User-Centered Social Information Retrieval Model Exploiting Annotations and Social Relationships

Chahrazed Bouhini, Mathias Géry, Christine Largeron

User-centered Social Information Retrieval Model Exploiting Annotations and Social Relationships

Chahrazed Bouhini, Mathias Géry, and Chritine Largeron

Université de Lyon, F-42023, Saint-Étienne, France
CNRS, UMR 5516, Laboratoire Hubert Curien, 42023, Saint-Étienne, France
Université de Saint-Étienne, Jean-Monnet, F-42023, Saint-Étienne, France
{Chahrazed.Bouhini,Mathias.Gery,Christine.Largeron}@univ-st-etienne.fr

Abstract. Social Information Retrieval (SIR) has extended the classical information retrieval models and systems to take into account social information of the user within his social networks. We assume that a SIR system can exploit the informational social context (ISC) of the user in order to refine his retrieval, since different users may express different information needs as the same query. Hence, we present a SIR model that takes into account the user's social data, such as his annotations and his social relationships through social networks. We propose to integrate the user's ISC into the documents indexing process, allowing the SIR system to personalize the list of documents returned to the user. Our approach has shown interesting results on a test collection built from the social collaborative bookmarking network *Delicious*.

Keywords: Social information retrieval model, social test collection, annotations, relationships, indexing

1 Introduction

The participants of social networks are not only allowed to share Web documents but also to annotate, to evaluate and to comment them [19]. Social tagging data, known as *folksonomies*, create social association between the users and the Web pages through the social annotations [20]. Social annotation is a set of tags (keywords) freely assigned by a user to describe the content of a Web document. To this end, annotations are widely considered as an effective means of enriching content with meta-data. *Folksonomies* can be considered as a fairly accurate source to discover user interests [18]. In this context, Social Information Retrieval (SIR), defined as the incorporation of information related to social networks and relationships into the information retrieval process [8], attempts to extend classical IR by taking into consideration the user's ISC within his social network. Indeed, social networks users may be seeking different informations expressed by the same queries. Thus, SIR systems can exploit the user's ISC to refine the user's retrieval. Hence, one aim of SIR consists in adapting usual IR models

and systems in order to deal with this user's social data (user's annotations and social relationships). We propose a SIR model, called $BM25F_S$, which integrates user's ISC during the indexing step. More specifically, the user's ISC is generated out of his annotations. Then, we benefit from social relationships of the user, that we call neighborhood, to enrich the user's ISC with the annotations of his neighborhood. Once the user's ISC generated, we investigate the way to integrate this social information into the SIR model.

This paper is organized as follows: we discuss some related works in Section 2 and we explain the motivations of our work in Section 3. In Section 4, we present our methodological framework by describing the approach that we used to generate the user's ISC and we detail the main contribution of our work which consists in integrating the user's ISC in the documents indexing step in order to build a personalized documents index. Further, we present some results of experiments done on a test collection generated from the collaborative bookmarking network **Delicious**[1] in Section 5, and we conclude with some perspectives.

2 Related Work

Social networks provide valuable additional information which have been used to improve the results of recommendation systems [6], collaborative filtering [5], or information retrieval [7]. Much related works use social informations for query expansion and disambiguation [15], [9], [4].

In this paper, we focus on the use of these social annotations and relationships in the SIR models for IR personalization. For example, with the aim to improve Web search, Bao et al. [2] propose two methods: SocialSimRank and SocialPageRank. The former allows to find the latent semantic association between queries and annotations, while the latter takes into account the popularity of web pages [2].

The first step of the IR personalization aims at modeling the user's profile and social context [1], [3], [18], [20], [17]. Indeed, several studies have proven that user's ISC can be effectively harvested from the social bookmarking systems [1], [17]. These works assume that the documents and the tags posted by users depend highly on their interests and provide rich information for building user profiles [3], [18], [20].

The second step aims at integrating the user's ISC into the SIR model by combining different weighting function. Authors in [3], [18], [20], [11], propose to personalize the user's search by ranking the resource based on, a matching between the user's interests and the documents' topics [20], or between the user's profile and the resource's profile [18], [3]. Rather than considering the resource's content, authors in [3], [18], [11] propose to build a *resource's profile* through the resource's annotations and compute a matching function between this resource's profile and the query terms, on the one hand and between the resource's profile and the user's profile, on the other hand.

[1] https://delicious.com/

Unlike in [3], [18] and [11], Xu et al., consider to match the document's content instead of matching only the document's profile over the query terms and the user's profile. The success they achieved is a strong support for our work [20]. Although, Cai et al., [3] discuss the limits of the weighting functions used in the previous cited works [18] [20], [11] as for a case study based only on a set of tags for resource recommendation, they propose a normalized term frequency to indicate the preference degree of a tag for the user and the representative degree of a tag for the resource [3]. We note that the user's profile generated in these works is based only on the user's annotations without exploring his relationships.

In our approach, we assume that the document content is useful for IR. Thus, using only the document profile is not enough. We assume also that it is important to exploit the social annotations of the user's relationships (neighborhood) in addition to his own social annotations. As Stoyanovich et al. [16] show, the predicted relevance of documents may be enhanced by exploiting the user neighbors' tagging actions [16]. Finally, we think that combining the document content with this ISC requires IR techniques that are able to handle large textual documents. This led us to introduce an original approach presented in the following sections.

3 Personalized IR exploiting folksonomies

3.1 Ambiguous queries

Almost all test collections, in IR research, assume that queries have a single interpretation representing the information need expressed by one user, which is implicitly defined in his relevance judgments [14]. However, in practice this is not necessary the case. For this reason, in this paper we propose a framework for personalized information retrieval based on folksonomies. Such a system should be able to handle ambiguous queries, i.e. queries having potentially several interpretations representing different information needs.

For example, suppose that two users u_1 and u_2 have the same query $q =$ "smartphone android" (cf. Table 1). We consider two documents d_1 and d_2; each document contains one query term, but *smartphone* is more important than *android* in the first document since d_1 contains only *smartphone*, and *android* is more important than *smartphone* in the second one since d_2 contains only *android*. Assuming that the two query terms have the same importance, a classical IR system should estimate that d_1 is equally relevant as d_2 for the query "*smartphone android*". However, depending on the user and his personal interests, the information need behind this query may focus either on the term *smartphone* or on the term *android*. The user u_1 is mainly interested in smartphone devices, then his information need is probably centered around smartphones with an opening on Android, and thus the query term *smartphone* should be more important than the query term *android*. On the other hand, the user u_2 is mainly interested by the Android operating system, consequently his information need is probably centered around Android, and thus the query term *android* should be more important than the query term *smartphone*.

	t_1 = smartphone	t_2 = android	t_3 = features
Query q	1	1	0
Document d_1	1	0	0
Document d_2	0	1	1
User u_1	2	1	0
User u_2	1	2	0

Table 1. Example: query, documents and user's profiles.

3.2 Personalized information retrieval

A personalized information retrieval system should be able to identify the user's personal interests, in order to better interpret the information need behind his queries, and returns lists of relevant documents to the users depending on their personal interests. In our example, a personalized IR system should consider d_1 as more relevant than d_2 for u_1, and the opposite for u_2.

3.3 Folksonomies and user's informational social context (ISC)

We assume that folksonomies may be exploited in order to build the informational social context of the user that could represent the user's interests and that could help the system to handle ambiguous queries return personalized results to the user.

As pointed out in related literature, the user's profile can be inferred from his social annotations ([1], [17]). The neighborhood's profile is defined by the annotations of his neighborhood. We assume that the user's profile can be enriched by the annotations of his neighbors to build the user's ISC.

3.4 Integrating user's ISC within the IR model

Since we exploit the social information about the user and his neighborhood to generate his ISC, we assume that the important terms representing the user's interests should appear in this ISC. Thus, reweighting such important terms when they are found within the document, should improve the document relevance score and allow to return the personalized relevant documents. We think that the integration of the user's ISC within the IR model is an important part of the personalization. One aim of this work is to handle textual documents containing thousands of terms, unlike most related work which only handle small sets of tags describing the document. Combining the user's ISC with this kind of textual data raises different issues than combining two sets of tags, like for instance in the work of Cai et al. [3]). Our work attempts to deal with this issue.

4 Social information retrieval model

We present a SIR model, called $BM25F_S$, that takes into account the user's ISC, to better describe the documents with respect to the user viewpoint.

4.1 Notations

We represent the "social tagging data", also known as *Folksonomies* [18], by a tuple $< U,\ Rel,\ T,\ D,\ A >$, where:

- $U = \{u_1, u_2, ..., u_x, ..., u_{|U|}\}$ is a set of social network users.
- $Rel \subseteq U \times U$ is a set of relationships between pairs of users, such that $(u_x, u_y) \in Rel$ iff there is a social relationship between a user u_x and another user u_y. The users related to u_x are typically those declared explicitly by u_x as his $neighbors$ where $neighborhood(u_x) = \{u_y \ / \ (u_x, u_y) \in Rel\}$.
- $T = \{t_1, t_2, ..., t_j, ..., t_{|T|}\}$ is a set of index terms.
- $D = \{d_1, d_2, ..., d_i, ..., d_{|D|}\}$ is a set of documents on the Web (images, videos, Webpages, etc.). A document d_i is represented by a set of terms ($t_j \in T$) and a term t_j may appear one or several times in a document d_i. We denote by tf_{ij} the term frequency of t_j in d_i. A weight w_{ij} of a term t_j for a document d_i is computed using this term frequency tf_{ij} of t_j in d_i.
- $A = \{a_1, a_2, ..., a_z, ..., a_{|A|}\}$ is a set of social annotations, i.e., $a_z = < d_i, u_x, T_z >$ is the annotation of the user u_x for the document d_i using a subset of terms $T_z \subset T$.

We define also:

- $Q = \{q_1, q_2, ..., q_l, ..., q_{|Q|}\}$, a set of users' queries, where each query q_l is represented by a set of terms.
- $Qrels = \{qrels_1, qrels_2, ..., qrels_l, ..., qrels_{|Q|}\}$, a set of global relevance judgments, where $qrels_l \subset D$ denotes the set of relevant documents for q_l.
- $Q_{UC} = \{(q_l, u_x) \subset Q \times U\}$, a set of couples (q_l, u_x) where the query q_l is issued by the user u_x to express his information needs.
- $Qrels_{UC} = \{qrels_{1,1}, qrels_{1,2}, ..., qrels_{l,x}, ..., qrels_{|Q|,x}\}$, a set of user-centered relevance judgments, where $qrels_{l,x} \subset D$ denotes the set of relevant documents for the query q_l and the user u_x.

4.2 Information retrieval model: BM25

We choose as baseline the IR weighting function BM25 [13], which is one of the most used indexation models in the IR research benchmarks such as INEX[2], TREC[3], etc. In this IR weighting function, the weight of a term t_j within a document d_i is computed according to the formula (1):

$$w_{ij} = \frac{(k_1 + 1) \times tf_{ij}}{k_1 \times ((b-1) + b \times (\frac{dl_i}{avgdl})) + tf_{ij}} \times log(\frac{N - df_j + 0.5}{df_j + 0.5}) \qquad (1)$$

where:

- dl_i is the document length of d_i and $avgdl$ is the average documents length.

[2] INEX (INitiative for the Evaluation of XML-Retrieval): https://inex.mmci.uni-saarland.de/

[3] TREC (TExt Retrieval Conference): http://trec.nist.gov/

- tf_{ij} is the term frequency of t_j within the document d_i.
- k_1 is the saturation parameter of tf_{ij}.
- b is the length normalization factor.
- N is the total number of documents in the corpus.
- df_j is the number of documents containing the term t_j.

In the $BM25$ model, the global score of a document d_i for a query q_l is computed as follows:

$$BM25(q_l, d_i) = \sum_{t_j \in q_l \cap d_i} w_{ij} \tag{2}$$

4.3 User informational social context (ISC)

The user's informational social context may contain different information types (annotations, comments, citations, social relationships, etc.). In this work we generate the user's ISC from the terms in his annotations and those of his neighbors. We present two variants of the user's ISC: $ISC_u(u_x)$, called "the user's profile" and $ISC_n(u_x)$, called "the neighborhood's profile":

The user's profile $ISC_u(u_x)$ is the set of terms which occur within the social annotations of the user.

$$ISC_u(u_x) = \{t_j \in T_z \ / \ a_z = < d_i, u_x, T_z > \in A_{u_x}\} \tag{3}$$

where: A_{u_x} is the set of social annotations of u_x.

The user's profile may contain several occurrences of the same term. Thus we can compute the term frequency tfu_{xj} for a given term t_j that has been used by u_x to annotate the documents. In the example provided in Table 1, the user profiles with the tfu_{xj} associated are represented as follows: $ISC_u(u_1) = ISC_u(u_2) = \{"smartphone","android"\}$ and the tfu_{xj} associated are (2,1) for u_1 and (1,2) for u_2.

The neighborhood's profile $ISC_n(u_x)$ is the set of terms which occur within the user neighborhood's annotations.

$$ISC_n(u_x) = \cup_{u_y \in U \ / \ (u_x,u_y) \in Rel} ISC_u(u_y) \tag{4}$$

Like previously, the neighborhood's profile may also contain several occurrences of the same term t_j, and we can compute the term frequency tfn_{xj} for a given term t_j that has been used by the neighborhood of u_x to annotate the documents. In our example, the $neighborhood(u_1)$ is composed of u_3 and the $neighborhood(u_2)$ is composed of u_4, with the user's profiles for u_3 and u_4. Then, we can compute the neighborhood's profiles for u_1 and u_2, given in Table 1.

4.4 Personalized index

Now, our aim is to take into account the user's ISC during the indexing step, in order to personalize the documents index.

	t_1 = smartphone	t_2 = android	t_3 = features
Query q	1	1	0
tf_{1j}	1	0	0
tf_{2j}	0	1	1
tfu_{1j}	2	1	0
tfu_{2j}	1	2	0
tfu_{3j}	2	3	0
tfu_{4j}	3	3	1
tfn_{1j}	2	3	0
tfn_{1j}	3	3	1

Table 2. Example: term frequencies of the user's ISC.

We combine the content of the document d_i, represented by a vector of term frequencies tf_{ij}, with the user's ISC, composed of two vectors of term frequencies tfu_{xj} and tfn_{xj}. Each document is indexed by a vector of weights ws_{xij}, with ws_{xij} the weight of the term t_j in the document d_i for the user u_x. The term weight ws_{xij} is a personalized version of w_{ij} for the user u_x.

We propose to combine the document's content and the user's ISC as three different fields of information (i.e. 3 vectors). As it has been shown more coherent than combining the score of each vector computed independently, in [12]. Thus, we built a personalized documents index based on these three fields:

- the *content of* d_i, represented by a vector of *field term frequencies* ftf_{xij} is equal to the classical tf_{ij}:
$$ftf_{xij} = tf_{ij} \tag{5}$$
- the *user's profile* $ISC_u(u_x)$, represented by a vector of *field term frequencies* $ftfu_{xij}$. Only the weights of the terms appearing both in the content of the document and in the user's profile should be considered:
$$ftfu_{xij} = \begin{cases} tfu_{xj} \text{ if } tf_{ij} > 0 \\ 0 \quad \text{else.} \end{cases} \tag{6}$$
where tfu_{xj} is the term frequency of t_j used by u_x in his annotations.
- the *neighborhood's profile* $ISC_n(u_x)$, represented by a vector of *field term frequencies* $ftfn_{xij}$. Similarly, only the weights of the terms appearing both in the content of the document and in the neighborhood's profile should be considered:
$$ftfn_{xij} = \begin{cases} tfn_{xj} \text{ if } tf_{ij} > 0 \\ 0 \quad \text{otherwise.} \end{cases} \tag{7}$$
where tfn_{xj} is the term frequency of t_j used by the neighborhood of u_x to annotate documents.

The BM25F weighting function has been proposed by [12] in order to index structured documents composed of several *fields* (e.g. *title*, *abstract*, *body*, etc.). $BM25F$ seems to be suitable for indexing our three fields-documents. This

function was extended by [21] in order to optimize the length normalization field-by-field. We chose to use this latter BM25F variant. Then, like Zaragoza et al., the first step of the BM25F function normalizes the term frequencies of each field by the field length [21]:

$$\overline{ftf}_{xij} = \frac{ftf_{xij}}{1 + b_d \times (\frac{dl}{avgdl} - 1)} \tag{8}$$

$$\overline{ftfu}_{xij} = \frac{ftfu_{xij}}{1 + b_{ux} \times (\frac{ul}{avgul} - 1)} \tag{9}$$

$$\overline{ftfn}_{xij} = \frac{ftfn_{xij}}{1 + b_{nx} \times (\frac{nl}{avgnl} - 1)} \tag{10}$$

where:

- b_d, b_{ux} and b_{nx} are some field-dependent parameters, similar to b (in BM25), for the d_i content field, the user's profile field and the neighborhood's profile field, respectively,
- ul and nl are the length of the user's profile field and the length of the neighborhood's profile field, respectively,
- $avgul$ and $avgnl$ are the average length of the user's profile field over the collection and the average length of the neighborhood's profile field over the collection.

Then, following Zaragoza's BM25F [21], we compute the term weight ws_{xij} for t_j within the document d_i for the user u_x using the weighting function in formula 11:

$$ws_{xij} = \frac{ctf_{xij}}{k_1 + ctf_{xij}} \times log\left(\frac{N - df_j + 0.5}{df_j + 0.5}\right) \tag{11}$$

where:

- ctf_{xij} is the combined term frequency of the three fields:

$$ctf_{xij} = w_d.\overline{ftf}_{xij} + w_{ux}.\overline{ftfu}_{xij} + w_{nx}.\overline{ftfn}_{xij} \tag{12}$$

- w_d, w_{ux} and w_{nx} are three field-dependent parameters used to tune the importance of the user's profile field and the neighborhood's profile field in relation to the importance of the document content field.

Finally, the personalized relevance score of a document d_i for the query q_l and the user u_x is given by equation 13:

$$BM25F_S(q_l, d_i, u_x) = \sum_{t_j \in q_l \cap d_i} ws_{xij} \tag{13}$$

Table 3 shows the field term frequencies for the example given in Table 1 and the relevance scores obtained using $BM25$ and $BM25F_S$. These scores have

U	D	ftf	T			$BM25(q_l, d_i)$	$BM25F_S(q_l, d_i, u_x)$
			t_1	t_2	t_3		
u_1	d_1	$ftfu_{11j}$	2	0	0	1.681	1.983
		$ftfn_{11j}$	3	0	0		
		ftf_{11j}	1	0	0		
	d_2	$ftfu_{12j}$	0	1	0	1.681	1.898
		$ftfn_{12j}$	0	3	0		
		ftf_{12j}	0	1	1		
u_2	d_1	$ftfu_{21j}$	1	0	0	1.681	1.898
		$ftfn_{21j}$	2	0	0		
		ftf_{21j}	1	0	0		
	d_2	$ftfu_{22j}$	0	2	0	1.681	1.983
		$ftfn_{22j}$	0	3	1		
		ftf_{22j}	0	1	1		

Table 3. Impact of the user's ISC on the indexing process.

been computed using the formula 11, with usual BM25 parameters values: $b_d = b_{ux} = b_{nx} = 0.75$ and $k = 1.2$.

The document relevance score increases when the frequencies of the user's profile terms are combined to those of the document terms. Furthermore, for two different user's ISCs with the same query terms, the ranking of the documents could vary according to the ISC of each user. For instance, when the user's ISC is considered, the document d_1 is more relevant than d_2 for u_1 (1.983 vs 1.898), whereas d_2 is more relevant than d_1 for u_2.

5 Experiments

Experiments have been carried out to evaluate the SIR model that considers the user's ISC compared to the classical IR model. Using the social test collection (Del_{SIR}) described bellow, and the evaluation measures *MAP* (Mean Average Precision) and *P[0.1]* (the precision at 10% of recall) [10], we evaluated the rankings produced:

- by a classical IR model with two kinds of data: the first one is composed of 79 global queries (Q) and their global relevance judgments ($Qrels$), the second one is composed of 244 user-centered queries (Q_{UC}) and corresponding user-centered relevance judgments ($Qrels_{UC}$).
- by our SIR model ($BM25F_S$) with only user-centered data (Q_{UC} and $Qrels_{UC}$), since the SIR model is not suited to handle global queries.

5.1 Social test collection

To the best of our knowledge, no SIR test collection exists providing a list of relevant documents for each user. So, we built a test collection Del_{SIR} based on Web

documents and user annotations extracted from the social collaborative bookmarking network *Delicious*. We collected 30,224 documents annotated by 370 users with 21,284 terms. To complete the social dataset with the user-centered data composed of pairs $(query, user)$ and user-centered relevance judgments, we created automatically 79 queries. Each query is composed by 2 terms occuring frequently together [4] in the annotations collected from Delicious. Then we generated 4,685 global relevance judgments and user-centered relevance judgments. A document is globally relevant if it has been annotated by any user with the 2 query terms in the same annotation. A document is user-relevant if it has been annotated by the user with the 2 query terms in the same annotation.

We kept only the pairs $(query, user)$ with at least 10 relevant documents ($|qrels_{l,x}| \geq 10$) so that we obtained 244 pairs (q_l, u_x) in the set Q_{UC}. This led to a reduction of the number of users (70 users left).

5.2 Evaluation results with the classical IR model $BM25$

For each given query, the classical IR model returns the same relevant documents whoever the user is. The results quality obtained with the $BM25$ decreases for both the MAP (0.0308 vs 0.1012) and the P[0.1] (0.0521 vs 0.1775) when the user-centered relevance judgments $Qrels_{UC}$ are considered. This confirms our expectations that the classical IR system has to adapt its models in order to deal with the user-centered data (Q_{UC} and $Qrels_{UC}$).

5.3 Evaluation results with the SIR model $BM25F_S$

In our experiments we have tuned the same way as in [21], using the grid-based 2D optimization, the b and k parameters for BM25 and for each field of $BM25F_S$. Table 4 shows results obtained with two fields-weight settings:

- $BM25F_S$, $settings_1$: $w_d = 1$, $w_{ux} \in]0..1]$ and $w_{nx} = 0$
- $BM25F_S$, $settings_2$: $w_d = 1$, $w_{ux} \in]0..1]$ and $w_{nx} \in]0..1]$

We selected the users having at least 5 queries and obtaining with classical IR a MAP result between 0.5% and 50%. We obtain a set of 10 users, corresponding to a set of 60 pairs (query, user) The results of our SIR model $BM25F_S$ compared to the baseline $BM25$ are shown in Table 4.

Considering the precision at 10% of recall (P[0.1]), the SIR model provides less good results than the baseline. While, using the Mean Average Precision measures (MAP), the SIR model $BM25F_S$ results (MAP = 0.0297 and MAP = 0.0293) are statistically better than the baseline (MAP = 0.0257). The significance has been checked by using statistical tests based on Wilcoxon matched-pairs signed-rank test at the 0.05 level, i.e. the improvement is significant when the p-value is less than 0.05. These results, obtained on a set of 60 pairs (q_l,u_x) for the 10 users in Table 4, confirm that a SIR model which takes into account the user's ISC, with or without neighborhood, enhance the relevance score results using the MAP measure which is considered as a global evaluation metric.

[4] In fact, the 79 couples of terms having the highest Jaccard Index

	$BM25$		$BM25F_S$			
			$settings_1$		$settings_2$	
	MAP	$P[0.1]$	MAP	$P[0.1]$	MAP	$P[0.1]$
u_1	0.0614	0.1310	0.0816	0.1426	**0.0819**	**0.1503**
u_2	0.0404	**0.2614**	0.0402	0.1265	**0.0416**	0.1203
u_3	0.0358	0.1076	**0.0486**	0.1438	0.0483	**0.1438**
u_4	0.0262	0.0922	**0.0278**	**0.0956**	0.0275	0.0948
u_5	**0.0287**	0.0569	0.0253	0.0484	0.0284	**0.0688**
u_6	0.0174	0.0529	**0.0199**	**0.0568**	0.0197	0.0550
u_7	**0.0183**	0.0207	0.0173	0.0298	0.0148	**0.0391**
u_8	0.0138	0.0296	0.0148	0.0316	**0.0156**	**0.0319**
u_9	0.0074	**0.0221**	**0.0093**	0.0215	0.0091	0.0188
u_{10}	0.0077	0.0095	0.0085	**0.0115**	**0.0103**	0.0101
Average	0.0257	**0.0784**	0.0293	0.0708	**0.0297**	0.0733

Table 4. $BM25F_S$ evaluation results.

6 Conclusion

We presented an approach that integrates the user's ISC into the documents. The user's ISC has been built using the user's annotations and those of his relationships. The aim is to personalize the user's search by considering his preferences and interests. Our approach allows highlight and reweight the important terms of the user's ISC when they are found in the document content. As we consider textual documents containing thousands of terms, we proposed to combine the user's ISC with textual content, in the weighting function. The SIR model, that considers the user's ISC, allows to better find the relevant documents for the user query than the classical IR model which does not consider the user's ISC. As future works, we plan to extend the user's ISC with further social data including the neighborhood of neighborhood's (friends of friends) annotations and build a bigger social test collection from *Delicious*' bookmarks. We would also like to study different parameters with further experiments to evaluate our SIR model.

References

1. Au-Yeung, C.m., Gibbins, N., Shadbolt, N.: A study of user profile generation from folksonomies. In: Workshop on Social Web and Knowledge Management. SWKM (2008)
2. Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., Su, Z.: Optimizing web search using social annotations. In: World Wide Web. pp. 501–510. WWW'07 (2007)
3. Cai, Y., Li, Q.: Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In: 19th Conference on Information and Knowledge Management. pp. 969–978. CIKM'10 (2010)
4. Chirita, P.A., Firan, C.S., Nejdl, W.: Personalized query expansion for the web. In: 30th Conference on Research and Development in Information Retrieval. pp. 7–14. SIGIR'07 (2007)

5. Ferrara, F., Tasso, C.: Improving collaborative filtering in social tagging systems. In: 14th Conference on Advances in Artificial Intelligence: spanish association for artificial intelligence. pp. 463–472. CAEPIA'11 (2011)
6. Guy, I., Zwerdling, N., Ronen, I., Carmel, D., Uziel, E.: Social media recommendation based on people and tags. In: 33rd Conference on Research and Development in Information Retrieval. pp. 194–201. SIGIR'10 (2010)
7. Hotho, A., Jaschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: The Semantic Web: Research and Applications, vol. 4011, pp. 411–426 (2006)
8. Kirsch, S.M.: Social Information Retrieval. Ph.D. thesis, Rheinische Friedrich-Wilhelms-Universitat Bonn (2005)
9. Lin, Y., Lin, H., Jin, S., Ye, Z.: Social annotation in query expansion. In: 34th Conference on Research and Development in Information Retrieval. pp. 405–414. SIGIR'11 (2011)
10. Manning, C.D., Raghavan, P., Schutze, H.: Introduction to Information Retrieval. Cambridge University Press; 1st edition (2008)
11. Noll, M.G., Meinel, C.: Web search personalization via social bookmarking and tagging. In: The Semantic Web, Lecture Notes in Computer Science, vol. 4825, pp. 367–380 (2007)
12. Robertson, S., Zaragoza, H., Taylor, M.: Simple BM25 extension to multiple weighted fields. In: 13th Conference on Information and Knowledge Management. pp. 42–49. CIKM'04 (2004)
13. Robertson, S.E., Walker, S.: Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In: 17th Conference on Research and Development in Information Retrieval. pp. 232–241. SIGIR'94 (1994)
14. Sanderson, M.: Ambiguous queries: test collections need more sense. In: 31st Conference on Research and Development in Information Retrieval. pp. 499–506. SIGIR'08 (2008)
15. Schenkel, R., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J.X., Weikum, G.: Efficient top-k querying over social-tagging networks. In: 31st Conference on Research and Development in Information Retrieval. pp. 523–530. SIGIR'08 (2008)
16. Stoyanovich, J., Amer-Yahia, S., Marlow, C., Yu, C.: Leveraging tagging to model user interests in del.icio.us. In: AAAI Spring Symposium: Social Information Processing. pp. 104–109 (2008)
17. Szomszor, M., Alani, H., Cantador, I., O'Hara, K., Shadbolt, N.: Semantic modelling of user interests based on cross-folksonomy analysis. In: Semantic Web Conference. pp. 632–648 (2008)
18. Vallet, D., Cantador, I., Jose, J.M.: Personalizing web search with folksonomy-based user and document profiles. In: 32nd European Conference on Advances in Information Retrieval. pp. 420–431. ECIR'10 (2010)
19. Volkovich, Y., Kaltenbrunner, A.: Evaluation of valuable user generated content on social news web sites. In: WWW (Companion Volume). pp. 139–140 (2011)
20. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: 31st Conference on Research and Development in Information Retrieval. pp. 155–162. SIGIR'08 (2008)
21. Zaragoza, H., Craswell, N., Taylor, M., Saria, S., Robertson, S.: Microsoft cambridge at TREC 13: Web and hard tracks. In: TExt Retrieval Conference. TREC'04 (2004)