Personalized Social Search Based on the User's Social Network

David Carmel, Naama Zwerdling,
Ido Guy, Shila Ofek-Koifman, Nadav Har'el,
Inbal Ronen, Erel Uziel, Sivan Yogev
IBM Research Lab in Haifa
Haifa, 31905, Israel
{carmel,naamaz,ido,shila,nyh,
inbal,erelu,sivany}@il.ibm.com

Sergey Chernov*
L3S Research Center, Leibniz Universität
Appelstr. 4, 30167 Hannover, Germany
chernov@L3S.de

ABSTRACT

This work investigates personalized social search based on the user's social relations – search results are re-ranked according to their relations with individuals in the user's social network. We study the effectiveness of several social network types for personalization: (1) *Familiarity-based* network of people related to the user through explicit familiarity connection; (2) *Similarity-based* network of people "similar" to the user as reflected by their social activity; (3) *Overall* network that provides both relationship types. For comparison we also experiment with *Topic-based* personalization that is based on the user's related terms, aggregated from several social applications.

We evaluate the contribution of the different personalization strategies by an off-line study and by a user survey within our organization. In the off-line study we apply bookmark-based evaluation, suggested recently, that exploits data gathered from a social bookmarking system to evaluate personalized retrieval. In the on-line study we analyze the feedback of 240 employees exposed to the alternative personalization approaches. Our main results show that both in the off-line study and in the user survey social network based personalization significantly outperforms non-personalized social search. Additionally, as reflected by the user survey, all three SN-based strategies significantly outperform the Topic-based strategy.

Categories and Subject Descriptors

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

General Terms

Algorithms, Experimentation

*Part of this work was done while Sergey Chernov interned at IBM Haifa Research Lab.

CIKM'09, November 2–6, 2009, Hong Kong, China.

Keywords

Personalization, Social search, Social networks

1. INTRODUCTION

1.1 Personalized search

Personalizing the search process, by considering the searcher's personal attributes and preferences while evaluating a query, is a great challenge that has been extensively studied in the information retrieval (IR) community but still remains a stimulating task [6]. It is of great interest since user queries are in general very short and provide an incomplete specification of individual users' information needs. For example, searching for "IR" by an information retrieval student has a completely different meaning than searching by another who is interested in infra-red radiation.

Search personalization requires the capability of modeling the users' preferences and interests. This is usually done by tracking and aggregating users' interaction with the system. In general, such aggregation includes users' previous queries [31], click-through analysis [20, 12], and even eye-tracking during the search session [20]. Users' interactions are structured into a *user profile* that can be utilized during search [2]. A user profile is usually employed in two main scenarios, either through *personalized query expansion*, i.e., adding new terms to the query and re-weighting the original query terms based on the user profile [9], or through *re-ranking* and *filtering* the search results while incorporating users' interests accordingly [30].

However the aggregation of user interactions comprises some difficulties. First, many users consider user profiling as an activity which may violate their privacy. Users may feel uncomfortable with a system that accumulates their interactions and can potentially exploit that data for malicious actions such as spamming, phishing, or exposing it to the general public. Privacy issues are the main reason for new regulations enforced by many countries that put constraints on systems' sufferance to aggregate users' activity [23].

Second, previous user interactions do not always provide a good indication of current needs. This is especially true for new users for whom only limited personal information exists, or when user preferences evolve over time. Moreover, the benefits that can be achieved through personalization vary accross queries [33]. For some queries, different people may expect the same results, whereas for others different results are expected by individuals even for identical queries. Finally, personalized search results make justifying the relevance of a specific result for a given query more difficult, as they are biased by query-independent personalized considerations. Some users may be confused when receiving different results for the same query due to the fact that their profile evolved during successive submissions.

1.2 Personalized social search

There are several alternative definitions of the concept *social search* [18, 7, 3]. In this work we use the notion of social search to describe the search process over "social" data gathered from Web 2.0 applications, such as social bookmarking systems, wikis, blogs, forums, social network sites (SNSs), and many others [7, 3]. Such a social search system represents different entity types (documents, persons, communities, tags) and their interrelations, and allows searching for all object types related to the user's query.

Social search provides an ideal testbed for personalization due to the explicit user interactions through Web 2.0 tools. A user profile that is derived from user feedback such as bookmarking, rating, commenting, and blogging, provides a very good indication of the user's interests. Furthermore, user profiles that are only based on explicit public social activity can be safely utilized without disrespecting the user's privacy[1]. Consequently, several previous works studied search personalization by profiling user interests based on public bookmarks aggregated from a social bookmarking system [22, 27, 35, 8].

In addition, when the user's social network (SN) is available, the preferences of the user's related people can be utilized to assist in obtaining the user's preferences, assuming closely related people have similar interests. This is the main assumption behind collaborative filtering methods for recommendation systems, when user interests and preferences are predicted based on the preferences of "similar" persons. User similarity relationships are typically inferred through user feedback in the form of item rating. However, more recent approaches leverage implicit interest indicators [11], such as tags, views, or comments, as well as direct familiarity relationships [21], e.g., as reflected through connections in SNSs. We note that we refer to social networks in their broad definition, i.e., networks of people. Connected edges may represent any type of relationship, not only explicit familiarity [15].

1.3 SN-based personalized social search

In this work we study personalized social search in the enterprise based on the social relations of the searcher. We focus on re-ranking of search results by considering their relationships to users that belong to the searcher's social network. The assumption behind this personalization approach is that the preferences of other people, who are expected to have "similar" interests as the searcher, provide a good prediction for the searcher's preferences and can thus assist in revealing the search results that might subjectively satisfy the searcher's needs.

Personalized re-ranking of search results is done as follows: given a list of (non-personalized) results retrieved for the user's query, and a list of related users extracted from his/her social network, search results are re-ranked by considering their relationship strength with those users. Thus, documents that are strongly related to the user's related people are boosted accordingly.

To retrieve the user's social network, and the user-document relationship matrix, we use *SaND* [29], an enterprise social search system used in our organization. For each user, *SaND* provides related people extracted through the user's SN. This is a ranked list of people, who relate to the user either through explicit familiarity connections (e.g., co-authorship of a wiki page or a connection within an SNS), or by some kind of similarity as reflected by their social activity (e.g., usage of the same tags or commenting on the same blog entry). People ranking is determined according to a weighting scheme that takes into account the overall related activity between two users [15]. In addition, *SaND* provides for each user all related documents (e.g., web-pages, blog entries), each associated with relationship strength to the user. A user may relate to a document through authorship, tagging or commenting, or by being mentioned in the page's content. The relative strength of each relationship type is determined by an appropriate weight [15].

We experiment with SN-based personalization considering three social network types: (1) *Familiarity-based* network, (2) *Similarity-based* network, and (3) *Overall* network that implies both relationship types. In addition to the user's SN, we consider the relevance of the search results to the user's topics of interest. These topics are approximated by a set of terms that are related to the user, including tags used by the user to bookmark documents, and tags used by others to bookmark that user [13]. We assume that these related terms represent the user's interests, thus can be used to personalize the search results accordingly. We note that this assumption only holds for active taggers, or for users that were heavily tagged by others. Personalization is achieved by promoting search results that were tagged with these user's terms, either by the user or by others. We call this approach *Topic-based* personalization. As mentioned above, this approach has been extensively studied by previous works on personalized social search which construct users' profiles based on the tags they used for bookmarking [22, 27, 35]. We use it as a comparative baseline for an SN-based personalization approach for social search.

The different personalization approaches are evaluated by an off-line study and by a user survey within our organization. In the off-line study we follow the work of Xu et al. [35], and Carman et al. [8], which evaluates search personalization as follows: given a user u who bookmarked a document d with the tag t, we assume that if u will search for t he will consider d relevant for t. Thus, any triplet (u, d, t) given by a social bookmarking system can be used as a personalized query for evaluation. The higher the rank of documents tagged by u with t, while simulating u searching for t, the better the personalization method is. The main drawback of this approach is that documents that were not tagged by u are considered irrelevant – a weak assumption that is not necessarily true. However, predicting u's tagging behavior indicates the system's personalization capability. We discuss the advantages and limitations of this evaluation approach in more detail in Section 4.

In the on-line study, we analyze a survey of 240 employees exposed to the different personalization approaches studied in this work. Our main results are: (1) Personalized social

[1]For full transparency, Web 2.0 tools should better clarify to their users that any public social data provided by them can potentially be utilized by other social software applications.

search based on the user's SN significantly outperforms non-personalized search. A maximal improvement was achieved by the Overall social network which integrates familiarity and similarity relations. (2) As reflected by the user survey, all three SN-based strategies significantly outperform the Topic-based strategy, which improves only slightly over non-personalized results. (3) The integration of related terms with related people in the user profile slightly improves the search results. (4) The off-line evaluation is consistent with the user survey in confirming the superiority of SN-based personalization strategies, and the contribution of additional related terms to the SN-based user profile. However, several discrepancies between the two evaluation methods raise concerns about its reliability in ranking alternative personalization approaches.

The rest of the paper is organized as follows. Section 2 discusses related work on search personalization in general and in particular for social search. Section 3 discusses the different personalization approaches we study in this work. Section 4 describes the off-line experiment and the on-line survey, and the results we obtained. Section 5 concludes and raises several direction for future work.

2. RELATED WORK

2.1 Personalized search

In recent years many researchers utilize query log and click-through analysis for web search personalization. In [28], the authors combine a topic-sensitive version of *PageRank* [16] with the history of user clicks data for search result personalization. Joachims et al. [20] study clicks applicability as implicit relevance judgments. They show that users' clicks provide a reasonably accurate evidence of their preferences. Tan et al. [31] propose a language modeling approach for query history mining. Their small-scale study demonstrates significant improvement of personalized web search with a history-based language model over regular search. The user modeling approach described in [30] is based on a decision-theoretic framework to convert implicit feedback into a user profile that is used to re-rank search results. Agichtein et al. [2] introduce an alternative user modeling method, in which a set of rules is applied to a query log. While user models are usually targeted at search personalization, they could also be applied for personalized information filtering, as was shown in [37] who analyze click history for the identification of regular users' interests. Recent work of Teevan et al. [34] on "groupization" shows that combining implicit user profiles from several related users has a positive impact on personalization effectiveness.

In addition to regular web log data, several works consider personalization using desktop data and external resources. For example, in [32] the authors index desktop information and experiment with different representations of users, documents and queries for personalized web search. Chirita et al. [9] explore personalized query expansion based on users' desktop information. Several approaches for personalized Web search are based on global interests using the *Open Directory Project (ODP)* categories [24, 10, 35]. In [24] the authors map previously visited pages to *ODP* categories and use this mapping to build a user profile. Another work [10] proposes a personalized version of *PageRank*, in which a hand-picked set of preferred users' categories are applied for result re-ranking.

Recently, new approaches for adaptive personalization focus on the user task and the current activity context . There are several approaches trying to predict applicability of personalization while considering the current context of the user's task on query submission [33, 12, 25].

2.2 Social search

The amount of social data is rapidly growing and has become a main focus of research on social search. Recent work [17] reports that in 2008 around 115 million bookmarks were available on the *del.icio.us* social bookmarking site. A page popularity measure, *SBRank*, proposed in [36], is proportional to a number of existing social bookmarks. Following the language modeling approach, a theoretically sound generative model for social annotations is presented in [38].

Tags and other conceptual structures emerging in social systems are called *folksonomies* and are typically modeled as graphs. A formal model for *folksonomies* and ranking algorithms called *Adapted PageRank* and *FolkRank* are defined in [18]. *FolkRank* is used for the generation of personalized rankings of entities within the *folksonomy* and for the recommendation of tags, users and resources. Lately, Bao et al. [5] propose two alternative algorithms, *SocialSimRank* and *SocialPageRank*. Both are based on social annotations and corresponding connections between pages, annotations and users. A comparative evaluation study of these algorithms and a few novel algorithms are described in [1].

2.3 Personalized social search

Several approaches for directly or indirectly employing users' social relations for personalization exist. A re-ranking method presented in [27] is based on users' tag profiles which are derived from his/her bookmarks in *del.icio.us*. The tags of each search result on the site are matched against the user's profile. The problem of automatic user profile generation is addressed in [4]. The authors investigate how accurate user profiles can be generated from *del.icio.us* data.

Another approach described by Bender at al. [7] directly exploits social relations by combining semantic and social factors in the ranking. The users, tags and documents are represented as nodes in a "friendship graph", in which edges are extracted from relationships like links, content, tagging and rating. Ranking is based on *UserRank*, an algorithm derived from the PageRank computation on the friendship graph. A document receives an extra "friendship" score when tagged by a user's "friend". Similarly to this approach, our work personalizes the score of a document for a specific user if it has been bookmarked, authored or commented by people related to the user, or tagged with terms related to her/him. We further analyze the value of personalization according to different relationship types, in particular familiarity and similarity.

Xu et al. [35] recently developed another personalization appraoch which uses social relations indirectly. Their *Topic Adjusting* algorithm is built on top of *ODP* data and *folksonomies* such as *del.icio.us* and *Dogear* [26]. Users' interests are inferred using the topics of their tagged pages. The relationship weights in the user-page matrix are defined based on the number of user annotations assigned to a page. This work has some similarity with our approach, however, our personalization method explicitly uses familiarity and similarity scores to model direct and indirect relations between users. For evaluating the *Topic Adjusting* algorithm, Xu et al. introduce a new method for automatic evaluation of personalized search, in which the user's tags are used as queries and all documents bookmarked by this user with that tag are considered relevant. More details on automatic evaluation of personalized search based on social bookmarking data can be found in [8]. We adopt a similar evaluation approach for our work and complement it with a large-scale user study.

3. USER PROFILES

In this section we describe the social search platform used for our study, and the types of user profiles we experimented with.

3.1 System description

IBM Lotus Connections (LC)[2] is a social software application suite for organizations that was introduced in 2007. It contains (as of version 2.0) five social software applications: profiles – of all employees, a social bookmarking system, a blogging service, a communities service, and activities (not discussed in this work). In our study we experimented with LC tools as used in our organization. Dogear[26], LC's social bookmarking application, allows users to store and tag their favorite web pages. Over 90% of the bookmarks are public (visible to all other users) and about half are intranet pages, while the other half are external internet pages. Dogear includes 743,239 public bookmarks with 1,943,464 tags by 17,390 users. Blog Central [19], LC's blogging system, has 16,337 blogs, 144,263 blog entries, with 69,947 users. LC's communities service contains over 2,100 online communities, each with shared resources and discussions, with a total of over 50,000 members.

Social Networks and Discovery (*SaND*) [29], is an aggregation tool for information discovery and analysis over the social data gathered from all LC's applications. It leverages complex relationships between content, people and tags, and its integrated index supports a combination of content-based analysis and people-based analysis. *SaND* provides several social aggregation services including social search, personalized item recommendations, personalized people recommendations, finding social paths between people, and additional social network services. *SaND* provides social search over the social data using a unified approach [3] in which all system entities (documents, persons, groups, tags) are searchable and retrievable. As part of its analysis, *SaND* builds an entity-entity relationship matrix that maps a given entity to all related entities, weighted according to their relationship strength. The entity-entity relationship strength is composed of two types of relations:

- **Direct relations**: Figure 1 shows all direct relations between entities that are modeled by *SaND*. For example, a user is directly related to: (1) a document: as an author, a tagger, or a commenter; (2) another person: as a tagger of, or tagged by that person, as a friend as stated in several SNSs that exist in the enterprise, or through the enterprise's organizational chart (direct manager/employee); (3) a tag: when used by the user for bookmarking, or when used by others to tag that user; (4) a group: as a member or an owner. Other direct relations and their corresponding relative weights are shown in the figure.
- **Indirect relations**: Two entities are indirectly related if both are directly related to the same entity[3]. For example, two users are indirectly related if both are related to another user, e.g. if both have the same manager, or if both tagged the same document.

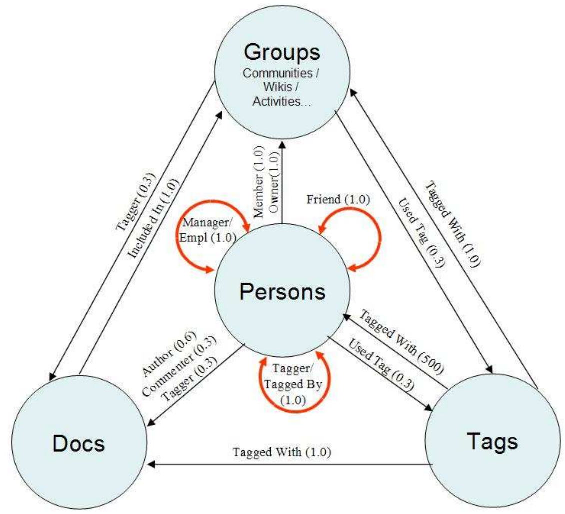

Figure 1: Direct relations between entities modeled by *SaND*. The relative relationship strengths appear on the graph's edges. Familiarity relations are colored red (bolded).

The overall relationship strength between two entities is determined by a linear combination of their direct and indirect relationship strengths. More details on score calculation and implementation issues are described in previous work on social network aggregation and social search [15, 3].

3.2 User profile types

We experimented with three types of social networks for personalization. Each network maps a user to a list of related users weighted according to their relationship strength.

Familiarity SN

Familiarity between two individuals is considered according to indicators that they know each other [15]. A direct familiarity relation exist if both persons are marked as friends in one of the enterprise SNSs, or when one is the direct-manger/employee of the other. In addition a person is familiar with those s/he tagged, but not vice versa. Indirect familiarity relations are defined when the two persons are both authors of the same paper, patent, or wiki-page, or when both have a common manager (team members).

In order to extract the user's Familiarity network, we use *SaND* to extract all the user's related people and to filter out all non-familiar people which do not obey the above constraints. In addition, the relationship strength between the two is modified to be based on familiar relations only. More details on the familiarity relationships and the calculation of the familiarity score can be found in [14].

[2]http://www-01.ibm.com/software/lotus/products/connections/

[3]Currently, only indirect relations of level two are considered, i.e., two-length path in the entity graph. However, the model can easily be extended to support indirect relations of any level .

Similarity SN

Similarity between two individuals is considered according to common activity in the context of LC's social software: co-usage of the same tag; co-tagging of the same document; co-membership of the same community, or co-commenting on the same blog entry. Similarly to the familiarity case, in order to extract the user's Similarity network, we use *SaND* to extract all related people and retrieve (and re-weight) only people which obey the above constraints.

Overall SN

Besides the Familiarity and Similarity networks, we also examine the user's Overall social network, which contain all related persons according to the full relationship model.

Topic-based

The user's topics of interests are represented by a set of terms that are closely related to the user. Directly related terms are tags used by the user to tag documents and other people, and tags used by others to tag that user. Indirectly related terms are those that are related to the user through other entities (e.g. all tags of a document bookmarked by the user). The user's top related terms retrieved by *SaND* serve as the user's Topic-based profile.

3.3 Personalizing the search

A user profile is constructed on the fly when a person logs into the system. For a user u, *SaND* retrieves $N(u)$ – the ranked list of users related to u, and $T(u)$ – the ranked list of related terms. These two lists are then used as the user profile to personalize the search results for all user's queries during the search session.

Given the user profile, $P(u) = (N(u), T(u))$, the search results are re-ranked as follows:

$$S_p(q, e|P(u)) = \alpha S_{np}(q, e) + (1-\alpha)[\beta \sum_{v \in N(u)} w(u, v) \cdot w(v, e) + (1-\beta) \sum_{t \in T(u)} w(u, t) \cdot w(t, e)] \quad (1)$$

$S_p(q, e|P(u))$ is the personalized score of entity e to query q given the profile of user u. $S_{np}(q, e)$ is the non-personalized *SaND* score of e to q. Since we only re-weight the search results, only entities with positive score are considered. $w(u, v)$ and $w(u, t)$ are the relationship strength of user v and term t to u, as given by the user profile. Similarly, $w(v, e)$ and $w(t, e)$ are the relationship strength between v and t to entity e, as given by *SaND*.

Thus, an entity is first scored by *SaND* according to its non-personalized scoring mechanism, and then the entity score is modified according to its relationship strength with users and terms in the user profile.

The equation has several parameters that control the amount of personalization. First $N(u)$ is determined according to the SN type used for personalization (Familiarity, Similarity, Overall). Second, the number of users and terms in the profile are configurable. Third, the parameter α controls the relative weight of the personalization score compared to the original non-personalized score, and β controls the relative weight between people and terms for personalization. In the next section we describe several experiments we conducted with some of these controllable parameters.

4. EVALUATION

In this section we describe the experimental methodology used to evaluate the SN-based personalization approach, the results of an off-line study using a bookmark-based evaluation, and a user survey we conducted in our organization.

4.1 Evaluation methodologies for personalized social search

Evaluating personalized search is a great challenge since relevance judgments can only be assessed by the searchers themselves – only the users can subjectively judge whether a specific result answers their personal need. Therefore, existing IR evaluation benchmarks based on judged queries, each associated with a set of relevant results objectively assessed by experts, cannot be utilized for personalized search evaluation.

Existing evaluation approaches for personalized search are often based on a user study, where participants are asked to judge the search results for their personal queries in a personal manner, thus alternative personalization techniques can be comparatively analyzed. However, appropriate user studies with a reasonable number of participants are very expensive to accomplish, therefore, such studies are uncommon and often limited to a small and a biased sample. Alternatively, users' implicit feedback such as clicking on a specific result (while un-clicking other results), can be interpreted as personal relevance judgment. Clicks, however, are not necessarily the best indicators for user satisfaction with results - clicking on a result does not necessarily mean it is relevant, while un-clicking does not always imply irrelevance. Furthermore, such evaluation is only feasible for a live system with enough users who use it on a regular basis.

Social search applications provide richer sources for user feedback that can be used for regular personalized search evaluation. User feedback such as rating, tagging, and commenting, indicates the user's interest in a specific document. Recently, several works utilized data from *Delicious* to evaluate personalized search methods [35, 8]. In this approach, any bookmark (u, d, t) which represents a user u who bookmarked a document d by a tag t, can be used as a test query for personalized search evaluation. The main assumption behind is that any document tagged by u with t (including d) is considered relevant for the personalized query (u, t) (i.e. u submits the query t).

Therefore, the bookmark triplets (u, d, t) extracted from a social bookmarking system provide an almost unlimited source of personalized test queries to be used for personalized search evaluation. Given the bookmark (u, d, t), a personalized search system is evaluated according to its ability to highly rank the corresponding documents. A good personalization policy is expected to differentiate between two similar tested queries (u_1, d_1, t) and (u_2, d_2, t), promoting d_1 while serving (u_1, t), and d_2 for the query (u_2, t).

There is a delicate issue with bookmark-based evaluation. The search system is already "aware of" the association between d and t, as realized by u, hence this information can be exploited for over tuning. For example, given the query (u, d, t), a personalization approach that retrieves only the documents tagged by u with t will inevitably outperforms other personalization alternatives, since any other document is considered irrelevant. However, this "over-tunned" personalization policy is restricted to queries that were previously used as tags by the user, hence it will totally fail for other personalized queries. This limitation cannot be disclosed by the bookmark-based evaluation methodology.

In order to eliminate the dependency between personalization and evaluation, and to simulate the personal query (u, d, t) with no prior knowledge on the user's association between t and d, we have to mask u bookmarking of d. Masking is done as follows: for each personal query (u, d, t), we first "hide" that bookmark from the search system before handling the query (u, t). The system is instructed as this specific bookmarking has never happened – d content is not enriched by the tag t (unless d was tagged with t by others), t is taken out from the user profile (unless t relations with u is derived from other resources) and u's relations with other entities that are based on this bookmark are modified accordingly. This masking guarantees that personalization is evaluated without any prior knowledge on u relations with d and t.

Note that personalized methods that better predict their users' interests, as reflected by their tagging activity, will be favored by that evaluation methodology. This is definitely one of the main characteristics that are expected from a personalized search system, hence such evaluation can successfully prioritize alternative personalization strategies. However, the bookmark-based evaluation approach still suffers form the incompleteness problem – not all documents tagged by u with t are relevant for u while searching for t, and not all documents not tagged by u with t are necessarily irrelevant. This limitation is partially handled by the huge amount of personalized queries available for evaluation. But we believe that conclusions based on such evaluation should be supported by alternative evaluation methods - an approach that was taken by us in this work. We first evaluate and tune our personalized social search system with the bookmark-based evaluation, using *Dogear's* bookmarks as personalized queries, and confirm our findings with an extensive user survey based on 240 participants that subjectively judge the results for their 577 personal queries. To the best of our knowledge, this is the first study that (1) eliminates the dependency between personalization and evaluation that inherently exists in bookmark-based evaluation; (2) validates the bookmark-based evaluation methodology for personalized search by comparing its findings with the results of an independent user survey.

4.2 Experimental setup

We experimented with several personalization methods that are based on the the user's social network, and on the set of the user's topics. A user profile, $P(u) = (N(u), T(u))$, is based on $N(u)$, a ranked list of the user's related people, as given by the user's SN, and $T(u)$, a ranked list of the user's related terms. The user profile is constructed on the fly while the user logs into system and is used to personalize (re-rank) the search results by Equation 1 throughout the search session[4].

4.3 Off-line study

In the off-line study we used Dogear's bookmarks as personal queries. For each personalization method, we randomly selected 2000 bookmarks, and for each bookmark (u, d, t) we masked its existence from the search index and the user profile, to completely hide the relations between u, d, and t. Then, t was submitted as a query to *SaND* and 1000 results (documents) were retrieved. Other retrieved entities such as persons and tags were ignored, as they are not suitable for evaluation by the off-line approach. The search results were re-ranked using u's profile, and were evaluated by measuring average-precision (AP) and reciprocal rank (RR), while considering all documents tagged by u with t as relevant answers. After completion, the hidden bookmark was returned to the collection before processing the next tested bookmark.

Note that due to the masking process, d will be retrieved for t only when t appears in the original content of d , or when d was associated with t by others. The personalization methods differ in the way they re-rank d. SN-based personalization methods will advance d when it is related to at least one person in u's social network. Topic-based personalization will boost d if tagged by at least one of the terms related to u.

4.3.1 *Off-line study - main results*

Table 1 shows the mean-AP (MAP) and mean-RR (MRR) results for the configurations we experimented with, setting $\alpha = \beta = 0.5$. The top rows show the results of SN-based personalization with top-5 related people and with no related terms. The bottom rows show the results with top-5 people and top-5 terms.

User Profile		MAP	MRR
Non-Personalized		0.156	0.187
No Terms	Familiarity-SN	0.389	0.444
	Similarity-SN	**0.423**	**0.476**
	Overall-SN	0.388	0.442
With Terms	Topic-based	0.426	0.475
	Familiarity-SN	0.412	0.461
	Similarity-SN	**0.452**	**0.510**
	Overall-SN	0.410	0.461

Table 1: Bookmark-based evaluation of personalized social search. User profile is based on the top-5 related people and top-5 related terms.

There are several interesting insights from these results. First, all personalization methods significantly outperform non-personalized search (one-tailed unpaired t-test, $p < 0.001$). The MAP of the Similarity-based strategy is almost three times higher than that of non-personalized search.

Second, the Similarity-SN significantly outperforms the Familiarity and the Overall networks, and maybe surprisingly, the Overall-SN is slightly inferior (almost identical) to the Familiarity-SN. This indicates that similarity relations better predict the user' preferences than familiarity relations. We do not have good explanation to the inferiority of the Overall network, especially when this result is in contrast to the results of the user survey discussed in the following. We hypothesize that better integration of the similarity and familiarity relations by *SaND* might result in better performance of the Overall network.

Third, Topic-based personalization with no SN data improves the search significantly, and outperforms the Familiarity and the Overall SN. Integrating the user's related terms with the related people improves the search performance of all network types. The best result achieved while

[4]In the off-line study, since a bookmark is hidden prior to handling the personal query, we re-construct the user profile after bookmark masking and before query submission, to guarantee that the user profile has no dependency on the tested bookmark.

integrating the top-5 similar people with the top-5 related terms. In the following we further experiment with that integration task.

4.3.2 User profile size

The size of the user profile is determined by the lists $N(u)$ and $T(u)$. These lists boost the search results through their relationship strengths with those related people and terms. There is a risk that adding too many people or terms to the user profile may personalize too much, disregarding new relevant items that have not been discovered yet by the user's community. Therefore, finding an "optimal" user profile size is an important factor that significantly affects personalization effectiveness.

The size of $N(u)$ is controlled by two parameters, max_N, which sets the maximum number of (top scored) related people in the profile, and θ_N which determines a threshold on the relationship score. This threshold guarantees that only closely related people will be part of the user profile. Therefore, "socially active" users will have max_N related people in their profile, while others may have much less. Similarly, max_T and θ_T determine the number of terms in the user profile.

We experimented with max_N and max_T, while fixing the θ values to 0.0 (i.e., each user has max_N people and max_T terms in the profile, unless *SaND* retrieves fewer related people or terms for that user). Figure 2 shows the MAP for the different SN types, averaged over 2000 personal queries, while fixing the number of related people to 5 and varying the number of related terms. Similarly, in Figure 3 we fix the number of terms to 5 and vary the number of people in the profile.

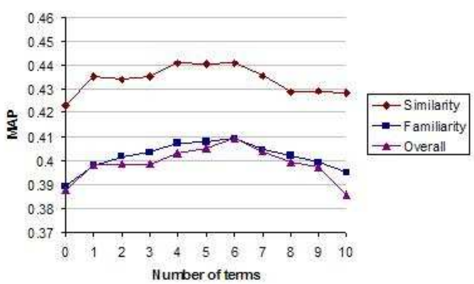

Figure 2: MAP for the different SN types, averaged over 2000 personal queries, for different number of related terms while fixing the number of related people to 5.

According to Figure 2, the maximum performance is achieved while adding 4-6 related terms to the the user profile, improving the MAP by 4-5% for all network types. Adding too many terms degrades the performance, even lower than with no terms at all, probably due to overstated personalization.

Figure 3 shows the performance of adding related people to a user profile with 5 related terms. We can see improvement only while adding similar people to the profile. Maximum improvement is achieved with 3 similar people, then the performance is dropped for additional people. In contrast, familiar people constantly harm the search performance while added to the profile. These results suggest that according to the bookmark-based evaluation, an "optimal"

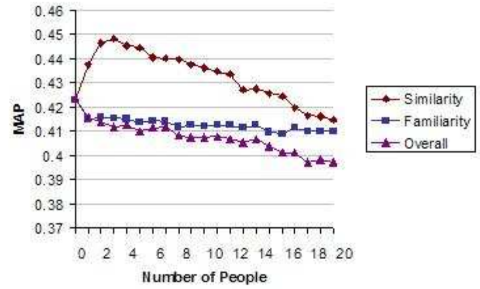

Figure 3: MAP for the different SN types, averaged over 2000 personal queries, for different number of related people while fixing the number of related terms to 5.

user profile should be based on a few similar people and a few related terms. However, we note that these results fit an "average user" while it is quite clear that user profiles should be subjectively adapted according to the user's personal characteristics. For future work we intend to experimenting with the θ values, adapting the optimal profile size for each user in a personal manner.

The off-line bookmark-based evaluation can be easily applied in very large scales, without any user intervention, hence it can be efficiently used for tuning the system parameters, and to efficiently examine alternative personalization strategies. However, due to the limitations of this approach, conclusions based on that evaluation should be validated by applying complementary evaluation methods.

4.4 User survey

To complement and validate the results of the off-line evaluation, we ran a user survey in our organization, asking participants to assess the search results for their queries in a personal manner. Each participant was asked to assess two personal queries and was given the opportunity to evaluate more queries, as much as s/he likes. In order to simulate personal queries, for which the user has personal information needs, we recommended the participant a set of tags s/he was tagged with in our organization's people tagging application [13], to be submitted as personal queries, assuming that such tags represent interesting topics, or at least familiar to the user. The participant was asked to select two terms out of the recommended tags, or alternatively submit their own (personal) queries for assessment. After completion, participants were encouraged to comment on the search experience with the system.

We considered users who had at least 30 people in both their Similarity and Familiarity networks, and at least 30 related terms. We note that this sample does not represent the entire population of employees, but rather active users of the LC system, who are the target population for our search system. We sent a link to the survey with a request for participation to a random sample of 645 of these users and got a response from 240, who judged 577 personal queries (91% of the queries were personal terms suggested by us while the rest were original queries selected by the participants). Our survey participants originated from 28 countries, spanning over the globe and over all our organizational divisions.

Each participant that took the survey was first classified randomly into one of eight classes, each associated with a different personalization strategy; the eight strategies we experimented with are shown in Table 2. Participants were not aware of the personalization type selected for them. For each participant who took the survey, the user profile was set according to the strategy of the class s/he was associated with, and for each user query, the search results retrieved by *SaND*, which were re-ranked according to the corresponding user profile, included 10 top relevant pages and 10 top relevant people, each judged by the user as non-relevant, relevant, or highly relevant. Figure 4 shows the entrance page users obtained while taking the survey, including the terms suggested as personal queries, and the results to be judged after the query was issued. Looking at the figure, please note that most terms suggested as personal queries for this specific user are ambiguous and can be interpreted in several ways. For example, the subjective meaning of "pasta" for this user is probably a code-name of a research project and not a noodle type.

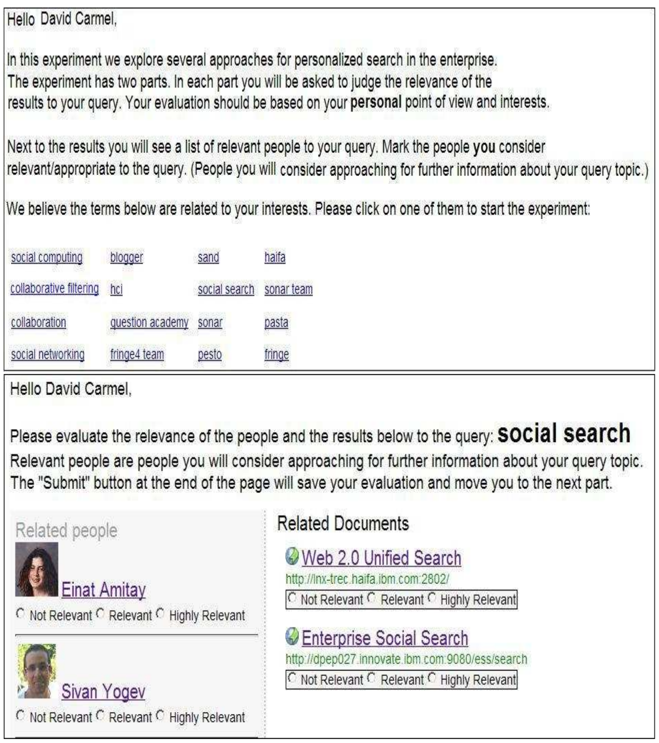

Figure 4: User survey pages. Top: the entrance page with instructions and the personal terms suggested for querying. Bottom: a snapshot of the results page.

Most of the comments we got were very positive. One participant wrote: *" [...] thanks for the opportunity to try out the research project! The results were quite interesting, I found content on the topic I didn't know to exist [...]"*. Another one wrote: *"[...] I am eager to see the evaluation metrics from these experiments. Such an outcome motivates to pay a lot more attention to social factors in all personalized applications [...]"*.

The quality of search results was measured by the normalized discount cumulative gain (NDCG) and by precision at 10 (P@10), averaged over the set of judged queries, for each of the classes. For DCG calculation we used gains (0,1,2) for the three relevance levels respectively, and the discount function used was $-log(rank + 1)$. Normalization (NDCG) was done by dividing the DCG value with an ideal DCG value calculated as all results are highly relevant. For P@10 calculation, we considered any positive judgment as relevant.

Table 2 shows the precision of the search results, as measured by NDCG@10 and P@10, for the eight personalization strategies. The general high satisfaction from the social search system is reflected by the high NDCG@10 (> 0.5) and P@10 (> 0.6) achieved in all classes.

User Profile		Judged Queries	NDCG @10	Delta (%)	P@10
Non-Personalized		79	0.511	–	0.61
No Terms	Familiarity	71	0.560	9.7	0.68
	Similarity	78	0.550	7.6	0.68
	Overall	69	**0.597**	**16.9**	**0.73**
With Terms	Topic-based	81	0.518	1.4	0.64
	Familiarity	68	0.561	9.9	0.69
	Similarity	69	0.565	10.7	0.71
	Overall	62	**0.581**	**13.8**	**0.72**

Table 2: User survey: The precision of the search results of the personalized search strategies, measured by NDCG@10 and P@10. The Delta column shows the improvement in NDCG@10 over non-personalized search.

The main outcomes of the survey are: (1) As in the off-line study, all personalization methods outperform the non-personalization strategy. These differences are found to be significant for all strategies except for the Topic-based one (one-tailed unpaired t-test, $p < 0.05$). (2) A maximal improvement was achieved by the Overall network, 16.9% improvement in NDCG@10 without terms and 13.8% with terms. (3) The Similarity network outperforms the Familiarity network with and without terms, and both significantly outperform the Topic-based strategy. (4) Related terms slightly improve search effectiveness when applied alone (1.4%), and when added to the Similarity and the Familiarity SN, in agreement with the off-line study; however, they decrease the performance of the Overall network. This result indicates that optimal integration between SN and personal terms should be further studied for each of the networks separately, as currently system parameters are commonly set to all user profile types.

There are several substantial differences between the two evaluation methods. Both methods confirm the significant contribution of personalization for social search, and the superiority of using similar people over familiar people in the user profile. However, the Overall network, the "shining star" of the user survey, performs the worst according to the off-line study. Similarly, the topic-based strategy, with marginal contribution in the survey, perform very well in the off-line study. In Section 5 we discuss possible reasons for these discrepancies and whether the conclusions derived from the bookmark-based evaluation have any value at all.

4.4.1 Personalized people search

Table 3 shows the distribution of the relevant people retrieved by the different SN based strategies (accumulating all positive judgments as relevant). On average, people retrieved by the Familiarity network were judged as more relevant for the user queries compared to to other networks and to the non-personalized search. We can clearly see an increase in the percent of relevant people while moving from non-personalized, to Overall, Similarity, Familiarity, respectively. However, this result is likely to be affected by the natural bias of users to people they are familiar with.

User Profile	Relevant people (%)
Non-Personalized	47.8
Familiarity	**55.8**
Similarity	52.2
Overall	51.8

Table 3: The relevance distribution of retrieved people for the different SN types.

Several participants mentioned the difficulty in judging the relevance of people to their query, mostly because of unfamiliarity. Someone wrote "*"…It would be good to include more information on the people that are shown on the results, like their Job Role/Title. This would help to identify on a first look their relevance or not."*.

Actually, participants had the opportunity to open the home-page created by *SaND* for each retrieved person, viewing his role, communities, list of publications, blogs, and more. However, it seems that judging unfamiliar people's relevancy is more difficult than judging unfamiliar documents' relevancy. Indeed, 21% of the retrieved people were not judged by the participants, relative to 9% only of retrieved pages. We therefore plan for future work to examine personalization of people search in more detail, e.g., by inspecting user behavior on a live social search engine.

5. SUMMARY AND DISCUSSION

In this work we investigated personalized social search based on the user's social relations. We studied the effectiveness of several social network types for personalization, and evaluated their contribution by an off-line study and by a user survey within our organization. Our results showed that according to both evaluations, social network based personalization significantly outperforms non-personalized social search. In addition, as reflected in our user survey, all three SN-based strategies significantly outperform the Topic-based strategy, which improves only slightly over non-personalized results.

The bookmark-based evaluation for search personalization has the advantage that it can be easily applied in very large scales, without any user intervention. To validate its outcomes we compared the results we got from the off-line study with those of the user survey. Our results show that there are several substantial discrepancies between the two evaluation methods. In particular, according to the off-line study, the Overall network is inferior to the Similarity and Familiarity networks, and to the Topic-based strategy, while in contrast, according to the user survey, the Overall network performs the best.

These disagreements are not unexpected – there are several differences between the two evaluation approaches. In the off-line study, participants were randomly selected from all Dogear users, while in the on-line study we focused on heavy users of the LC system. In addition, off-line queries were based on the user's tags while on-line queries were based on the tags the users were tagged with. Furthermore, the bookmark-based evaluation method predicts the user's bookmarking activity while the on-line survey measures directly the users' personalized relevance judgments. As a result, the off-line approach discriminates against authored or commented documents, and biases tagged documents, while this discrimination does not exist in the user survey.

The extreme success of the Similarity network in the off-line study, in contrast to its comparable performance with other networks in the user-survey, can be explained by the fact that social activity of similar people better predicts the user's social activity than the activity of familiar people. This also interprets the difference in performance of the topic-based strategy, which performs reasonably well in the off-line study while exhibiting inferior effectiveness compared to SNs in the user survey. It seems that similar related people and related terms are strongly associated with the tested bookmark's document, therefore, in the off-line run, this document is advanced even after bookmark masking. In contrast, according to the survey results, interesting/relevant documents are associated with similar and familiar related people much more than with related terms.

The disagreements between the bookmark-based evaluation and the user survey put a question-mark on its reliability for personalized search evaluation, especially for ranking different personalization approaches. Considering the survey results as ground truth, some of the "conclusions" derived from the off-line evaluation were proved to be wrong. However, we believe that it might have some benefits, mostly for parameter tuning while fixing the personalization strategy. For example, finding a good combination of related people and terms in the user profile, or searching for appropriate α and β for Equation 1. In any way, it seems that conclusions based on bookmark-based evaluation should better be confirmed by an independent evaluation method.

In this work we mostly focused on document retrieval, while abandoning other retrievable entities such as people, tags, and groups. Our user-survey evaluated people search quality, and indeed showed the superiority of Familiarity network over other networks for personalized people search. However, this results should be confirmed as many retrieved people were not judged due to participants' unfamiliarity (20%). In addition, we assume that familiar people were favored by participants in their judgments over non-familiar ones. Therefore, reducing that bias is needed in order to objectively evaluate personalized people search. However, we believe that people search will benefit from emphasizing familiar people for the searcher, as these people are the most reasonable sources of additional information, as expected from people search results.

As previous work showed not all queries should be personalized [33]. We hypothesize that this is also true for personalized social search. In this work we simulated personal queries with tags used for bookmarking by the user, in the off-line study, and with tags the user was tagged with, in the user survey. In both cases these types of personal queries are limited and do not cover the whole spectrum of possible personal queries, but rather a subset that is likely to benefit from personalization and which can be judged by the methods in use. For future work we would like to expand our study to other types of personal queries, to better understand what types of queries should be personalized, and for the long run, to enhance our personalized social search system to be able to decide on-line, per each user and query pair, the most suitable search policy.

Our personalization approach is extremely simple and is based on re-ranking the search results based on their relationship strength with the user's related people and topics. The high effectiveness of this approach for social search implies that the social relations used for personalization, as derived from the user's social network, are highly reliable in predicting user interests and preferences. This claim holds for enterprise social data, as shown in this work. The high quality social data that is available today for individuals in the enterprise, allows the identification of social relations that can be utilized for search personalization and for other applications. The question whether social data out of the firewall, typically with lower quality, can be used effectively for search personalization remains open for further research.

6. REFERENCES

[1] F. Abel, N. Henze, and D. Krause. Ranking in folksonomy systems: can context help? In *Proceedings of CIKM*, pages 1429–1430. ACM, 2008.

[2] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *Proceedings of SIGIR*, pages 3–10. ACM Press, 2006.

[3] E. Amitay, D. Carmel, N. Har'El, S. Ofek-Koifman, A. Soffer, S. Yogev, and N. Golbandi. Social search and discovery using a unified approach. In *Proceedings of HyperText*, pages 199–208. ACM, 2009.

[4] C. M. Au-Yeung, N. Gibbins, and N. Shadbolt. A study of user profile generation from folksonomies. In *Proceedings of the WWW Workshop on Social Web and Knowledge Management*, April 2008.

[5] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *Proceedings of WWW*, pages 501–510. ACM, 2007.

[6] N. J. Belkin. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42(1):47–54, 2008.

[7] M. Bender, T. Crecelius, M. Kacimi, S. Michel, T. Neumann, J. X. Parreira, R. Schenkel, and G. Weikum. Exploiting social relations for query expansion and result ranking. In *Proceedings of ICDE Workshops*, pages 501–506. IEEE, 2008.

[8] M. J. Carman, M. Baillie, and F. Crestani. Tag data and personalized information retrieval. In *Proceedings of the CIKM workshop on Search in social media*, pages 27–34. ACM, 2008.

[9] P. A. Chirita, C. S. Firan, and W. Nejdl. Personalized query expansion for the web. In *Proceedings of SIGIR*, pages 7–14. ACM, 2007.

[10] P. A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschütter. Using ODP metadata to personalize search. In *Proceedings of SIGIR*, pages 178–185. ACM, 2005.

[11] M. Claypool, P. Le, M. Waseda, and D. Brown. Implicit interest indicators. In *Proceedings of IUI*, pages 33–40. ACM Press, 2001.

[12] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of WWW*, pages 581–590. ACM, 2007.

[13] S. Farrell and T. Lau. Fringe contacts: People-tagging for the enterprise. In *Proceedings of the WWW Workshop on Collaborative Web Tagging*, 2006.

[14] I. Guy, M. Jacovi, N. Meshulam, I. Ronen, and E. Shahar. Public vs. private: comparing public social network information with email. In *Proceedings of CSCW*, pages 393–402. ACM, 2008.

[15] I. Guy, M. Jacovi, E. Shahar, N. Meshulam, V. Soroka, and S. Farrell. Harvesting with SONAR: the value of aggregating social network information. In *Proceedings of CHI*, pages 1017–1026. ACM, 2008.

[16] T. H. Haveliwala. Topic-sensitive pagerank. In *Proceedings of WWW*, pages 517–526. ACM, 2002.

[17] P. Heymann, G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? In *Proceedings of WSDM*, pages 195–206. ACM, 2008.

[18] A. Hotho, R. Ja"schke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In Y. Sure and J. Domingue, editors, *The Semantic Web: Research and Applications*, volume 4011 of *LNAI*, pages 411–426, Heidelberg, June 2006. Springer.

[19] J. Huh, L. Jones, T. Erickson, W. A. Kellogg, R. K. E. Bellamy, and J. C. Thomas. Blogcentral: the role of internal blogs at work. In *Proceedings of CHI*, pages 2447–2452. ACM, 2007.

[20] T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of SIGIR*, pages 154–161, 2005.

[21] H. Kautz, B. Selman, and M. Shah. Referral web: combining social networks and collaborative filtering. *Communication of the ACM*, 40(3):63–65, 1997.

[22] H. R. Kim and P. K. Chan. Personalized search results with user interest hierarchies learnt from bookmarks. In *WEBKDD,*, volume 4198 of *LNCS*, pages 158–176, 2005.

[23] A. Kobsa. Privacy-enhanced personalization. *Communication of the ACM*, 50(8):24–33, 2007.

[24] F. Liu, C. Yu, and W. Meng. Personalized web search for improving retrieval effectiveness. *IEEE Transactions on Knowledge and Data Engineering*, 16(1):28–40, 2004.

[25] J. Luxenburger, S. Elbassuoni, and G. Weikum. Matching task profiles and user needs in personalized web search. In *Proceedings of CIKM*, pages 689–698. ACM, 2008.

[26] D. R. Millen, J. Feinberg, and B. Kerr. Dogear: Social bookmarking in the enterprise. In *Proceedings of CHI*, pages 111–120. ACM, 2006.

[27] M. G. Noll and C. Meinel. Web search personalization via social bookmarking and tagging. In K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, editors, *The Semantic Web*, volume 4825 of *LNCS*, pages 367–380. Springer, 2007.

[28] F. Qiu and J. Cho. Automatic identification of user interest for personalized search. In *Proceedings of SIGIR*, pages 727–736. ACM Press, 2006.

[29] I. Ronen, E. Shahar, S. Ur, E. Uziel, S. Yogev, N. Zwerdling, D. Carmel, I. Guy, N. Har'El, and S. Ofek-Koifman. Social Networks and Discovery in the Enterprise (SaND). In *Proceedings of SIGIR*, pages 836–836, 2009.

[30] X. Shen, B. Tan, and C. Zhai. Implicit user modeling for personalized search. In *Proceedings of CIKM*, pages 824–831. ACM, 2005.

[31] B. Tan, X. Shen, and C. Zhai. Mining long-term search history to improve search accuracy. In *Proceedings of KDD*, pages 718–723. ACM Press, 2006.

[32] J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *Proceedings of SIGIR*, pages 449–456. ACM Press, 2005.

[33] J. Teevan, S. T. Dumais, and D. J. Liebling. To personalize or not to personalize: modeling queries with variation in user intent. In *Proceedings of SIGIR*, pages 163–170. ACM, 2008.

[34] J. Teevan, M. R. Morris, and S. Bush. Discovering and using groups to improve personalized search. In *Proceedings of WSDM*, pages 15–24. ACM, 2009.

[35] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu. Exploring folksonomy for personalized search. In *Proceedings of SIGIR*, pages 155–162. ACM, 2008.

[36] Y. Yanbe, A. Jatowt, S. Nakamura, and K. Tanaka. Can social bookmarking enhance search in the web? In *Proceedings of JCDL*, pages 107–116. ACM, 2007.

[37] B. Yang and G. Jeh. Retroactive answering of search queries. In *Proceedings of WWW*, pages 457–466. ACM Press, 2006.

[38] D. Zhou, J. Bian, S. Zheng, H. Zha, and C. L. Giles. Exploring social annotations for information retrieval. In *Proceedings of WWW*, pages 715–724. ACM, 2008.