



## Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>  
Eprints ID : 15229

The contribution was presented at UMAP 2014 :  
<http://www.um.org/umap2014/umap2014/index.html>

**To cite this version** : Kacem, Ameni and Boughanem, Mohand and Faiz, Rim  
*Time-Sensitive User Profile for Optimizing Search Personalization*. (2014) In:  
22nd International Conference on User Modeling, Adaptation and  
Personalization (UMAP 2014), 7 July 2014 - 11 July 2014 (Aalborg, Denmark).

Any correspondence concerning this service should be sent to the repository administrator: [staff-oatao@listes-diff.inp-toulouse.fr](mailto:staff-oatao@listes-diff.inp-toulouse.fr)

# Time-Sensitive User Profile for Optimizing Search Personalization

Ameni Kacem<sup>1</sup>, Mohand Boughanem<sup>2</sup>, and Rim Faiz<sup>1</sup>

<sup>1</sup> LARODEC, IHEC, Carthage Presidency, 2016 Tunis, Tunisia  
ameni.kacem@gmail.com, rim.faiz@ihec.rnu.tn

<sup>2</sup> IRT SIG, 118 Route de Narbonne, 31062, Toulouse CEDEX 9, France  
bougha@irit.fr

**Abstract.** Thanks to social Web services, Web search engines have the opportunity to afford personalized search results that better fit the user’s information needs and interests. To achieve this goal, many personalized search approaches explore user’s social Web interactions to extract his preferences and interests, and use them to model his profile. In our approach, the user profile is implicitly represented as a vector of weighted terms which correspond to the user’s interests extracted from his online social activities. As the user interests may change over time, we propose to weight profiles terms not only according to the content of these activities but also by considering the freshness. More precisely, the weights are adjusted with a temporal feature. In order to evaluate our approach, we model the user profile according to data collected from Twitter. Then, we rerank initial search results accurately to the user profile. Moreover, we proved the significance of adding a temporal feature by comparing our method with baselines models that does not consider the user profile dynamics.

**Keywords:** Personalized search, User profile, Freshness, Interests Dynamics, Kernel function.

## 1 Introduction

Personalization in information retrieval consists of providing search results that fit the individual user’s information needs and match his/her interests instead of providing the same results to a query for all users [1]. To build the user profile, that is necessary to perform personalization, many works focused on leveraging user’ activities for inferring interests.

To represent the user profile, the most common way is the vector space model (VSM) where interests are represented as a vector or vectors of keywords. The weight of the keywords, in most of the prior works, is assigned using the TF-IDF model or its variants [2] [3]. Other approaches extract taxonomies from the Open Directory Project (ODP) hierarchy to represent the user profile [4]. All these approaches assign more importance to the frequent terms no matter the moment of use. They did not explicitly integrate the time when the words

appeared in the user activities, but time is often used to discern short-term and long-term user profiles. Short-term profile considers current session while the long-term profile is built according to several search sessions. In fact, discerning the short- and long-term interests requires the use of a time interval that may include several interests [5], or session’s boundaries mechanisms where a session is defined by a set of queries related to the same information need [6] [7].

However, the short-term profile integrates only interactions extracted from the current session but lays long-term interests aside. The long-term profile captures old interests without considering the actual user needs [8]. To overcome this shortcoming, we propose to consider both type of profile (short- and long-term) in the same framework and to assign a novel temporal weight to terms of the profile.

More precisely, we study how to leverage user’s activities for user modeling and evaluate how do temporal dynamics affect the quality of user models in the context of personalized search. We propose a personalized search framework where the user profile is implicitly constructed from the user social Web activities and represented as a vector of weighted terms which correspond to the user’s interests. We propose to weight the profiles terms according to both the freshness and the frequency in order to unify both the recent and persistent interests instead of using the delimitation of the session activities.

The remainder of this paper is organized as follows. In Section 2, we review related works focusing on personalized search systems and user profiling modeling and evolution. In Section 3, we propose a temporal-frequency user profile that adjusts the frequency notion by a temporal function. In particular, our model integrates a freshness feature in order to track changes implicitly without discerning explicitly the long-term and short-term user profiles. Section 4 describes the experimental methodology adapted to evaluate our proposed approach followed by the corresponding results and their discussion. The final Section presents a summary of our work and future dicrections.

## 2 Related Work

In this section, we first review some existing works on personalized search and social-based user profiling. Next, we examine and discuss the user profile evolution in terms of discerning the short-term and long-term user profiles.

### 2.1 Personalized Search

The user profiling strategies extract information about the user from different sources and model them into a profile using multiple representations. Current systems tend to collect information about the user by considering folksonomies as a primary source to define the user’s profile. Many information sources have been used to model the user’s profile such as users’ personal information like users’ manual input hobbies, search history [9], click-through records [10], Web browser history [11], folksonomies [2], annotations [12], Web communities [13] and groups [14].

User's interests are also sourced from the social Web. Users interact with each other by creating and sharing content and by expressing their interests on different social Websites [15]. Social-based users' profiles exploit those data in order to extract knowledge about the searcher's preferences and interests. Noll and Meinel [16] examined two types of profiles: the user's profile and the document's profile in order to define related tags that were used to rerank the non-personalized search results. Xu et al. [3] proposed a folksonomy-based personalized search using tags extracted from Delicious and Dogear. They applied the topic matching between the user's interests and the document's topic in order to rerank the web pages rather than using only the term matching between the query and the document's content. Carmel et al. [17] explored the user's connections in social networks. They re-ranked search results based on their connection strength with the user's related persons and topics. In fact, they used three types of profiles: with explicit familiarity connections, with connections obtained through common social activities and finally merging both of the previous types.

After collecting information about the user, his profile is often represented as a set of weighted keywords corresponding to the user interests [12] [11], as categories extracted from the Open Directory Project (ODP) hierarchy [4] [18] [7], or as semantic networks [19].

## 2.2 User-Profile Evolution

The evolution of the user profile is represented in prior works as either a short-term profile or a long-term profile. The short-term profile describes the interests and needs of users related to activities of the current search session. Some approaches [6] [22] [7] define the short-term user profile as all the interactions and interests related to a single information need. However, other studies [5] [20] define it as multiple interests emerged in a single time slot.

The second type of profile is the long-term profile, which refers to the use of specific information such as the users education level and general interests, user query history and past user clickthrough information. Teevan et al. [23] developed rich long-term user models based on desktop search activities to improve ranking. They altered the query term weights from the BM25 weighting scheme to incorporate user interests as captured by her desktop index. Similarly, Chirita, et al. [24] capture personalized query expansion terms for web search using three different desktop approaches: the entire desktop data, only the desktop documents relevant to each user query, and a natural language processing techniques to extract diffusing lexical compounds from relevant desktop resources. Besides, Tan et al. [9] studied long-term language model-based representations of users' interests based on queries, documents and clicks. They considered different amount of history and found that for fresh queries recent history was the most important, but for recurring queries long-term history was more significant.

Most of the existing approaches [8] exploit temporal factor after building the user profile in order to track its changing, rather than using it in the modeling step. For instance, [25] captured temporal dynamics of the individual profiles inferred from Twitter activities. They specify two types of users based on their

interactions regarding a specific public topic. The first one is the continuously active users that interact during a long period. The second type is the sporadically active users that interact within a short period.

However, the fact of discerning the short-term and long-term user profiles does not necessarily reflect the user's needs. For users who are not very active on social services, the short-term profile can eliminate relevant results which are more related to their personal interests. This is because their social activities are few and separated over time. In addition, for users who are very active, the aggregation of recent activities without ignoring the ancient interests would be very interesting because this kind of profile is usually changing over time. Furthermore, in Sugiyama, et al. [11], the current interests are used in the active session while the persistent interests are stored for use in later search sessions. For all those reasons, we assume that a user profile can reflect both the recurrent (persistent) and the current (recent) interests but with different scales based on freshness.

In this paper, the user profile is represented as a vector of keywords terms corresponding to the user interests implicitly inferred from his activities on social Web systems. We will adjust the importance of each keyword according to the time of its use unlike non-time-sensitive approaches that do not consider the time but only the frequency [2] [3]. The specificity of the profile resides in weighting the terms by combining their frequency and their freshness. We assume that this way we naturally combine short-term profile and long-term profile into a single one in which we give importance to the recent interests without ignoring the continuous ones.

### 3 Time-Sensitive Personalization Approach

In the classical non-time sensitive approaches, the relevance of an interest in the user profile is assumed to be only decided by the counts of terms in the profile, but not by their position in time. As the user interests evolve, we propose to build a time-sensitive user profile under the assumption that older frequent terms should not outperform current and not frequent terms.

After collecting keywords from the user interactions in social Web, their weights are computed by combining both their frequency and their appearing moment. We consider the social Web but any other source does not affect our approach. More formally, given a document  $D^{S_i} = (t_1, t_2 \dots t_N)$  generated at moment  $S_i$  (day, hour or minute...) ,such as a tag or a microblog. By document, we mean any content generated by the user such as a tag or a microblog. We extract documents' terms and generate their normalized term frequency (nTF) described as:

$$nTF(t_i)^{S_i} = \frac{freq^{S_i}(t_i)}{\sum_{\forall k \in D^{S_i}} freq^{S_i}(t_k)} \quad (1)$$

with:  $freq^{S_i}(t_i)$  is the relative frequency of a term  $t_i$  in  $D^{S_i}$  and  $\sum_{\forall k \in D^{S_i}} freq^{S_i}(t_k)$  represents the sum of the frequencies of all terms appeared in  $D^{S_i}$ .

To measure the freshness of a term, we review the notion of term frequency by adjusting it with a temporal-biased function. In fact, we assume the more the term is closer to the current date  $S^C$ , the more its temporal frequency would be significant. We use the Kernel Gaussian function as a temporal-biased function [26] [27]:

$$K(S^C, S_j) = \frac{1}{\sqrt{2 \cdot \Pi} \cdot \sigma} \cdot \exp \left[ \frac{-(S^C - S_j)^2}{2 \cdot \sigma^2} \right] \quad (2)$$

where  $\sigma$  is the interpolation coefficient,  $S^C$  is the current date and  $S_j$  is a prior date.

Figure 1 illustrates three terms distributions using first a simple cumulative term frequency of three terms (see Fig 1-a), compared with their revised cumulative frequency using Kernel (Fig1-b). We notice that term TF1 starting with high frequency (Fig1-a) its kernel version (CF1) increases slowly. However, term TF3 starting from low frequency (0 in this case), continue to increase until it reaches the same cumulative frequency than TF1. Its kernel version (CF3) over-passes CF1. Term TF2 which has a uniform distribution continue to increase uniformly. Thus, in each date  $S_j$ , we define the user profile as a vector  $\vec{U}$  of terms and their corresponding global weights  $W$ :

$$\vec{U} = (t_1^{S_j} : W_1^{S_j}, t_2^{S_j} : W_2^{S_j}, \dots, t_m^{S_j} : W_m^{S_j}) \quad (3)$$

where the temporal weight  $W(t)^{S^c}$  of a term  $t$  in the profile is the sum of its time-biased relative frequency defined as follows:

$$W(t_k)^{S^c} = \sum nTF(t_k)^{S_j} \cdot K(S^C, S_j) \quad (4)$$

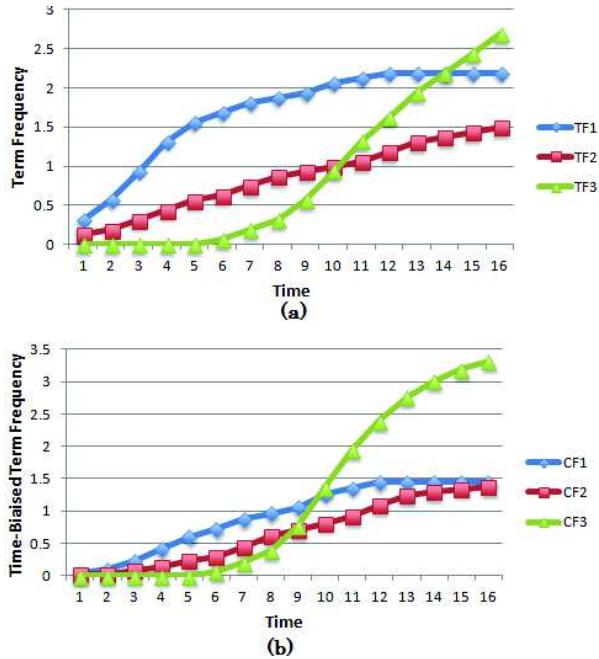
The personalization strategy that we adopt consists of submitting a query to a standard search engine and measuring the similarity between the user profile and each returned Webpage-profile  $\vec{WP} = (t_{wp1}, t_{wp2}, \dots, t_{wpk})$  thanks to the cosine similarity measure. The time-sensitive based user profile can be further refined by smoothing it with the Webpage-query similarity obtained to personalize search results for all user's queries during the search session. Finally, the search results are re-ranked as follows:

$$Score(U, Q) = \alpha \cdot Sim(\vec{U}, \vec{WP}) + (1 - \alpha) \cdot Sim(\vec{WP}, Q) \quad (5)$$

where  $Sim(\vec{WP}, Q)$  is the score obtained from the original results reflecting the matching between the query and the Webpage, and  $Sim(\vec{U}, \vec{WP})$  denotes the user-Webpage similarity. Both of the similarities are computed through the cosine function.

## 4 Experiments and Results

In this Section, we investigate the impact of the time-sensitive user profile strategy in the context of personalized search. More specifically, we examine the



**Fig. 1.** Example of terms distribution using cumulative term frequency and a kernel version of the frequency

impact of our proposed temporal pattern in improving the accuracy of the Web search. Accordingly, our aim is to analyze and compare our approach with two different approaches. The first one is the ranked results returned by Google <sup>1</sup>. The second one is the non-time-sensitive metric nTF [16]. We particularly analyze how the proposed Time-Sensitive User Profile (TSUP), brought in Section 3, affects personalization and achieves better performances in comparison to non-time sensitive variants.

#### 4.1 Data Set and Evaluation Methodology

Over a period of the first two weeks of December 2013, we crawled the microblogging system Twitter<sup>2</sup> posts (tweets) via the Twitter 4j API to randomly select 800 users and extract their 69000 tweets. For each tweet, we combine the relative frequencies with the temporal biased function that evaluates the distribution of the terms over time. The main details of our data set are presented in Table 1.

We select a unique query for each user profile related to his areas of interests defined on Twitter totaling 800 queries. Our queries are randomly selected from the online Twitter categories of interests (computer science, politics, chemistry, ...). Each query is submitted to a Web search engine (Google in our case). We select the top 100 documents per query. These documents are then proceeded to the stop words removal, stemming and tokenization of documents and users' extracted

<sup>1</sup> [www.google.com](http://www.google.com)

<sup>2</sup> [www.twitter.com](http://www.twitter.com)

terms thanks to Apache Lucene<sup>3</sup> classes including Porter Stemming Filter class before weighting the result keywords. Each Webpage is represented as vector of terms  $\vec{WP} = (t_{wp1}, t_{wp2}, \dots, t_{wpk})$ . Each term in the Webpage profile is weighted using the TF-IDF model [3]. Finally, our personalizing approach consists of re-ranking the initial list of Webpages in accordance with each user profile.

**Table 1.** Data statistics

Number of Users	800
Period	01/12/2013 - 15/12/2013
Total Number of Tweets	69000
Average Number of Tweets per participant	86.25
Average Number of Tweets per participant per day	5.75

In order to measure the quality of the results, we use the Normalized Discounted Cumulative Gain (NDCG) at 10 (Jrvelin and Keklinen, 2000) for all the judged queries. Results are judged by 40 voluntary assessors with three levels of relevance, namely highly relevant (value equal to 2), relevant (value equal to 1) or irrelevant (value equal to 0). The assessors are graduate students in different fields, i.e., computer science, chemistry, tourism, electrical engineering, and medical. Each assessor evaluates 20 users' profiles. In addition, we used a second metric which is the Precision at top 10. We considered any positive judgment as relevant.

## 4.2 Results and Analysis

In this Section, we lay out the findings of our analysis. First, we present results obtained by comparing our model with the two baselines described in the beginning of this Section. Then, we try to specify the impact of growing information about the user's activities on the social Web.

**Baselines Comparison Results.** We compare our time-based user profile with the non-personalized results returned by Google and with the re-ranked list of documents returned by modeling the user profile according to the nTF scheme. From this comparison ( $\alpha = 0.6, \sigma = 4$ ), we obtained the values summarized in Table 2 where we use two metrics namely the P@10 and the NDCG@10.

From the results presented in Table 2, we notice that the TSUP approach overcomes the results given by both of Google and the nTF approach for both of NDCG and P@10.

From our point of view, the reason of these promising values is the fact that the term frequency does not reflect the freshness of an interest but gives an overview of how often the user mentioned a term when interacting with the online social systems. However, standard search engines return relevant results to the user query's terms but they are indifferent to the users' interests especially when

<sup>3</sup> [www.lucene.apache.org](http://www.lucene.apache.org)



**Table 2.** Comparison results

	P@10 (Average %)	NDCG (Average %)
<b>Without Freshness</b>		
Google	57.87	45.67
nTF	62.68	58.80
<b>With Freshness</b>		
TSUP personalization	74.72	78.15

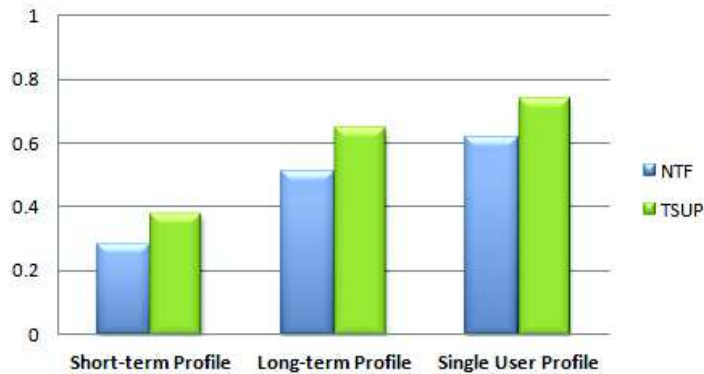
the queries are short [28] or ambiguous [29]. Hence, the time-based user profile strategy defines current interests and needs of a user better than the non-time sensitive one. Furthermore, the standard search engine (e.g, Google) gives the same list of results without considering the user’s individual needs because the ranking is based only on the matching of the document’s terms to the query’s keywords.

Consequently, merging both the freshness-feature and the term-frequency into our proposed weighting scheme has proved its effectiveness. The temporal-function allows considering the actual interests which are used to enhance the current search without overlooking the persistent interests and helps to personalize recurrent information needs.

**Impact of User’s Profile Information Amount.** In order to better evaluate the influence of the temporal feature, we use the same personalization methodology to compare the time-sensitive user profile (TSUP) with the nTF-based user profile in terms of three profiles’ temporal aspects namely:

- Short-term profile: all tweets extracted during the current day,
- Long-term Profile: all previous tweets except for those in the current session (before current day),
- Single Profile: all the recent and old tweets as a single user profile.

As we can see in Figure 2, when we merge both of the interests into a single profile, we have a growing amount of profiling information that leads to better

**Fig. 2.** Comparison of mean of P@10

improvements in retrieval relevance. A single user profile that exploits all user's interests give better results than using profiles based solely on short- or long-term interests. Indeed, our approach outperforms the nTF approach with the three temporal aspects.

## 5 Conclusion

In this paper, we explored the problem of personalized search and developed a user-modeling framework for Twitter microblogging system. In fact, the integration of the social data on the user model is accurate and efficient because people are likely to write a blog or bookmark a Webpage about something that interests them. Further-more, we investigate how the temporal-based user profile influences the accuracy of personalized search. We used a vector-based representation that takes into account the temporal-frequency measured by merging the term frequency and the freshness of each keyword using the Kernel function.

We find encouraging results when we compared our approach to two non-temporal sensitive approaches: the standard search engine Google and the user profiling using the Normalized Term Frequency scheme. In addition, we analyzed the aggregation of the current and recurrent interests. We found that increasing amount of profiling information yields to greater improvement in retrieval performance.

Future work will further research the temporal aspects when enriching the user pro-file by including diverse users' social behaviors by gathering data from multiple sources such as timelines from social networks namely Facebook<sup>4</sup> or bookmarks of interesting Webpages such as Delicious<sup>5</sup>. Moreover a comparison with other temporal models [17] [30] would be required to evaluate the effectiveness of our proposed measure.

## References

1. Qiu, F., Cho, J.: Automatic identification of user interest for personalized search. In: 15th International Conference on World Wide Web, pp. 727–736. ACM, New York (2006)
2. Vallet, D., Cantador, I., Jose, J.M.: Personalizing Web Search with Folksonomy-based User and Document Profiles. In: Gurrin, C., He, Y., Kazai, G., Kruschwitz, U., Little, S., Roelleke, T., Rüger, S., van Rijsbergen, K. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 420–431. Springer, Heidelberg (2010)
3. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 155–162. ACM, New York (2008)
4. Speretta, M., Gauch, S.: Personalizing search based on user search histories. In: 2005 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 622–628. IEEE Press, New York (2005)

<sup>4</sup> [www.facebook.com](http://www.facebook.com)

<sup>5</sup> [www.delicious.com](http://www.delicious.com)

5. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., Robbins, D.: Stuff I've seen: a system for personal information retrieval and re-use. In: 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, pp. 72–79. ACM, New York (2003)
6. Shen, X., Zhai, C.: Exploiting query history for document ranking in interactive information retrieval. In: 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, pp. 377–378. ACM, New York (2003)
7. Daoud, M., Boughanem, M., Tamine-Lechani, L.: Detecting Session Boundaries to Personalize Search Using a Conceptual User Context. In: Ao, S.I., Gelman, L. (eds.) Springer 2009. Advances in Electrical Engineering and Computational Science, vol. 39, pp. 471–482. ACM, New York (2009)
8. Bennett, P.N., White, R.W., Dumais, S.T.: Modeling the Impact of Short- and Long-Term Behavior on Search Personalization. In: 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 185–194. ACM, New York (2012)
9. Tan, B., Shen, X., Zhai, C.: Mining long-term search history to improve search accuracy. In: 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 718–723. ACM, New York (2006)
10. Dou, Z., Song, R., Wen, J.: A large-scale evaluation and analysis of personalized search strategies. In: 16th International Conference on World Wide Web, pp. 581–590. ACM, New York (2007)
11. Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive Web search based on user profile constructed without any effort from user. In: 13th International Conference on World Wide Web, pp. 675–684. ACM, New York (2004)
12. Cai, Y., Li, Q.: Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In: 19th ACM International Conference on Information and Knowledge Management, pp. 969–978. ACM, New York (2010)
13. Kritikopoulos, A., Sideri, M.: The compass filter: search engine result personalization using web communities. In: Mobasher, B., Anand, S.S. (eds.) ITWP 2003. LNCS (LNAI), vol. 3169, pp. 229–240. Springer, Heidelberg (2005)
14. Teevan, J., Morris, M., Bush, S.: Discovering and Using Groups to Improve Personalized Search. In: Second ACM International Conference on Web Search and Data Mining, pp. 15–24. ACM, New York (2009)
15. Orlandi, F., Breslin, J., Passant, A.: Aggregated, Interoperable and Multi-Domain User Profiles for the Social Web. In: 8th International Conference on Symantic Systems I-SEMANTICS, pp. 41–48. ACM, New York (2012)
16. Noll, M.G., Meinel, C.: Web Search Personalization via Social Bookmarking and Tagging. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 367–380. Springer, Heidelberg (2007)
17. Carmel, D., Zwerdling, N., Guy, I., Ofek-Koifman, S., Har'el, N., Ronen, I., Uziel, E., Yogev, S., Chernov, S.: Personalized social search based on the user's social network. In: 18th ACM Conference on Information and Knowledge Management, pp. 1227–1236. ACM, New York (2009)
18. Liu, F., Yu, C., Meng, W.: Personalized Web Search For Improving Retrieval Effectiveness. IEEE Transactions on Knowledge and Data Engineering 16(1), 28–40 (2004)

19. Micarelli, A., Sciarrone, F.: Anatomy and Empirical Evaluation of an Adaptive Web- Based Information Filtering System. In: *User Modeling and User-Adapted Interaction*, vol. 14(2-3), pp. 159–200. Kluwer Academic Publishers Hingham, MA (2004)
20. Pretschner, A., Gauch, S.: Ontology Based Personalized Search. In: *11th IEEE International Conference on Tools with Artificial Intelligence*, pp. 391–398. IEEE Press, New York (1999)
21. Shen, X., Tan, B., Zhai, C.: Context-sensitive information retrieval using implicit feedback. In: *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 43–50. ACM, New York (2005)
22. Zemirli, W.: *Modèle d'accès personnalisé à l'information basé sur les diagrammes d'influence intégrant un profil multidimensionnel*. Paul Sabatier University, Toulouse (2008)
23. Teevan, J., Dumais, S.T., Horvitz, E.: Personalizing search via automated analysis of interests and activities. In: *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 449–456. ACM, New York (2005)
24. Chirita, P., Frien, C.S., Nejdl, W.: Summarizing local context to personalize global web search. In: *15th ACM International Conference on Information and Knowledge Management*, pp. 287–296. ACM, New York (2006)
25. Abel, F., Gao, Q., Houben, G., Tao, K.: Analyzing temporal dynamics in Twitter Profiles for Personalized Recommendations in the Social Web. In: *3rd International Conference on Web Science*, pp. 1–8. ACM, New York (2011)
26. Lv, Y., Zhai, C.: Positional language models for information retrieval. In: *32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 299–306. ACM, New York (2009)
27. Gerani, S., Carma, M.J., Crestani, F.: Proximity Based Opinion Retrieval. In: *33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 403–410. ACM, New York (2010)
28. Jansen, B.J., Spink, A., Saracevic, T.: Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management* 36(2), 207–227 (2000)
29. Cronen-Townsend, S., Croft, W.B.: Quantifying query ambiguity. In: *Second International Conference on Human Language Technology Research*, pp. 104–109. Morgan Kaufmann Publishers Inc., San Francisco (2002)
30. Billsus, D., Pazzani, M.J.: User Modeling for Adaptive News Access. In: *Journal of User Modeling and User-Adapted Interaction*, vol. 10, pp. 147–180. Kluwer Academic Publishers Hingham, MA (2000)