rated higher by the user than personalized Google search.

EXPERIMENTAL EVALUATION
To evaluate the different search indices we designed, we performed a user study with 7 subjects (2 female) in the age group 20-30, all students at our university and regular users of web search. We recruited users who self-reported themselves as having been active on Twitter for a period of at least 6 months, and to have access to an email account they had actively used for more than 1 year. On Twitter, our users followed a fairly typical assortment of friends, researchers, celebrities, journalists, university accounts, popular brands, etc.

To work around the limits imposed by the Twitter API and gather a sizeable corpus of tweets, we collected tweets over 4 rounds, at 7-10 day intervals for each user. For the email links index, users selected a subset of email folders (in addition to their sent email folders), which they judged to contain high-quality information. As mentioned earlier, this process is conducted entirely locally on users' own machines in order to assure them about privacy.

Users in our study had

- 5,397 to 11,398 emails over a couple of years,
- 807 to 1,442 unique domains in their email,
- 185 to 334 Twitter followees,
- 1,038 to 1,548 unique domains in their Twitter feed.

More details about these parameters are shown in Figure 2. The number of domains in each index is a few thousand, representing a tiny percentage of web sites that are actually present on the web and crawled by commercial search engines.

	Min	Median	Max	Mean
Number of Twitter friends	185	262	334	260
Number of Email messages	5,397	7,618	11,398	7,847
Unique domains from Twitter	1,038	1,201	1,548	1,231
Unique domains in email	807	1,161	1,442	1,150

Figure 2. Statistics on the personal search indices

We conducted the experiment with each participant independently. All our participants used Google as their search engine by default. They were also generally logged in during their routine searches, except when using mobile devices. To lend ecological validity to our experiment, we asked them to look at their search history, (gathered by Google for signed-in users and available at `http://www.google.com/history`), and select about 10 queries to try out with SLANT. Users were thus trying out our search indices with queries they had already organically performed on their own.

Users were provided with the test interface shown in Fig. 1 and asked to enter their queries, while remaining signed in to Google. They were asked to carefully evaluate each set of search results (limited to the top 10 results), and rank the quality of each set of results on a scale of 1 to 5. Participants were not told which column corresponded to which search index, in order to elicit unbiased feedback. Typically, users spent about 60 minutes performing 8 to 9 queries and evaluating and rating the results from each search index. The whole process including the mining of email and Twitter took about 6 hours per user. In all, we gathered detailed information on 59 queries across our users.

We recorded all queries and ratings for analysis. We employed a think-aloud methodology and recorded qualitative user comments to gain additional insights. While we recognize that this process may somewhat affect which queries users picked from their history, we judged it necessary; the specific queries and comments provide valuable insight into the efficacy of each search index.[1]

[1]The supplemental materials included with this paper contain a table listing all queries, along with their categorization, ratings and user comments on the results.

For the TopTweets search index, we gathered links posted by the TopTweets account once a week, over 6 weeks. This process ensured that the index had a sizeable corpus to work with; in all, it consisted of 1,073 domains.

RESULTS

In this section, we present a quantitative analysis of the results of our user study. Our analysis of the results starts by testing the hypothesis that users tastes are very different, and that their search queries are satisfied by different sources. We then analyze how the different search indices in SLANT perform and classify queries to understand the relationships between query types and evaluation results. Next, we evaluate whether different search indices provide different results and show evidence that combining these results can be very effective.

Similarity in Content of Social Chatter

Restricting search to domains mentioned in social chatter achieves two objectives: (1) the domains are human curated and (2) they are personalized. If most users are interested in the same set of the most popular sites, then there would be little need for personalization and we could just adopt a single human-curated index.

	User 2	User 3	User 4	User 5	User 6	User 7
User 1	0.16	0.05	0.09	0.05	0.14	0.12
User 2		0.12	0.03	0.07	0.13	0.11
User 3			0.18	0.10	0.06	0.08
User 4				0.03	0.12	0.08
User 5					0.04	0.07
User 6						0.13

Figure 3. Cosine similarity between weighted email indices of different users.

Fig. 3 shows the cosine similarity between the domain-weight vectors for the email indices of each pair of users. The vectors are normalized so all referenced domains have a positive weight while preserving the relative ordering with each other. The cosine similarity ranges from 0 to 1, with 0 indicating orthogonality and 1 indicating equivalence. As we can see, the similarity between users is fairly low. This is true despite the fact that our seven users, all being students at the same university, are more likely to have similar interests than people in the general population. This indicates that different users tend to have different sites in their social chatter and therefore personalization is likely to be highly beneficial.

Rating the Search Indices

We first summarize the ratings users gave to the different SLANT indices. Our baseline, personalized Google search, receives favorable average ratings from most users, ranging from 2.9 to 4.1. Both Twitter and email-based indices also obtained similar ratings. The Twitter-based search index rates between 3.2 to 4.1, while the email-based search index rates between 3.3 and 4.1. For 4 out of 7 users, the email-based search was rated the best overall, compared to two for the Twitter-based index, and one for personalized Google search. This shows that social chatter based in-

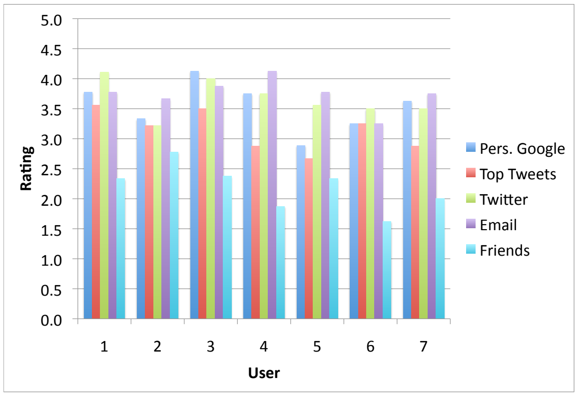

Figure 4. Average ratings of searches using different indices.

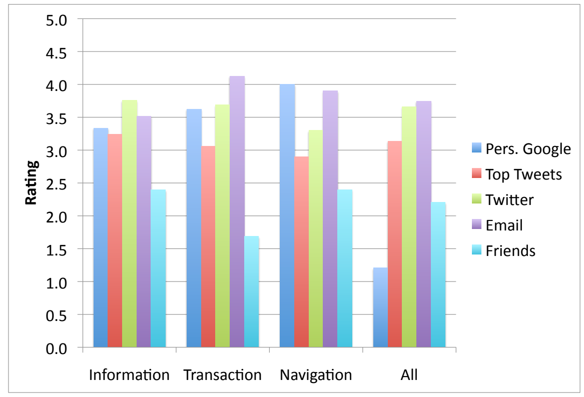

Figure 5. Average ratings across types of queries.

dices can match or outperform normal personalized search in terms of user satisfaction with results.

The search indices based on TopTweets and friends' names are not directly competitive with the other two personalized indices. The TopTweets index achieved average ratings from 2.7 to 3.6, and the index with friends names was rated between 1.6 and 2.8.

Categories of Queries

To gain better insight into the strengths and weaknesses of the different search indices, we used Broder's web search taxonomy [6]. This taxonomy classifies queries into three categories: *informational*, where users are motivated to get more information about a subject; *transactional*, where users are interested in performing a certain transaction, such as shopping or finding a file to download, and *navigational*, where users are looking for the URL of a specific site. Broder's analysis of query logs found that the distribution between informational, transactional, and navigational queries was 48%, 30%, and 20% respectively.

We manually classified the 59 query terms generated by our 7 users into one of these three categories. In some cases, if the categorization was somewhat ambiguous, we contacted the user and asked about his or her original intent. In these queries, the distribution between informational, transactional, and navigational queries is 56%, 27%, and 17% respectively, which is somewhat similar to the ratio reported by Broder.

Figure 5 shows the average ratings across the three types of queries, as well as the overall average. The overall averages across Google, Twitter, and email are about the same: 3.5, 3.7, and 3.8, respectively. Looking at the results of the different categories, we observe the following:

- *Informational.* Both Twitter and email based-indices fare slightly better than personalized Google search in the informational category. Informational is the most important category, carrying over half of all the queries, but it is also the hardest to do well on.
- *Transactional.* Email is significantly better in the transactional category, with an average rating above 4.
- *Navigational.* Both Google and email are better than Twitter in this category.

Qualitative Analysis

Both the Twitter and email-based search indices in SLANT index a tiny fraction of the world wide web. It is therefore surprising that their results are comparable to, or even better than, regular web search which indexes every one of the billions of web pages. In fact, before the study, several of our participants indicated skepticism about our search indices due to this reason. To understand why and how these indices work well, we now discuss some actual queries and comments from users.

Informational Queries

Examples of queries where personalized Google search performs best are *health hazards of clementine fruits* and names of famous personalities like athletes and political activists. In these cases, Google provides the general information sought by the user, such as a Wikipedia entry or a news page. For example, a search on *wikileaks* also earns Google the highest score because the user was happy to see the latest news about it.

Email has the most number of unique high scores in this category. A primary characteristic is that the user typed in just a word or two and was presented with information of specific interest. As we reported earlier, a user who searched for *"310"* was looking for a course numbered ME310 and this was captured in the user's email custom search. In this case, the user exclaimed *"Wow this needed so little typing"*. Similarly, a search for *iphone5* returned an article from the technology blog TechCrunch, a favorite of the user, and a search for the single word *"dock"* returned the Bower and Wilkins dock, related to the user's interest in audio docks.

The Twitter search also returns several unique high scores. For example, the search *Simpsons Monday joke* returns the exact joke that the user was looking for. Users also tended to like blogs and articles on the topics searched.

This suggests that these search indices have different strengths. Google search tends to find general information for all queries. The Twitter links index returns results for what is trending, as well as interesting blogs and articles written by authors who the user is personally interested in. Finally, email contains a lot of links to domains of personal interest. We speculate that social chatter is useful for informational queries because users are often searching for information discussed in their conversations. SLANT is also particularly useful for query terms that are commercially valuable and therefore likely to attract web spam. An interesting effect we have observed is that SLANT results tend to be more personalized and colorful than those from regular web search. One user remarked, *"I don't want results to be dull, like Wikipedia showing up on every query."*

Transactional Queries

SLANT's email-based index was clearly the best in the transactional category, and was rated 5 on half of the queries. We surmise that this is because email typically captures the user's personal context well. Email often includes personal information such as bank transactions, addresses, receipts, shopping information, etc. When users query for transactions related to a party they have already communicated with, this index is likely to generate good results. For example, a query like *lost cheque book* performed well with this index, without the user needing to specify details like the name of the associated bank. One user who searched for a query *disable unc check* related to compliance software was surprised to see a good result from a domain which had been emailed to him. Another user searched for *Russian roulette rihanna download* and remarked how she liked the result from a site she uses regularly.

Notably, Google search is rated the highest on all map and direction requests, particularly because it directly displays a map. It also provides the only 5-point rating on the query *matlab registry key*. Google is likely to perform well when looking for very specific information about a transaction. The Twitter-based index does better on queries like *Cauvery fishing camp* and *mac iso image mount*. It is likely that the human-curated links tweeted about these topics are of higher quality than the top Google links, and once again, avoid attracting spam.

Navigational Queries

Navigational queries are relatively unremarkable – we recorded almost no comments on the results relating to the top three indices. For these queries, all the Twitter ratings were between 2 and 4, while both Google and email score 5 on 40% of the queries. This is perhaps not surprising since Twitter is primarily used to share noteworthy news, and less for mundane matters.

Friends Names Index

Although the search index based on friends names performed poorly overall, it has turned up a number of unexpected and serendipitous results that delighted users. For example, one user searched for *parasailing san diego* and was amazed to find a Picasa photo album of a friend's visit to San Diego. Another searched for *Boston University* and discovered that a colleague of his was going there for his MBA. In yet another instance, a professor was looking up *Zigbee* and found that one of her students had written a paper on the topic, a fact she had not known. This kind of serendipitous finding, while rare, is remarkable. A topic for future work would be explore methods to improve the chances of this happening, and to estimate the quality of results from this index, so that they can be shown only if they are likely to surprise the user.

Query Length

It is well known that average search query length has been going up over the last decade [15], and that users need to formulate more precise queries to obtain accurate results. It is customary for users to iterate on queries when the results do not turn up answers they need. A couple of our users remarked how little typing was needed to get the results they wanted. Twitter and Email take advantage of the users' context to help disambiguate information, thus minimizing typing. The *310* query mentioned earlier in this paper is one such example. The ability to specify shorter queries is particularly useful when searching on mobile devices and other limited function input devices such as TV remotes, where typing is inconvenient.

Correlation Between Search Engines

Although the average ratings of the top three indices are quite similar, we observed that the ratings for each individual query vary quite widely. In addition, the qualitative results suggest that different searches perform well for different reasons. This leads us to ask if the domains present in the search indices and their weights are correlated to each other. The Pearson correlation coefficients between the ratings for each pair of search indices are shown in Figure 6. Except for a slight positive correlation between Google and TopTweets, there is very little correlation between any other pair of search indices. This lack of correlation suggests that the indices are complementary.

	TopTweets	Twitter	Email	Friends
Pers. Google	0.21	0.05	-0.08	-0.08
TopTweets		0.13	-0.04	0.10
Twitter			-0.10	0.05
Email				0.08

Figure 6. Correlation between different search indices.

Combining results from multiple indices

Can we improve search by combining multiple customized searches? While investigating such a design is beyond the scope of this paper, our results can be used to compute the theoretical upper bound of a combined search with the maximum rating of its components. In our current study, each search returns ten results. A straightforward way of creating a combined engine is to simply return ten results from each search. Combining two searches will generate 20 results; combining three will generate 30. While it is preferable to return satisfactory results in the fewest number of search hits, further investigation of ways to combine the indices is desirable, given the likely increase in user satisfaction.

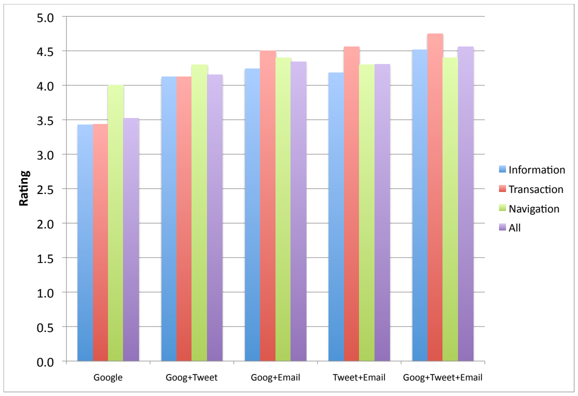

Figure 7. Combining the top performing search engines.

Figure 7 shows the possible ratings for four combinations, compared to the baseline of personalized Google search. The baseline obtains an average of about 3.5 for all categories, except for scoring a 4.0 average on the navigational category. Adding Twitter-based search boosts the average to above 4.1 in all categories; similarly, adding email-based search to Google boosts the average about 4.2, with transactional queries scoring a high 4.5 average. Interestingly, we can combine just Twitter and email-based results to achieve a similarly high rating. In other words, combining any two of the top three search indices will provide a significantly improved result, bumping up the average from 3.5 by 0.7 points, a gain of 20%. While users in our study are active on both email and Twitter, this result suggests that users who have a good corpus on only one of email or Twitter can benefit by adding their medium of choice to augment Google search.

Finally, we can combine results from all the top three engines. In this case, the average rating improves from 3.5 to 4.6, a gain of about 31%. Specifically, out of the 59 queries, in 37 queries at least one index gets the highest possible score, in 20 queries the highest score awarded is 4 and in only 2 queries the highest score assigned to any index is a 3. With over 60% of the queries receiving the best possible score, combining the three search indices can produce an exceptionally powerful search engine.

Summary

We now summarize the major findings from our evaluation.

1. The personalized SLANT indices have little similarity between individuals. This suggests that personalization is important.

2. Each user found the quality of email and Twitter-based searches in SLANT similar to personalized Google, even though the results returned are very different, and from a small subset of domains.

3. Google is better at returning general results for terms and helping navigate to specific web sites when users type in its distinguishing characteristics. The Twitter-based index is better at returning trending topics and opinions, blogs and articles of interest to users. The email-based index, being personal, helps in disambiguating terms, thus minimizing the need for typing long queries and search refinement. Email-based search is useful for transactional queries because the domains that users routine interact with are often embedded in email.

4. Combining the best results from different search indices can significantly improve user satisfaction. Combining results from any pair of Google, Twitter-based, and email-based indices boosts the average rating from 3.5 to above 4, and combining all three indices increase it even further.

5. Although the friends names index generates generally poor results, it occasionally returns highly serendipitous results that surprise the user.

DISCUSSION

Our results are evidence of the somewhat surprising fact that custom search engines that index a tiny fraction of the entire world wide web can perform comparably with traditional search engines for general-purpose web search. Our form of social search could be implemented as a mostly client-side solution, with the search engine knowing only summary details such as the domains of interest to the user. While search engines have been personalizing search results for users, there is a sense of discomfort among users in giving out detailed information such as their email and clickthrough patterns to “Big Brother” portals. One of our users commented: *“I would not like to share my personal information with Google, or for that matter any company, as I always feel that it is a definite risk of loss of privacy. There could also be accidental information leak by a company with whom I might share my details, like Blippy accidentally leaked credit card information.”*

In future, we envisage that users may see multifaceted search results. They can get authoritative and definitive results from a global search engine, along with the more personal and entertaining results from personal search engines that are automatically assembled from their social chatter.

Of course, social search indices also have limitations. We believe that the approach of generating socially curated results complements traditional search engines, but does not replace them. The quality of the indices in SLANT is dependent on the extent and quality of email archives, or links tweeted by the user’s followees. One way of increasing and enriching the size of the corpus of the social indices would be to explore tweets from multiple levels within the network, such as a followee’s followees, and assign trust and weights to the content generated by them.

CONCLUSIONS

Using SLANT, we have shown that social chatter from email and social media can be leveraged to improve web search.

Social data is personalized, free of spam and can be analyzed directly by users on the client side, alleviating privacy concerns.

User studies with SLANT reveal that search results derived from socially curated indices can be as good if not better than the results of commercial search engines, even when they use personalized search. We envision that combining results from these different indices will generate superior results that satisfy users for many types of queries.

ACKNOWLEDGMENTS

We thank Rakesh Ghiya for useful discussions, Diana MacLean and the anonymous reviewers for feedback on the paper, and all the users of SLANT. This research was funded in part by NSF POMI (Programmable Open Mobile Internet) 2020 Expedition Grant 0832820 and the Stanford MobiSocial Computing Laboratory.

REFERENCES

1. E. Adar, J. Teevan, and S. T. Dumais. Large scale analysis of web revisitation patterns. In *Proceedings of CHI '08*. ACM, 2008.
2. M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi. Eddi: interactive topic-based browsing of social status streams. In *Proceedings of UIST '10*. ACM, 2010.
3. M. S. Bernstein, D. Tan, G. Smith, M. Czerwinski, and E. Horvitz. Personalization via friendsourcing. *ACM TOCHI*, 17:6:1–6:28, May 2008.
4. Blekko. Friends make search better! `http://blekko.com/likes-info`, Retrieved Aug. 29, 2011.
5. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of WWW '98*. Elsevier, 1998.
6. A. Broder. A taxonomy of web search. *SIGIR Forum*, 36:3–10, September 2002.
7. M. Cassidy. An update to Google social search. *The Official Google Blog*, Feb. 17, 2011.
8. J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: experiments on recommending content from information streams. In *Proceedings of CHI-2010*. ACM, 2010.
9. E. Cutrell, D. Robbins, S. Dumais, and R. Sarin. Fast, flexible filtering with phlat. In *Proceedings of CHI '06*. ACM, 2006.
10. S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff I've seen: a system for personal information retrieval and re-use. In *Proceedings of SIGIR '03*. ACM, 2003.
11. B. M. Evans, S. Kairam, and P. Pirolli. Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Information Processing and Management*, 46(6):679 – 692, 2010.
12. Google. `http://google.com/cse`.
13. S. Hangal, M. S. Lam, and J. Heer. Muse: Reviving memories using email archives. In *Proceedings of UIST-2011*. ACM, 2011.
14. D. Horowitz and S. D. Kamvar. The anatomy of a large-scale social search engine. In *Proceedings of WWW '10*. ACM, 2010.
15. F. Lardinois. Hitwise: Search queries are getting longer. *ReadWriteWeb*, Feb. 24, 2009.
16. J. Luxenburger, S. Elbassuoni, and G. Weikum. Task-aware search personalization. In *Proceedings of SIGIR '08*. ACM, 2008.
17. Y. Mehdi. Facebook friends now fueling faster decisions on Bing. *Microsoft Bing Search Blog*, May 16, 2011.
18. M. R. Morris and E. Horvitz. Searchtogether: an interface for collaborative web search. In *Proceedings of UIST '07*. ACM, 2007.
19. M. R. Morris, J. Teevan, and S. Bush. Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting. In *Proceedings of CSCW '08*. ACM, 2008.
20. M. R. Morris, J. Teevan, and K. Panovich. A comparison of information seeking using search engines and social networks. In *Proceedings of ICWSM '10*. AAAI, 2010.
21. M. R. Morris, J. Teevan, and K. Panovich. What do people ask their social networks, and why? A survey study of status message Q&A behavior. In *Proceedings of CHI '10*. ACM, 2010.
22. The Radicati Group. Email Statistics Report, 2010-2014, 2010.
23. F. Radlinski and S. Dumais. Improving personalized web search using result diversification. In *Proceedings of SIGIR '06*. ACM, 2006.
24. L. Rao. Twitter Seeing 90 Million Tweets Per Day, 25 Percent Contain Links. *TechCrunch*, Sept. 14, 2010.
25. D. Segal. The dirty little secrets of search. *The New York Times*, Feb. 13, 2011.
26. M. Smith, V. Barash, L. Getoor, and H. W. Lauw. Leveraging social context for searching social media. In *Proceedings of the 2008 ACM workshop on Search in social media*. ACM, 2008.
27. D. Sullivan. As deal with Twitter expires, Google realtime search goes offline. *Search Engine Land*, Jul 4, 2011.
28. J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *Proceedings of SIGIR '05*. ACM, 2005.
29. Twitter. `http://www.twitter.com/TopTweets`.