

Exploring Online Social Activities for Adaptive Search Personalization

Qihua Wang
IBM Almaden Research Center
650 Harry Road
San Jose, California, USA
qwang@us.ibm.com

Hongxia Jin
IBM Almaden Research Center
650 Harry Road
San Jose, California, USA
jin@us.ibm.com

ABSTRACT

The web has largely become a very social environment and will continue to become even more so. People are not only enjoying their social visibility on the Web but also increasingly participating in various social activities delivered through the Web. In this paper, we propose to explore a user's public social activities, such as blogging and social bookmarking, to personalize Internet services. We believe that public social data provides a more acceptable way to derive user interests than more private data such as search histories and desktop data. We propose a framework that learns about users' preferences from their activities on a variety of online social systems. As an example, we illustrate how to apply the user interests derived by our system to personalize search results. Furthermore, our system is adaptive; it observes users' choices on search results and automatically adjusts the weights of different social systems during the information integration process, so as to refine its interest profile for each user. We have implemented our approach and performed experiments on real-world data collected from three large-scale online social systems. Over two hundred users from worldwide who are active on the three social systems have been tested. Our experimental results demonstrate the effectiveness of our personalized search approach. Our results also show that integrating information from multiple social systems usually leads to better personalized results than relying on the information from a single social system, and our adaptive approach further improves the performance of the personalization solution.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms

Algorithms, Experimentation

Keywords

Personalized search, Online social activities, Adaptivity

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'10, October 26–30, 2010, Toronto, Ontario, Canada.
Copyright 2010 ACM 978-1-4503-0099-5/10/10 ...\$10.00.

1. INTRODUCTION

The past decade has witnessed the rapid growth of the Internet. The numerous services on the Web have become an indispensable part of many people's life. Personalized applications, which customize services based on a person's preferences, have the potential to significantly improve user experiences. There are many applications that can benefit from personalization, including web search, advertisement targeting, news filtering, tag recommendation, and so on. Extensive efforts have been made on personalizing web applications and services.

A major challenge to perform personalization is to learn about a user's interests and preferences. Researchers [11, 14, 18, 13, 3] have proposed solutions that extract users' interests from a variety of information sources, such as search history, emails, desktop documents, and so on. While selecting the right sources of user information is critical to the success of a personalization solution, there are three criteria that should be considered when making a choice.

- *Availability* of user information: Information that is easy to get facilitates the creation and update of users' interest profiles. Publicly accessible information has the highest availability. In contrast, privately owned resources, such as emails and desktop documents, are available only upon users' approval and may not be easy to acquire. Moreover, information stored by a search engine provider, such as users' search and click-through history, is normally unavailable to outsiders, such as another search engine or independent researchers, due to commercial and privacy reasons. A personalization solution, either on web search or other tasks, is more flexible if it learns about users' interests from highly available information sources.
- *Privacy*: Some resources, such as emails and desktop documents, may contain sensitive information about their owners. Users may have privacy concerns on sharing those data with a search engine or even with a client-side search tool that communicates with the search engine. A personalization solution may have better psychological acceptance, if it does not explore those data with potential privacy concerns.
- *Accuracy*: When learning about a user's preferences, it is important to select those information sources that can reflect the user's interests accurately. Data sources, such as users' emails and click-through history, are good choices, but sometimes there could be much noise in the data. For example, a user's emails may contain unsolicited advertisements and broadcasting messages that do not match the user's interests. For another example, not every webpage a user clicks through among search results is truly useful to her, as the user

might be very well going through the results to find what she really wants. Information sources with low noise-level on users' preferences are highly desirable, as they can potentially save much efforts later on noise filtering.

Social networks and Web 2.0 collaboration systems have experienced explosive growth in the past few years. Outstanding examples include Facebook, Twitter, Del.icio.us, and so on. A large portion of Internet users are actively engaged in a variety of online social or collaborative activities, such as blogging, mini blogging, mutual comments, and social bookmarking. Such online activities carry valuable information about users' backgrounds and interests, which can be exploited to personalize various applications. Researchers [17, 10] have recently explored folksonomy from social bookmarking systems, such as Del.icio.us, to develop personalized search systems. Their systems learn about a user's interests based on the user's bookmarked webpages and the tags applied to those webpages. Experimental results demonstrated the effectiveness of their solutions.

In this article, we propose a personalization framework that adaptively infers users' interests and preferences through their public activities on a variety of online social systems. Our personalization system retrieves information from multiple social systems and creates an interest profile for each user by integrating different streams of information. As an example, we will personalize the search results returned by an underlying search engine based on how well they match the query issuer's interest profile. We will also show how to improve and update the user interest profiles by learning from user feedback.

In the current implementation, our system collects data from three types of social information sources; they are blogs, social bookmarks, and mutual tags. Blogging and social bookmarking are very popular online activities. Mutual tagging, which is often referred to as people tagging in existing literature [7, 6], allows users to apply tags on each other, and tags from different users are aggregated and viewable on one's profile. Tags applied to a user usually describe the user's attributes, such as her affiliations, expertise, and the projects she has been involved in. Hence, people tags is also a quality source for studying users' backgrounds and preferences. By extracting information from multiple social systems, we are more likely to acquire comprehensive information about users' interests than existing solutions that rely on a single social information source.

Exploiting information on social systems for personalization purposes has a number of advantages. First of all, social resources, such as blogs, mini blogs, bookmarks, and mutual tags, are often-times publicly assessable. The primary goal for most people to write blogs or tag a webpage is to share their thoughts, interests, or findings with others. Hence, information from social systems has very high availability, as opposed to private information such as emails and browsing history. Second, exploring published information and resources on social systems does not violate user privacy. Since the main objective of online social activities is to share, users' published information and resources on social systems are meant to be assessed by others. Finally, very accurate information about users' interests can be learned from their online social activities. Many online social activities, such as bookmarking and blogging, are actively initiated by users (in contrast to emails, which could be passively received). Therefore, noise level, regarding user preferences, is relatively low. For example, people actively express their opinions when they write a blog, and they are unlikely to bookmark something that they do not care about.

There are a couple of challenges in exploring data on various social systems to performance personalization. The first one is the in-

tegration of data from multiple social systems. In the real world, all users are not equally active in each type of online social activities. For instance, a user may be a frequent blogger but rarely bookmark webpages. This indicates that the various information sources we consider may not be equally important for deriving the interests of different users. The second challenge is to determine to what extent a user's interests should affect the underlying services, such as the ranking of search results. Personalization can have negative impacts on user experiences, if we overdo it. We tackle the above challenges by proposing an adaptive solution. Our system automatically adjusts the weights of different information sources as well as the degree of personalization based on past performance. For each individual user, our system aims to find a setting that works best for him/her.

We have performed experiments to evaluate our personalized search system. Our data set consists of blogs, social bookmarks, and people tags from real-world online social systems. Over two hundred users from worldwide who are active on the three types of social activities have been tested. Our experimental results demonstrate the effectiveness of our solution. Our results show that integrating the information from multiple social systems leads to better performance than relying on a single source. The results also support the advantages of being adaptive in personalization degree and source weights.

Finally, we emphasize that while the discussion and evaluation in the rest of this paper are on personalized search, the techniques we describe can be applied to other personalized applications. Our key contribution is developing an adaptive approach that derives users' interests from public information in various social systems.

The rest of this article is organized as follows. In Section 2, we describe the design of our personalized search system in detail. In Section 3, we present the experimental results of our solution. In Section 4, we discuss future work and related topics. Finally, we study related work in Section 5 and conclude in Section 6.

2. SYSTEM DESIGN

In this section, we describe our personalized search system. We will first give an overview of our solution and then describe the major components and steps in detail.

2.1 Overview

Our personalized search solution consists of the following steps.

First, the system extracts information on users' interests from their activities on different online social systems. An *interest profile* is built and maintained for each user.

Second, upon receiving a search query from a user, our system processes the query as follows.

1. The system forwards the query to an underlying search engine. The search engine then returns a number of webpages relevant to the query. Each webpage is associated with a *relevance score*; the more relevant and important a result is with regards to the query, the higher its relevance score.
2. The system retrieves an *interest vector* from the user's interest profile. Each element of the interest vector contains a word and a score indicating the degree of interest. When constructing the interest vector for a user, contributions from different social information sources may carry different weights.
3. For each of the top n webpages (say, $n = 100$) returned by the search engine, the system computes an *interest score* based on how well the webpage matches the user's interest vector.

4. For each webpage, the system computes a final score by combining its relevance score and its interest score. There is a parameter called *personalization degree* that determines how much the final score is affected by the interest score.
5. The system sorts the webpages according to their final scores and returns the stored list of results to the user.

Third, the system makes adjustments based on the user’s response to the search results. Parameters that may be adjusted include the personalization degree and the weights of different social information sources. The personalization degree will be increased (or decreased), when personalization leads to better (or worse) search results for the user. Similarly, the weight of an information source increases (or decreases), when the information source is found to be contributing good (or bad) information about the user’s interests. The goal of the adjustments is to find a set of parameters that works well for a specific user.

2.2 User Interest Profile

We create an interest profile for each user in the system. A user interest profile is represented as $\langle V, W, p \rangle$, where $V = \{v_1, \dots, v_k\}$ is a set of *interest vectors*, $W = \{w_1, \dots, w_k\}$ is a set of real-number weights of the corresponding interest vectors, and p is a real number called the *personalization degree*. An interest vector v_i stores the user’s information retrieved from the i th social system. Each item in an interest vector v_i is a pair consisting of a keyword t and a real-number score s . The larger the score s is, the more interested the user is at the keyword t . The weights in W are used when combining the interest vectors in V into an overall interest vector for a user. The personalization degree p determines to what extent a user’s interest profile may affect the ranking of search results.

In the rest of this subsection, we focus on the creation and combination of interest vectors. Personalization and the adjustment of W and p will be discussed in later subsections.

Creating interest vectors. For every user, we extract information on her interests from a variety of online social activities. Potential social information sources include, but are not limited to, blogs, mini blogs, social bookmarks, and mutual tags. There are different ways to create an interest vector, depending on what the corresponding information source is. For text resources, such as blogs, we extract the most important keywords from every blog and add those keywords to the user’s interest vector. The score of a keyword may be the number of the user’s blogs that contain this keyword. For tag-based resources, such as mutual tags (i.e. people tags), the tags are treated as keywords in the interest vector. The score of a keyword can be the number of people that have tagged the user with such a keyword. For example, if Alice has been tagged with “biker” by five other people, then “biker” is a keyword with score five in Alice’s interest vector on people tagging. Certain social information contains both texts and tags. For instance, social bookmarks include bookmarked webpages and the tags applied to those webpages. We combine the previous two methods to create interest vectors for such kinds of resources.

For each user, we normalize the scores in her interest vectors into $[0, 1]$. This is to facilitate the combination of multiple interest vectors into an overall interest vector for a user.

Combining interest vectors. As stated earlier, every user’s interest profile contains multiple interest vectors, one from each social information source. To get a comprehensive picture of a user’s interests, we need to combine those individual interest vectors into an overall interest vector for the user. Assume that there are k interest

vectors in the user’s profile. Let T_i be the set of keywords in the i th ($i \in [1, k]$) interest vector, and let T be the set of keywords of the overall interest vector. We have $T = \bigcup_{i=1}^k T_i$. For every word $t \in T$, its score $s(t)$ in the overall interest vector is computed as $s(t) = \sum_{i=1}^k s_i(t) \cdot w_i$, where $s_i(t)$ is the score of t in the i th interest vector ($s_i(t) = 0$, if $t \notin T_i$) and w_i is the current weight of the i th interest vector. In other words, those interest vectors are combined with different weights.

As mentioned in Section 1, all users are not equally engaged in all the online social activities we consider. Therefore, when combining individual interest vectors for a user, different interest vectors should carry different weights. For example, if a user has written a lot of blogs but rarely bookmarks anything, we may attach more importance to the interest vector constructed from her blogs than the one from her bookmarks. The reason is that, a more abundant information source tends to carry more accurate and comprehensive information about the user’s preferences. This motivates us to assign different initial weights to different interest vectors, and adjust the weights over time according to how well each interest vector contributes to finding what the user really wants.

For efficiency consideration, we adopt a simple approach on source weight initialization. We initialize the weights of various information sources by assigning a weight proportional to the total amount of information provided by the corresponding information source. For instance, if the user has 100 blogs and 50 bookmarks, then the initial weight of the blog vector may be twice of the initial weight of the bookmark vector. We will discuss the adjustment of source weights in Section 2.4.

Updating interest vectors. Users will continue using the online social systems and generate more data after our search system has created interest profiles for them. In order to stay updated, we may periodically crawl new data from social systems. Integrating new information into a user’s interest profile is simple: we just need to add new keywords to the corresponding interest vector and/or update the weights of existing keywords in it. There is no expensive computation or re-training for profile updates in our solution. Furthermore, if a new social information source becomes available, we may simply add a new interest vector to the interest profile of each user and make use of the new data. Finally, we give higher priority to new data by decaying old data periodically, since new information better represents the current interests of users.

With the growing popularity of social networks, more and more people are actively involved in one or more online social systems. Taking advantages of new information from social activities enables us to maintain an updated interest profile for almost everyone, including those users who do not use our search system very often.

2.3 Search Result Personalization

In this subsection, we study how to personalize search results based on a user’s interest profile. Our solution employs result processing to perform search personalization and may be used with most existing search engines.

First of all, upon receiving a search query from a user, we forward the query to the underlying search engine. The search engine will then return a list of webpages for the query. The search engine will also associate a relevance score to each returned webpage, such that webpages which are ranked higher (i.e. those considered to be more relevant and important to the query) have higher relevance scores. The relevance scores will be normalized into $[0, 1]$. In case the underlying search engine does not return relevance scores for search results, we may estimate a relevance score based on the ranking of the results. For example, we may use $1/(1+k)$ as the relevance score of the k th webpage in the result list.

Second, we combine the individual interest vectors in the user’s profile to acquire an overall interest vector for the user, using the approach presented in Section 2.2. For each of the top m webpages (say, $m = 100$) returned by the underlying search engine, we compute an interest score based on how well the webpage matches the overall interest vector. In our current implementation, the interest score is proportional to the cosine similarity between the word vector of the webpage (or the snippet of the webpage returned by the search engine) and the overall interest vector of the user. Note that we compute an interest score only for those high ranked results so as to make our personalized process more efficient. Lower ranked results have low relevance scores and are thus unlikely to be ranked high even if they get relatively high interest scores. Furthermore, it is undesirable to bring results that are not very relevant to the search query to the front of the list, even if they match the user’s interests well.

Third, we compute a final score for each of the top m webpages. Let $g_r(x)$, $g_i(x)$, and $g_f(x)$ be the relevance score, the interest score, and the final score of webpage x , respectively. We have

$$g_f(x) = g_r(x) \times (1 - p) + g_i(x) \times p$$

where p is the personalization degree in the user’s profile. A user’s personalization degree may be initialized to a value in $[0, 1]$ (say, 0.5) and will be adjusted over time, depending on whether the current personalization setting is effective for the user. Different users may end up having different personalization degrees. A higher personalized degree leads to larger influence of the user’s interest profile on search results.

Finally, the top m webpages are re-ranked based on their final scores, and the new list of sorted results is returned to the user.

2.4 Adaptive Adjustments

In this subsection, we discuss the adjustments on the parameters W and p in interest profiles so as to better serve each user. After answering a search query, our systems observes the user’s reactions and determines which results are really useful to the user. A straightforward approach is to mark those results clicked by the user as useful. In practice, the user sometimes needs to go over a few webpages before finding what she really wants. But most of the time, links that are clicked are more relevant to what the user is looking for than those that are not. Researchers have proposed improvements over the straightforward solution to better determine useful results, such as observing whether the user has to revise the original query and search again after clicking through a few results. Further discussion on determining which search results are truly useful to the user is beyond the scope of this paper. For simplicity, our system considers those clicked-through links as useful to the user.

Adjusting personalization degree. For some users, their interest profiles correctly represent their preferences, and our personalization scheme is able to rank what they are looking for high in the result lists for most search queries. In those cases, it may be desirable to increase the personalization degree so as to bring more interesting webpages to the front for future search queries. On the contrary, personalization may not be effective for some users. There are a couple of potential reasons why our scheme may not work well. First, our system may fail to extract the right information from the users’ online social activities. For example, the keywords we extract from some blogs may be not correctly reflect the topics of the articles. Second, the information we acquire from online social activities and the users’ search queries may be in different contexts. For instance, a user may write about topics on her personal life in blogs, while issuing search queries mostly related to her business.

In that case, adjusting search results based on what we get from the user’s blogs could make it more difficult for her to find the desired business information. When the personalized search system detects that a user’s interest profile is affecting the search results more than it should, the system may decrease the personalization degree for the user.

To make adjustment after answering a query, the first step is to evaluate the effectiveness of personalization with the current personalization degree. Let S_u be the set of search results that are actually clicked by the user u . Also, let L_g be the original list of results returned by the underlying search engine and L_p be the final list of results returned by our personalized search system. We compute two matching scores, one between L_g and S_u , and the other between L_p and S_u . Our system checks how many results in the top x spots of L_g (or L_p) are included in S_u , and it gives more weight to those elements ranked higher in the list. More specifically, the system uses the Normalized Discounted Cumulative Gain (NDCG) as the matching score between list L and set S :

$$NDCG(L, S) = Z_x \sum_{i=1}^x (2^{r_i} - 1) / \log_2(i + 1)$$

where r_i is 1 if the i th element of L is in S and r_i is 0 otherwise. Z_x is chosen so that a perfect ranking has NDCG value of 1.

Let p_u^0 be the current value of the personalization degree of user u . We compute the new value p_u^1 of personalization degree as follows.

$$\begin{aligned} p_u^1 &= p_u^0 - p_u^0 \cdot \alpha \cdot \beta && \text{when } m(L_p, S_u) < m(L_g, S_u) \\ p_u^1 &= p_u^0 + (1 - p_u^0) \cdot \alpha \cdot \beta && \text{when } m(L_p, S_u) > m(L_g, S_u) \end{aligned}$$

where α is a real-number parameter controlling adjustment speed and $\beta = \frac{|NDCG(L_p, S_u) - NDCG(L_g, S_u)|}{NDCG(L_p, S_u) + NDCG(L_g, S_u)}$. Intuitively, the larger the difference between $NDCG(L_p, S_u)$ and $NDCG(L_g, S_u)$, the faster the personalization degree is adjusted (due to larger β).

Adjusting source weights. Different people may use online social systems differently. In Section 2.2, we pointed out that interest vectors from various information sources are combined with different weights. Furthermore, we have presented our method on estimating the initial weights of different information sources based on the total amount of information provided by a source. However, the initial values may not be accurate, and moreover, new data from those information sources may be added to the user’s profile over time. To make our system more adaptive, we design a scheme to automatically adjust the weights of different information sources in a user’s profile.

After answering a search query from user u , let S_u be the set of search results that are actually clicked by u . We adjust the weight of an information source according to how well its interest vector matches the webpages in S_u (recall that a user’s interest profile maintains an interest vector for each information source). Let v_i be the interest vector of the i th information source, the similarity score between v_i and S_u is computed as $h(v_i, S_u) = \sum_{s \in S_u} \cos(v_i, s)$, where $\cos(v_i, s)$ is the cosine similarity between interest vector v_i and the word vector of webpage s .

We then compute the average value \bar{h} of the matching scores of the interests vectors in u ’s profile. We increase the weight w_i of the i th information source, if $h(v_i, S_u) > \bar{h}$; we decrease w_i , if $h(v_i, S_u) < \bar{h}$. Let w_i^0 be the current weight of the i th information source in u ’s profile. We compute the new value w_i^1 as follows.

$$\begin{aligned} w_i^1 &= w_i^0 - w_i^0 \cdot \alpha' \cdot \beta' && \text{when } h(v_i, S_u) < \bar{h} \\ w_i^1 &= w_i^0 + (1 - w_i^0) \cdot \alpha' \cdot \beta' && \text{when } h(v_i, S_u) > \bar{h} \end{aligned}$$

where α' is a real-number parameter controlling adjustment speed

and $\beta' = |h(v_i, S_u) - \bar{h}|/\bar{h}$. Intuitively, the further the $h(v_i, S_u)$ deviates (either being larger or smaller) from average value, the faster the weight w_i is adjusted (due to larger β').

3. EVALUATION

We have implemented our personalized search approach and performed experiments on real-world data to evaluate its performance. There are four objectives in our experiments.

1. We would like to check whether our personalized search solution improves search quality over its underlying non-personalized search engine.
2. We would like to see how the amount of available data on social systems affects the performance of our solution. In other words, whether our solution performs better on users who are highly active on social systems (and thus has a richer interest profile) than those who are less active.
3. We would like to evaluate if using the information from multiple online social systems is better than relying on the information from a single source with regards to search personalization.
4. We would like to determine whether the adaptive schemes improve the performance of our solution.

In the rest of this section, we describe the experiment setup and present our findings.

3.1 Experimental Data

We tested our personalized search approach using data from three types of online social activities. They are blogs, social bookmarks, and mutual tags (also called people tags). Our experimental data was collected from a large-scale social computing system that contains a blogging sub-system called MyBlogs, a social bookmarking system called MyFaves, and a user profiling system called TagU¹. Our data set from MyFaves consists of more than 420,000 bookmarks contributed by over 6,000 users from worldwide. The profiling system TagU supports mutual tagging which allows users to tag each other. In our TagU data set, more than 53,000 users have been tagged with over 170,000 tags.

Not every user is involved in all the three types of online social activities we considered. In our experiments, we chose those users who have written at least 10 blogs, received no less than 10 people tags, and bookmarked 20 webpages or more. There are 208 users in our data sets that meet our selection criteria. Our selection contains users who are very active in online social systems as well as those who are less active. On average, each of the selected user has 26 people tags, 129 blogs, and 391 bookmarks. The maximum numbers of people tags, blogs, and bookmarks among the selected users are 137, 2635, and 5050, respectively.

3.2 Evaluation Method and Metrics

Common methods to evaluate personalized search approaches include user interviews and search log examination. While users' opinions are crucial to personalization solutions, conducting study with a large number of people is oftentimes expensive and time-consuming; not to mention that those 208 users meeting our selection criteria are from all over the world. Examining search logs to compare personalized search results with non-personalized results is an effective approach. However, in practice, acquiring

¹MyBlogs, MyFaves, and TagU are not the real names of the systems we collected data from; we anonymized their names for blind review purpose.

search logs are difficult for most researchers, because search engine providers do not normally release such data. Furthermore, we would need user identity information so as to match the search logs with the user interest profiles we build from online social systems. Getting search logs with identity information is almost impossible due to privacy concerns. Therefore, we have to resort to other evaluation methods.

The major challenge to evaluate a personalized search system is to determine which results are considered useful and relevant to a search query by a specific user. Here, we employ the evaluation method proposed in [17], which uses people's bookmarks and tags to represent their opinions. The intuition is that, if a user u bookmarked a webpage and tagged the webpage with a word t , then u must consider the webpage useful and relevant to t . We may issue a query consisting of the keyword t on behalf of u , and then check whether those webpages tagged with t by u are ranked high in the result list returned by the personalized search system. A drawback of this evaluation method is the potentially high false negative rate: a webpage not being tagged with t may still be considered useful by u . In spite of the drawback, this method still allows us to get a pretty good idea on the performance of a personalized search system. Furthermore, the method has a couple of advantages that make it a good choice for us. First, this evaluation method only refers to users' social bookmarks and tags, which are easy to get (as opposed to search logs). Second, since user participation is not required in the evaluation process, we can run as many experiments as needed to achieve the evaluation objectives given at the beginning of this section. To answer the questions we have, we will need to test the personalized search system over a large number of queries in various settings.

We tested the 208 chosen users one by one. For each user, 25% of her bookmarks from MyFaves together with her blogs and people tags were used to create her interest profile; the other 75% of the user's bookmarked webpages in MyFaves (which do not overlap with the 25% bookmarks used in profile building), together all the webpages bookmarked by other MyFaves users, were used as the testing corpus. For the i th user u_i , we randomly selected 30 words u_i has applied as tags on her bookmarks. For each of those 30 words, a search query consisting of the word was issued on behalf of u_i . Assume that a search query consists of a word t . Let $L_t[1, k]$ be the list of top k results returned by the search system and S_t be the set of webpages that have been tagged with t by u_i . We compute the *recall@k* as $|L_t[1, k] \cap S_t|/|S_t|$. To measure the performance of the search system with regards to u_i , we compute the average value of the recall over the 30 search queries issued for u_i . The *improvement percentage* of search approach A over approach B for u_i is computed as $(\bar{r}_a - \bar{r}_b)/\bar{r}_b$, where \bar{r}_a and \bar{r}_b are the average recall of approaches A and B , respectively. To evaluate the overall performance of the search system, we compute the average recall value over the tested users. The average improvement percentage against a baseline search solution over tested users may also be computed. Finally, when comparing two search solutions, we say that a solution *wins* for a user, if it has a higher average recall for the user's queries than the other solution.

3.3 Experimental Results

In this subsection, we present our experimental results. For convenience, we refer to our personalized search approach as Personalized and the underlying non-personalized one as Normal.

Personalization v.s. Non-personalization. The first set of experiments are designed to check whether Personalized improves

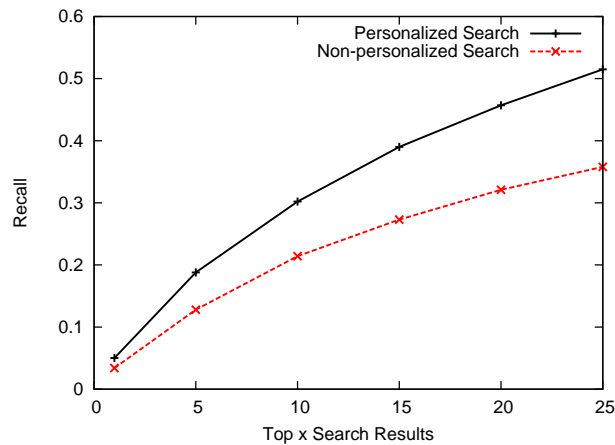


Figure 1: Comparison between personalized search and non-personalized search.

the search quality over Normal. Experimental results are presented in Table 1 and visualized in Figure 1.

As we can see from Figure 1, Personalized outperformed Normal with regards to recall in all cases. On average, Personalized’s recall was 26% to 61% higher than that of Normal, depending on how many results we consider for each query. In particular, when we consider more than 5 returned results, Personalized’s improvement percentage over Normal was larger than 50%.

In addition to comparing recall, we would also like to know if Personalized outperforms Normal for most users. The numbers of wins for Personalized and Normal are given in Table 1. When we consider at least the top 10 results for each query, Personalized outperformed Normal on 117 to 134 (or 56% to 64%) users, while Normal outperformed Personalized on around 40 (or 19%) users; for about 20% of the tested users, Personalized and Normal had roughly the same performance. This indicates that, for most users, it is beneficial to personalize their search results.

In general, the experimental results demonstrated that our personalized search approach is effective in identifying individual users’ preferences and adjusting search results accordingly.

Active users v.s. Less active users. The goal of the second set of experiments is to evaluate the performance of Personalized for users with different amounts of data on social systems. For each of the 208 selected users, we computed the sum of the numbers of her social activities (i.e. the number of blogs, the number of bookmarks, and the number of people tags). We chose the 35 users with the largest sums of activities and named the group High. In other words, High contained those users who are most active on social systems. Also, we selected the 35 users with the smallest sums and named the group Low. Finally, we chose 35 users with sums around the average value over the 208 users and named the group Medium. The average sums of social activities of users in High, Medium, and Low are 1888, 608, and 135, respectively.

Experimental results are presented in Table 2 and visualized in Figure 2. As we can see from Table 2, Personalized had higher average recall than Normal in all the three groups of users. The percentage of improvement on recall ranged from 23% to 60%. Furthermore, in all the three groups, Personalized won much more users than Normal did. In general, our experimental results show that Personalized is effective both for users who are highly active and those who are less active.

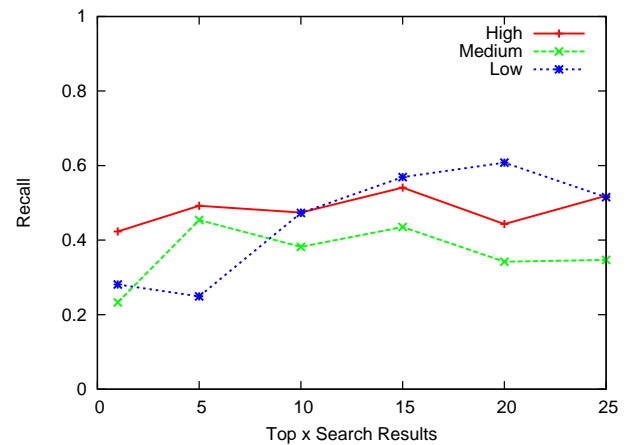


Figure 2: Comparing the effectiveness of personalized search among groups of users with different amount of social activity data. The figure compares their average improvement percentage on recall over the baseline non-personalized search approach.

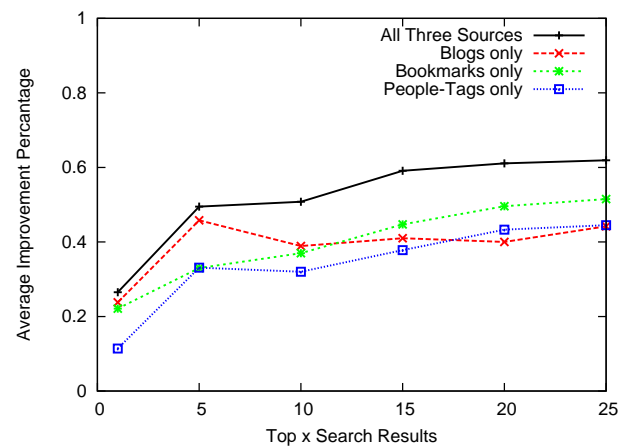


Figure 3: Comparing the approach that uses multiple information sources with those using a single source. The figure compares their average improvement percentage on recall over the baseline non-personalized search approach.

Next, we compare the performance of Personalized among different groups. From Figure 2, Personalized performed better for High than for Normal. Personalized also performed better for High than for Low when we consider only the first or top 5 search results for each query. However, when we consider at least the top 10 results, Personalized had similar performance for High and Low. We cannot draw a conclusion that Personalized performed better for users who are more active on social systems than those who are less active. But the decent performance on Low does show that Personalized is very effective even for users who do not have much available data in online social systems.

Multiple sources v.s. Single source. The objective of the third set of experiments is to study whether using information from multiple social systems is superior to relying on information from a single source with regards to search personalization. As we know, Personalized gathers information from three social systems. We would like to see if referring to more information sources indeed leads to better understanding of a user’s backgrounds and

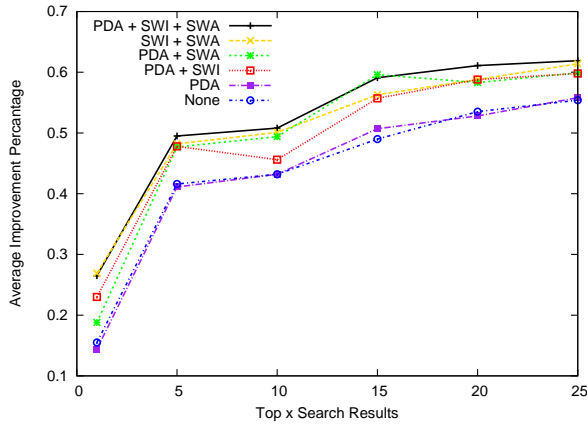


Figure 4: Compare the effectiveness of adaptive learning mechanisms. The figure compares their improvement percentage on recall over the baseline non-personalized search approach. In the figure, PDA stands for Personalization Degree Adjustment, SWI stands for Source Weight Initialization, and SWA stands for Source Weight Adjustment. “None” indicates no adaptive scheme is enabled, while “PDA, SWI, and SWA” indicates all three schemes are enabled.

preferences. In our experiments, we compare *Personalized* with its special versions, each of which uses information from only one social system. For convenience, we call the special versions of *Personalized* that use only blogs, only social bookmarks, and only people tags, *Blogs*, *Bookmarks*, and *People-Tags*, respectively. In the experiments, we compared the four personalized search approaches over the baseline approach *Normal*. Experimental results are presented in Table 3 and Figure 3. In the table and the figure, the improvement percentage of a personalized search approach is its improvement percentage on recall over *Normal*.

First, we compare the three single-source personalized search approaches, *Blogs*, *Bookmarks*, and *People-Tags*, with *Normal*. As we can see from Figure 3, all the three personalized approaches outperformed *Normal*, as all of them had positive improvement percentage over *Normal*. This indicates that our personalized scheme is effective even if only one social information source is available. Furthermore, the three single-source approaches had similar performance, which shows that the amounts of user information extracted from the three information sources may be close to each other.

Second, we compare *Personalized* with its three special versions. From Figure 3, *Personalized* performed better than *People-Tags* and *Bookmarks* in all cases. The recall of *Blogs* was similar to that of *Personalized* when we consider only the first or top 5 results per query. But *Personalized* considerably outperformed *Blogs* when we take into account more than the top 5 results. Our results demonstrate that utilizing information from multiple sources does allow us to gather more comprehensive information about a user.

Effectiveness of adaption. The objective of the last set of experiments is to determine whether the adaptive schemes improve the performance of *Personalized*. *Personalized* employs three adaptive schemes: personalization degree adjustment, estimation on the initial weights of information sources, and adjusting the weights of those sources after answering search queries. We would like to know if the three schemes have positive impact on the perfor-

mance of *Personalized*. We compare *Personalized* with its special versions, in which one or more of the adaptive schemes are disabled. For convenience, we refer to the special version that disables all the three adaptive schemes as *None*. Similar to the last set of experiments, we use *Normal* as a baseline approach for comparison purpose. Experimental results are presented in Table 4 and Figure 4. For simplicity, we use PDA, SWI, and SWA to denote the three adaptive schemes, personalization degree adjustment, source weight initialization, and source weight adjustment, respectively.

First, the three adaptive schemes, when used together, have positive impact on the performance on *Personalized*. As we can see from Figure 4, PDA + SWI + SWA outperformed *None*. The improvement percentage of PDA + SWI + SWA over *Normal* is 7 to 10% more than that of *None* over *Normal*. This shows that adaptability is advantageous in general.

Next, we study whether all the three adaptive schemes are effective. We found that adjusting personalization degree alone is not very effective, as PDA had roughly the same performance as *None*. Performance enhancement is observed when PDA is paired up with SWI or SWA. Among the three adaptive schemes, source weight adjustment appears to be most effective, as both SWI + SWA and PDA + SWA had very similar performance with PDA + SWI + SWA. The performance of our personalized search approach gets worse when SWA is disabled. From Figure 4, PDA + SWI had slightly lower improvement percentages over *Normal* when compared with PDA + SWI + SWA.

4. DISCUSSION AND FUTURE WORK

In this section, we discuss topics related to our personalized search solution and future work.

Deployment feasibility in practice. To deploy our personalized search scheme in practice, a natural question is whether search engine providers indeed have easy access to information on users’ online social activities. Our answer is positive. Right now, a number of major search engine providers are offering various social networking services. For example, Yahoo! provides blogging service on user profiles and has also acquired the popular social bookmarking site Del.icio.us. For another example, Microsoft is running Windows Live Spaces, which is a popular blogging and social networking platform. Similarly, Google offers Blogger and other social networking services. For such search engine providers, acquiring information on users’ online social activities is not a problem, as they can always crawl information from their own services. Furthermore, several major social networking websites have started to offer web search as auxiliary services. For instance, when a user issues a search query on Facebook, relevant web results will be displayed in addition to Facebook-specific content such as user profiles. These social networking sites can easily benefit from our ideas. Finally, nowadays, a lot of information is publicly accessible on major social networking systems, such as Twitter, LinkedIn, and Del.icio.us. Many people are increasingly aware of the importance of gaining visibility on the web. For example, having a profile on LinkedIn with detailed professional information may lead to career opportunities from time to time. It is increasingly practical for search and other service providers to acquire information on users’ social online activities from the Internet.

In addition to Internet, our personalized search approach may be used in enterprise environments as well. Enterprise social computing systems, such as IBM’s Lotus Connections, enable companies to create social networks among their employees. Enterprise users may connect with their colleagues and share their ideas through

blogs, bookmarks, tags, and other activities. Popular systems, such as Lotus, are being used by thousands of companies worldwide. Personalization solutions taking advantages of enterprise social networks could have broad impacts in the business world.

Generalizing our solution for other applications. Beside web search, personalization is also useful in applications such as advertisement targeting, news filtering, tag recommendation, and so on. Learning about users' interests is critical to any personalization application. We have described a way to adaptively infer user interests from public information on multiple social sources. The interest information derived by our approach is not restricted to web search and can thus be easily consumed by all kinds of personalization applications. The only step in our solution that is application-dependent is the way user feedback is acquired. For web search, our approach observes the click-through results. This can be generalized as observing whether an interest-based recommendation is taken or not. Our adaptive components require no or minor change when used on other applications.

Refined learning. Our personalized search system employs a number of adaptive schemes. Currently, the adjustment of personalization parameters, such as personalization degree and information source weights, does not take topics into account. In practice, certain information sources may be particularly good at recommending content on certain topics, while being bad at others. For example, a user may often blog her traveling experiences but rarely write about technical topics. In this case, information from the user's blogs may be good at personalizing queries related to traveling but weak at suggesting technical contents. For search queries related to traveling, such as "cruise" and "Hawaii", the personalized search system may assign more weight to the interest vector from the user's blogs. In contrast, for search queries on technical stuffs, such as "database" and "python", the system may give less weight to blogs and rely more on other information sources. To achieve such fine-grained adjustments, in a user's interest profile, each information source will have a number of weights, one on each topic. And the weights on different topics will be adjusted independently. Our next step is to study whether such a fine-grained adaptive scheme can boost the performance of our personalized search system even further.

5. RELATED WORK

Personalized search has attracted significant amount of interests in the research community. Existing solutions differ in various aspects, such as the information sources for user interests and how to perform personalization. At the high level, there are two ways to acquire information about users' interests. The first approach is to ask users to explicitly state their preferences [8, 4], while the second approach is to let the personalized search system to infer users' preferences from one or more data sources. Our personalized search solution falls into the second category.

Researchers have explored all kinds of sources to acquire the context information needed for search personalization. Among those data sources that have been explored, search and browsing history are probably the most popular one in existing solutions [12, 1, 13, 11, 9]. Qiu and Cho [11] proposed a personalized search framework based on users' past search history. They designed a user model to formalize users' interests in webpages and correlate them with users' clicks-on search results. In [1], Cao et al. applied Hidden Markov Model on users' search logs to learn about the contexts of users' search queries. Their approach can be scaled to process very large volumes of search log data. In [9], Luxenburger et al. introduced a statistical language model that may represent user

tasks in different granularity levels. Their scheme allows dynamically switch of the course of personalization and may detect when a user's search and browse history is not appropriate for aiding search quests.

Besides search and browsing history, emails and desktop documents are other popular data sources for context information. In [3], Chirita et al. proposed to infer user interests by exploiting a user's personal information repository, which include personal collection of text documents, emails, and cached webpages. Their approach achieved search personalization through query expansion, i.e. modify the original search query to include information on query issuer's preferences. Our solution employs a different strategy by re-ranking search results rather than revising search queries. Also, since emails and desktop documents may contain sensitive information, users may have privacy concerns on those data. Xu et al. [18] have proposed a privacy-enhanced solution, which allows a user to specify which parts of the information retrieved from her personal data may be shared with a search engine for personalization purposes. Such a selective sharing scheme grants users control over their interest profiles.

Researchers have also studied building rich models of interests from a combination of information sources. Teenvan et al. [14] explored the idea of creating user interest models from a variety of information, including search history, browsing history, and emails. They mentioned using machine learning to select the best parameters as their future work. In [16], White et al. performed a systematic study on the effectiveness of variant sources of contextual information for user interest modeling. They investigated context information, such as user's browsing history and the combined interests of other users that also visit the current webpage. However, online social activity is not among the context sources they examined. Their study showed that different sources may perform differently at different time and context overlap outperforms any isolated source. Their findings support the use of adaptive schemes and imply the benefits of integrating information from multiple sources.

None of the above work exploits information from online social systems to perform personalized search. With the growing popularity of social networks, more and more valuable information can be retrieved from users' online social activities. Researchers have recently started to harvest data from social systems to infer users' interests. Xu et al. [17] proposed to learn about users' interests from their bookmarks and tags on social bookmarking system. They also proposed a method to evaluate personalized search approaches using social bookmark information. We adopted their evaluation method to test our personalized search solution. Similarly, Noll and Meinel [10] presented a personalized search approach that is based on user information from social bookmarks and tagging. Both [17] and [10] considered only a single source of social information, i.e. social bookmarks.

Carmel et al. [2] investigated personalized social search based on the user's social relations. In their approach, search results are re-ranked according to their relations with individuals in the user's social network. Preference is given to search results that are related to individuals who are closely connected to the query issuer in the social network. While the solution in [2] also utilizes information from multiple social systems, they mainly explore the relationship among users. It may be difficult to generalize the relation-based personalization scheme in [2] to web search on the Internet, where most webpages do not have clear relation with the people in a user's social network. In contrast, our personalized search solution focus on deriving topics of interest, rather than relationship, from users' social activities. Our topic-based personalized search scheme can be easily applied to internet web search, as we do not assume re-

lations among users and search items; we just check whether an item contains topics that are interesting to a target user. Finally, our personalized solution is adaptive and is able to automatically adjust personalization degree and the weights of different information sources. Such adaptability is not provided in the approach in [2].

Finally, researchers have systematically studied the effectiveness of personalized search in real life. Dou et al. [5] examined whether personalization is consistently effective on different queries for different users. They found that personalized search has significant improvement over common web search on some queries, but it has little effect on other queries. Teevan et al. [15] went one step further; they characterized queries using a variety of features and, using those features, they built predictive models to identify queries that can benefit from personalization. Determining when to personalize is an interesting future direction. We will explore how to integrate that into our solution.

6. CONCLUSION

In this article, we have proposed a personalization framework that infers users' preferences from their activities on a variety of online social systems. We have described how to create user interest profiles, how to integrate information from different information sources, and as an example how to personalize search results. Furthermore, our system is adaptive; it observes users' choices on search results and automatically adjusts parameters such as personalization degree and the weights of different information sources so as to better meet the user's needs. We have implemented our approach and performed experiments on real-world data collected from three large-scale social systems. Over two hundred users from worldwide have been tested. Our experimental results demonstrated the effectiveness of our personalized search approach. Our results also showed that integrating information from multiple social systems usually leads to better personalized search results than relying on the information from a single social system, and our adaptive approach further improves the performance of the search system. Finally, our techniques on adaptively mining user interests from public information on social systems may be applied to other applications in addition to web search.

7. REFERENCES

- [1] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li. Towards context-aware search by learning a very large variable length hidden markov model from search logs. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 191–200, New York, NY, USA, 2009. ACM.
- [2] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'el, I. Ronen, E. Uziel, S. Yogev, and S. Chernov. Personalized social search based on the user's social network. In *CIKM '09: Proceeding of the 18th ACM conference on Information and knowledge management*, pages 1227–1236, New York, NY, USA, 2009. ACM.
- [3] P. A. Chirita, C. S. Firan, and W. Nejdl. Personalized query expansion for the web. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 7–14, New York, NY, USA, 2007. ACM.
- [4] P. A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschütter. Using odp metadata to personalize search. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 178–185, New York, NY, USA, 2005. ACM.
- [5] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 581–590, New York, NY, USA, 2007.
- [6] S. Farrell and T. Lau. Fringe contacts: People-tagging for the enterprise. In *WWW 2006: Collaborative Web Tagging Workshop*, Edinburgh, Scotland, 2006.
- [7] S. Farrell, T. Lau, S. Nusser, E. Wilcox, and M. Muller. Socially augmenting employee profiles with people-tagging. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pages 91–100, New York, NY, USA, 2007. ACM.
- [8] P. Ferragina and A. Gulli. A personalized search engine based on web-snippet hierarchical clustering. In *WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 801–810, New York, NY, USA, 2005. ACM.
- [9] J. Luxemburger, S. Elbassuoni, and G. Weikum. Matching task profiles and user needs in personalized web search. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, pages 689–698, New York, NY, USA, 2008. ACM.
- [10] M. Noll and C. Meinel. Web search personalization via social bookmarking and tagging. *The Semantic Web*, pages 367–380, 2008.
- [11] F. Qiu and J. Cho. Automatic identification of user interest for personalized search. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 727–736, New York, NY, USA, 2006. ACM.
- [12] M. Speretta and S. Gauch. Personalized search based on user search histories. In *WI '05: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 622–628, Washington, DC, USA, 2005. IEEE Computer Society.
- [13] J.-T. Sun, H.-J. Zeng, H. Liu, and Y. Lu. Cubesvd: A novel approach to personalized web search. In *In Proc. of the 14th International World Wide Web Conference (WWW)*, pages 382–390. Press, 2005.
- [14] J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 449–456, NY, USA, 2005.
- [15] J. Teevan, S. T. Dumais, and D. J. Liebling. To personalize or not to personalize: modeling queries with variation in user intent. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 163–170, New York, NY, USA, 2008. ACM.
- [16] R. W. White, P. Bailey, and L. Chen. Predicting user interests from contextual information. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 363–370, New York, NY, USA, 2009. ACM.
- [17] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu. Exploring folksonomy for personalized search. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 155–162, New York, NY, USA, 2008. ACM.
- [18] Y. Xu, K. Wang, B. Zhang, and Z. Chen. Privacy-enhancing personalized web search. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 591–600, New York, NY, USA, 2007. ACM.

Table 1: Comparison between personalized search and non-personalized search

Top x Search Results		1	5	10	15	20	25
Recall	Non-Personalized	0.034	0.128	0.214	0.273	0.321	0.358
	Personalized	0.050	0.188	0.302	0.390	0.457	0.515
Improvement Percentage	Non-Personalized	0.0	0.0	0.0	0.0	0.0	0.0
	Personalized	26.5	49.5	50.8	59.1	61.1	61.9
Number of Wins	Non-Personalized	22	33	39	35	40	42
	Personalized	67	101	118	117	132	134

Table 2: Comparison among users with different amount of social activity data

Top x Search Results		1	5	10	15	20	25
High	Non-Personalized Recall	0.035	0.129	0.235	0.303	0.374	0.424
	Personalized Recall	0.057	0.198	0.363	0.507	0.611	0.717
	Improvement Percentage	42.3	49.2	47.4	54.1	44.3	51.9
	Non-Personalized Wins	5	11	8	5	8	7
	Personalized Wins	12	17	19	23	25	27
Medium	Non-Personalized Recall	0.050	0.196	0.322	0.416	0.488	0.551
	Personalized Recall	0.067	0.263	0.422	0.545	0.629	0.701
	Improvement Percentage	23.3	45.4	38.2	43.5	34.2	34.7
	Non-Personalized Wins	2	4	7	5	4	5
	Personalized Wins	11	20	25	21	24	24
Low	Non-Personalized Recall	0.029	0.097	0.151	0.174	0.201	0.217
	Personalized Recall	0.041	0.134	0.189	0.224	0.255	0.274
	Improvement Percentage	28.1	24.9	47.3	56.9	60.8	51.5
	Non-Personalized Wins	2	2	4	3	6	7
	Personalized Wins	11	12	19	16	19	17

Table 3: Comparison between multiple sources and single source

Top x Search Results		1	5	10	15	20	25
Recall	All Three	0.050	0.188	0.302	0.390	0.457	0.515
	People-Tags Only	0.041	0.163	0.265	0.341	0.402	0.450
	Bookmarks Only	0.047	0.167	0.274	0.355	0.425	0.475
	Blogs Only	0.048	0.176	0.280	0.364	0.425	0.484
Improvement Percentage	All Three	26.5	49.5	50.8	59.1	61.1	61.9
	People-Tags Only	11.4	33.1	32.0	37.8	43.3	44.5
	Bookmarks Only	22.1	33.0	37.0	44.7	49.6	51.5
	Blogs Only	23.8	45.8	38.9	41.0	40.0	44.2

Table 4: The effectiveness of adaptive learning mechanisms. In the table, PDA stands for Personalization Degree Adjustment, SWI stands for Source Weight Initialization, and SWA stands for Source Weight Adjustment.

Top x Search Results		1	5	10	15	20	25
Recall	PDA + SWI + SWA	0.050	0.188	0.302	0.390	0.457	0.515
	SWI + SWA	0.050	0.188	0.301	0.388	0.455	0.516
	PDA + SWI	0.048	0.188	0.298	0.385	0.452	0.510
	PDA + SWA	0.046	0.184	0.298	0.383	0.447	0.502
	PDA	0.044	0.175	0.283	0.364	0.430	0.481
	None	0.044	0.177	0.284	0.366	0.433	0.481
Improvement Percentage	PDA + SWI + SWA	26.5	49.5	50.8	59.1	61.1	61.9
	SWI + SWA	26.9	48.2	50.1	56.3	58.8	61.4
	PDA + SWI	23.0	47.8	45.6	55.7	58.8	59.8
	PDA + SWA	18.8	47.7	49.4	59.6	58.3	59.9
	PDA	14.4	41.1	43.2	50.7	52.8	55.8
	None	15.5	41.6	43.2	49.0	53.5	55.4