Introducing a Personal Semantics in the User Profile Modeling for a Personalized Information Retrieval

Tahar RAFA[1,2], Samir KECHID[2]
[1]Department of Mathematics and Computer science
University of MEDEA, Algeria
rafa.tahar@gmail.com, rafa.tahar@univ-medea.dz

[2]LRIA Laboratory, Department of Computer Science
USTHB, Algiers, Algeria
kechidsam@yahoo.fr, skechid@usthb.dz

Journal of Digital Information Management

ABSTRACT: *The objective of personalized information retrieval systems is to infer the real user information need hidden behind his query, and then tailor the search results to this need. The user information need may be relative to one of the user's multiple search contexts, such as user current location, user social context or an important event that occupies the public opinion of the user community. In this paper we present a new personalized information retrieval approach in which we design and use a user profile. This profile takes into account the different social, situational and temporal contexts of the user previous searches, and we introduced a personal semantics in the representation of these search contexts. The introduced personal semantics is interested by the identification and the valorization of the different co-occurrence relationships between the different relevant components of the user search history. The user profile, boasting its semantic vision, is used to infer the user current information need by the enrichment of the initial user query without going beyond the user current interest.*

Subject Categories and Descriptors
[H.3.3 Information Search and Retrieval]; Retrieval models: **[H.3.4 Systems and Software]**; User profiles and alert services

General Terms: Information Retrieval, User Profile, Personal Information Service, Semantics

Keywords: Personalized Information Retrieval, User Profile, Personal Semantics, Geo-Social Context, Event Extraction

Received: 1 November 2018, Revised 2 December 2018, Accepted 19 December 2018

Review Metrics: Review Score 4.95/6, Revise Scale: 0-6, Inter-reviewer Consistency: 72%

DOI: 10.6025/jdim/2019/17/2/61-74

1. Introduction

The basic hypothesis of the personalized information retrieval (PIR) is that the same answer for the same query sent by two different users may not satisfy both (Bouhini et al., 2016). So, the objective of PIR systems is to infer the real user information need hidden behind his query, and then tailor the search results to this need (Daoud et al., 2011).

According to (Jansen & Rieh, 2010; Cole, C. 2011), the user information need must be contextualized in the user situation in order to be meaningful. To achieve this contextualization of the current user information need, we need to use some user data like the user search history, the user current location and the user socio-temporal context. Usually, the PIR approaches use a structure called user profile, allowing a detailed description of the user and his interests (Nazim- Uddin et al., 2012). This profile gathers data explicitly given by the user such as identity, interest domains, preferences etc. and other data implicitly extracted from his relevant search history, his history of social, situational and temporal information etc. (Bouhini et al., 2016; Buidguaguen, 2011; Basile et al, 2014). The user profile is integrated in the search process to improve the adaptation of the search results to the user need.

Several works in personalized information retrieval represent the user profile using relevant terms of the user search history (Nazim-Uddin et al., 2012). Nevertheless, a search process based only on the syntax (keywords / terms), seems closer to a pattern recognition operation, than to an operation of assimilation and satisfaction of a user cognitive need often expressed by a short and ambiguous query (Aimé et al., 2010). So, without exploitation of the semantic relationships between terms, the information retrieval remains unable to benefit to the richness of the human language for a deep understanding of the user query, the user need and the targeted documents (Fouad et al., 2012). Indeed, many documents whose content is semantically similar to the query may not be selected, because they use other terms to describe differently the same subject.

In recent years, many PIR approaches are oriented to the introduction of semantics in the user profile to improve the accuracy and the relevance of the search results by better understanding the user's intentions (Soha and Abdelmoty, 2017; Duong et al., 2013). Indeed, semantics does not consider documents as a simple set of unrelated terms, but rather it takes into account the meaning conveyed by these terms. The semantic-based PIR approaches rely, for the semantic links extraction, on the exploitation of external semantic resources such as ontologies, thesaurus, dictionaries etc. (Zargayouna et al., 2015).

The semantics introduced in this work is different of the ontological semantics. We are interested by the semantics latent in the user's search history. Our basic hypothesis is that the co-occurrence relationships between terms are meaning carriers. From our viewpoint, the frequent cooccurrence, in many documents, of two terms may mean that they collaborate to describe the same subject. For example, we can deduct from the co-occurrence of "*apple*" and "*computer*" that concerning the "*computer science*" subject, and from the co-occurrence of "*apple*" and "*tree*" that concerning the "*agriculture*" subject. So, we believe that we can extract some semantics based on the co-occurrence links between relevant components of the user's search history, and that this semantics, which we call "*personal semantics*", can help to better represent the user interests.

The idea of personal or non-ontological semantics is not without precedent. Indeed, in the state of the art on the semantic information retrieval of Zargayouna et al. (2015) is mentioned the subject of ontology populating, which consists of enriching an ontology with new semantic annotations. Aimé et al. (2010) propose a personalization of domain ontology, introducing the notion of *over-concept* that represents the concept's consensual interpretation given by the individuals of a specific domain.

Our new idea, inspired from these previous works, is to introduce personal semantics in a user profile. The semantic vision applied here is not limited to considering the cooccurrence relationships only between terms, but rather, we extend it to taking into account different user's social, temporal and situational search contexts. These contexts are qualified as importance providers for a large part of relevant terms of the user search history. So we consider also, the co-occurrence relations linking terms to the user's social annotations, to the important events and to the different information about user's current location.

In this paper, we propose a new PIR approach based on designing and exploiting a user profile. In this user profile we represent and we use, with a semantic vision, the social, the situational and the temporal information derived from the relevant user's search history. The goal of our approach is to exploit the relevance sensors provided by the user profile to improve the enrichment of the user query to better express the user information needs, which improves the search results relevance.

After this introduction, the remainder of the paper is organized as follows: Section 2 is dedicated to related works. Then, event definition and modeling are presented in section 3. In Section 4 our proposed user profile is detailed and the proposed search process is presented in section 5. Section 6 is devoted to the approach evaluation. Finally, section 7concludes the presented work and highlights future works.

2. Related Works

2.1. On User Profile Modeling

The user-oriented information retrieval literature is full of approaches that offer an important variant of models to designing user profiles. In addition to the characteristics explicitly introduced by the user, each approach has its viewpoint concerning implicit relevance sensors that can better characterize the user.

Social Information: Some works consider the social context as a discriminating factor to identify the user interests. These approaches potentially use the user relevant history of social activities (annotations, marking, sharing, friendship links ...) to build user profile. Bouhini et al. (2016) propose an approach to search personalization based on the combination of user social annotations and the relevant terms of his profile. Saoud and Kechid (2016) present, in a context of distributed information retrieval, a personalization approach combining the user folksonomy and user social relations for the construction of user profile. Lin et al. (2014) propose, in a personalized recommendation framework, a user profile combining social annotations and the thematic content of the relevant documents. Jeon et al. (2010) exploit the social relations between users for the search personalization. They use, if necessary, the profiles of the user's most similar friends to guide the user searches.

Situational Information: Other works exploit situational information (GPS coordinates, location name, location type etc.) to improve the search results relevance. Akermi et al. (2015), propose a personalized recommendation approach in a mobile environment guided by a situational user profile. This profile is based on the user's locations and preferences, and aims to identify the right information to recommend depending on the user current location. Bouidguaguen (2011) proposes an approach of the research personalization based on the construction of a situational user profile, composed of the different search locations and the attached interest centers. In (Bila et al., 2008), the authors propose a mobile user profile in which the user's locations are represented by the most visited places, and the relations between them are represented as a tree. The user interests are explicitly acquired using a quiz.

Combination of Social and Situational Information: Halfway between these two precedent types, some approaches combine the relevance factors from the user's two social and situational contexts, in order to better represent the user profile. In our previous works (Rafa T. & Kechid S. 2016), we presented a personalized search approach based on a geo-social user profile that combines relevant information issued from the thematic, social and situational aspects of the user's previous searches. Aneja and Gambhir (2014) propose a geo-social profile for a user of an ad-hoc social network. They build a profile for each search location by combining the browsing history and location information, and provide a profile matching algorithm to infer the user's dynamic interests. Bao et al. (2012) present an approach for a personalized recommendation of locations. This approach is based on a user profile that combines the user's preferences and social annotations on the most visited places, to recommend places that meet the user interests.

Events and Temporal Information: Differently of these precedent works, some approaches of personalized information retrieval are interested by the temporal and event aspect of the user interests. In these approaches, temporal information is considered as relevance sensor that can help to infer the user's information needs. In (Bouidguaguen, 2011), the temporal information taken into account by the user profile is that concerning the search time (i.e. the query sent time). The authors of (Vandenbussche and Teissèdre, 2011) propose an information retrieval approach which combines the temporal information on the document (creation, update…) and that of its content. In (Kanhabua and Norvag, 2012), is proposed a semantictemporal approach for learning the extraction from text of events and named entities. The work presented in (Basile et al., 2014), proposes a system for the extraction and indexing of events and the retrieving of their related subjects from texts.

2.2. On Using Semantics in Personalized Information Retrieval

To improve the representation of the user's dynamic interests, some PIR approaches introduce the semantics in the user profile. These approaches exploit external semantic resources (ontologies, thesaurus, dictionaries, knowledge bases, lexical bases ...) for an automatic understanding of the meanings conveyed by the terms of user query, the terms of the targeted documents, and the terms of the user's search history.

Soha and Abdelmoty (2017) propose a spatio-semantic user profile in a social network based on localizations. This profile is constructed by combining information about the different locations visited by the user, and his social annotations used to annotate these locations or to annotate his activities performed in these locations. The authors use the "Foursquare" location database to extract the semantic information from the locations (name, category, region ...), and the "*WordNet*" ontology to filter and categorize the user's annotations. Duong et al. (2013) present a semantic personalized search approach based on the ODP ontology (open directory project). The concepts of ODP ontology are used for semantic indexing of documents and user profile to create a semantic search space. Daoud et al. (2011), in their approach of semantic personalization propose to model the user profile and the documents in the form of a graph of concepts. They use the concepts of ODP ontology.

Differently of these previous approaches where the semantic links are extracted only from external semantic resources, the idea proposed by Aimé et al. (2010) is somewhat close to our idea of personal semantics. They propose a semantic approach for the enrichment of user query using personalized domain ontology. They introduce the notion of "*over-concepts*" which represent the consensual interpretations of a concept given by individuals of a specific domain. These interpretations may be different to those of ontology, but they reflect the personal perception of the domain individuals, and they can help to build a personalized ontology that can be used for enriching the user query.

3. Event Modeling

The temporal context considered in this paper is represented by the events most important for the user. For this, we need firstly to highlight what an event means.

3.1. Event Definition

The notion of *Event* does not lead to a strict, precise and consensual definition (Arnulphy, 2012). There is very wide variety of event definitions relative to different fields (history, philosophy, linguistics, computer science ...). Nevertheless, they contain a common characteristic of event "*anchored in time*" (Ludovic et al., 2013).

For the present paper, we adopt the definition proposed in (Arnulphy, 2012): "*the events are finite spatiotemporal* *entities*". And this proposed in (Serrano et al., 2012): "*an event E is the combination of three components E(S, SP, I): a semantic property S, a temporal interval I, and a spatial entity SP*".

3.2. Events Classification

The event models proposed by the different event extraction approaches allow extracting from texts any event without regarding to its importance for the user. As we are interested only by important events for the user and his close community, we will classify the event in only two classes:

Past Event: Concerns past events having a periodicity of reproduction (for example, the football world cup), or those having a periodicity of celebration (for example, the world war).

Current Event: Concerns an event that has recently passed and still holds the community's opinion, an event in progress (arrived and not yet completed), or an event that will come very soon (planned) but it attracts attention now.

3.3 Proposed Event Model

For us, an event (figure 1) may be described by the 3 following information:

(i) *Event Identifier*: Can be represented by a proper name ("*Nobel Prize*" for example), named entity ("*football world cup*" for example), or even by a sentence ("*the launch of the Algerian geostationary satellite alcomsat-1*" for example).

(ii) *Event Date*: May be in a full or partial form of date, a period or a time.

(iii) *Event Location*: Can be represented by the location name, the city or the country.

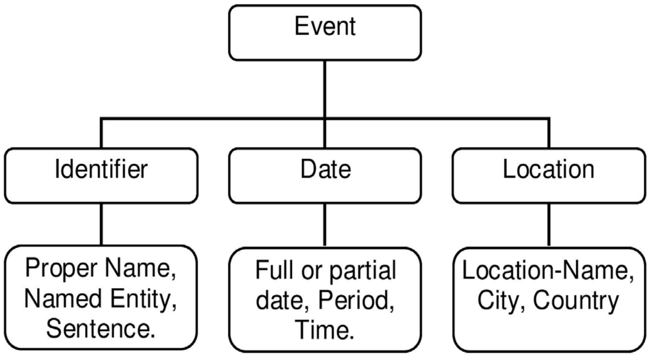

Event
Identifier
Date
Location
Proper Name, Named Entity, Sentence.
Full or partial date, Period, Time.
Location-Name, City, Country

Figure 1. The Event Model

To extract these event information from the text, we can use the temporal markup language "TimeML" (Pustejovsky et al., 2003), particularly the attributes: <EVENT> which allows identifying the event identifier and <TIMEX3> which allows extracting and normalizing the temporal expressions related to the event (date or time).

4. Proposed User Profile

The user profile that we propose combines the relevance sensors from the 4 different thematic, social, situational and event contexts of the user previous searches. To improve the user interests' representation, we reinforce the user profile by introducing a personal semantics based on cooccurrence links between its different components.

The user profile is composed of a base of research situations used to store the relevant search history according to the search locations, and a semantic representation space used to filter and classify the relevant terms of search history according to different user's search contexts qualified as importance providers for these terms.

4.1. Base of Research Situations

The base of research situations *(B_RS)* is used to save the user relevant search history. It groups the set of research situations related to the different locations from where the user has made information searches.

A research situation (RS) represents the relevant history related to the set of queries sent from the same location. Each RS consists, as shown in (Figure 2), of 4 components:

(i) Situational Component:
It includes data about the user location. One location is represented by its parts: (location name, location type, city and country). Location data can be derived by input the user's GPS coordinates in a geographic database.

Research situation (RS)
<u>**Location**</u>: Nam_loc: .... Typ_loc: .... Cty_loc: .... Ctry_loc: ....
<u>**Relevant Documents:**</u> D1, D2, D3, ...
<u>**Relevant Terms:**</u> (t1, wt1), (t2, wt2), ...
<u>**Frequent Tags:**</u> (g1,wg1), (g2, wg2), ...
<u>**Important Past Events:**</u> (pev1, wpev1), (pev2, wpev2), ...

Figure 2. The Research Situation Model

(ii) Thematic Component:
It is used to store the relevant search history of the concerned RS. This history is represented by the set of user's relevant documents (long visited, printed, shared, annotated, marked, etc.), and the set of the most weighted terms extracted from these documents.

The weight of a term t in the search history of a given RS, noted $W(t)$, is calculated using the $tf*idf$ formula of the vector model (Salton and Buckley, 1988) as following:

$$W(t)=(0.5 + 0.5 * tf/Max\ tf) * \log(N/n) \quad (1)$$

With:

- *tf*: Total frequency of term *t* in the RS’s documents.
- *Max tf*: The highest frequency of terms in RS’s documents.
- *N*: Number of documents in the RS’s history.
- *n*: Number of documents containing the term *t* in the RS’s history.

(iii) Social Component:
It is represented by the set of K’ tags most used by the user to annotate the RS’s relevant documents.

The weight of a tag *g,* noted *W(g)*, is also calculated using the *tf * idf* formula of the vector model as following:

$$W(g)=(0.5 + 0.5 * gf / Max\ gf) * \log (N/n) \quad (2)$$

With:

- *gf*: total frequency of tag g in the RS’s documents.
- *Max gf:* the highest total frequency of the tags in the RS’s

documents.

- *N*: number of RS’s documents.
- *n*: number of RS’s documents annotated by the tag *g*.

(iv) Temporal Component:
It is represented by a set of the most important events extracted from the RS’s relevant documents. These events are considered as past events, while the current events are captured from the profiles of the user friends as explained below.

The importance of a past event is expressed by its frequent presence in many documents of the user search history. This importance, is represented by a weight noted *W(p_ev)* and calculated as follow:

$$W (p\_ev) = (|d (p\_ev)| / |d|) * (|RS (p\_ev)| / | RS|) \quad (3)$$

With:

- $| d (p\_ev)|$: Number of documents in *B_RS* containing the past event *p_ev*.
- $| d |$: Total number of documents in *B_RS*.
- $| RS (p\_ev) |$: Number of RSs in B_RS in which at least one document contain *p_ev*.
- $| RS |$: Total number of *RSs* in *B_RS*.

Only a set of top past events is retained in each *RS*.

4.2. Semantic Representation Space of the Relevant Search History

As we have already explained, we introduce in the user profile a personal semantics which is interested by the identification and the valorization of the latent semantics in the user search history.

Concerning this personal semantics, we believe that the following 4 facts are meaning carriers: (i) the joint appearance of two relevant terms in many documents, (ii) annotate by the same tag several documents containing the same term, (iii) the frequent occurrence of a relevant term in many documents in the same location, and (iv) the frequent co-occurrence of a relevant term and an important event in many documents.

This personal semantics may not exist in ontologies, and so our approach aims to identify and valorize this personal semantics through the representation of different co-occurrence relationships between: relevant terms themselves, terms and the most used tags in the user’s social annotations, terms and most important events, as well as terms and the different locations from where the user has made information searches.

In this phase of semantic representation we proceed to classify the relevant terms of the search history into a set of homogeneous classes. Each class serves to store the weighted semantic links between relevant terms and a specific user context. Our goal through the classification by contexts of relevant terms is to facilitate the selection of the most appropriate profile part to guide the new searches according to the current user context. For example, if the user changes his interests for a given period, the new relevant terms will appear in the search history, and their importance increases according to their frequency, their semantic links are calculated and stored in the different bases of the semantic space. These new relevant terms and their linked tags, locations or events may be used to guide the user’s new searches that concern the same subject.

The space of semantic representation that we propose (Figure 3) gathers 5 bases (classes) as is presented bellow.

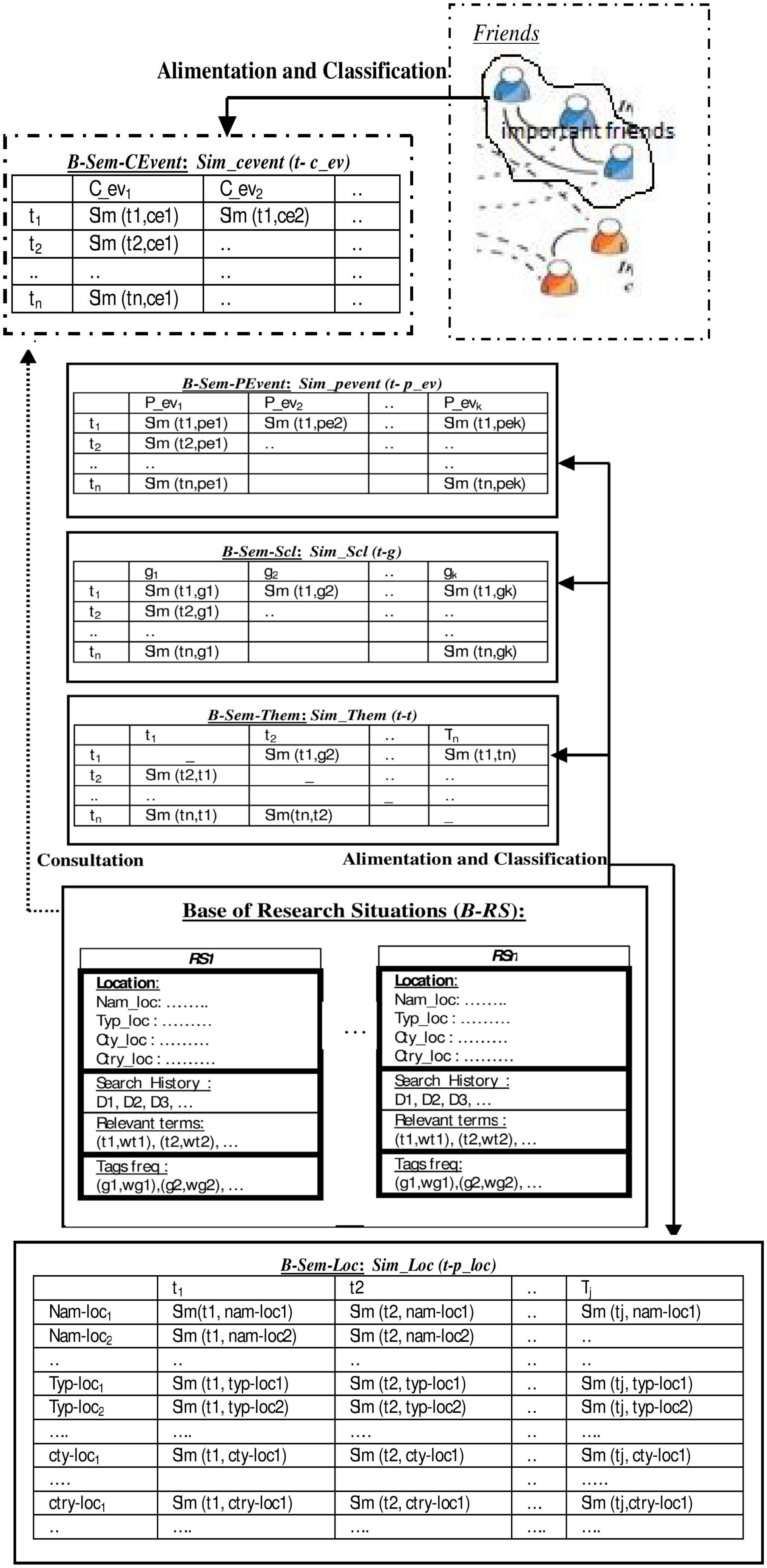
4.2.1. Base of the Thematic Semantics (term-term link)

The thematic semantics is represented by a link connecting each relevant term t_i to each one of other relevant terms t_j of the user profile history. We consider the frequent cooccurrence relationship between terms as semantic link. The thematic semantics between two relevant terms is saved in a base called (*B_Sem_Them*), and is expressed by a similarity noted $sim\_them\ (t_i, t_j)$ and calculated as follows:

$$Sim\_them\ (t_i,t_j)= |d(t_i,t_j)| / [|d(t_i)|^2 + |d(t_j)-d(t_i, t_j)|^2] \quad (4)$$

With:

- $| d (t_i, t_j) |$: Number of documents containing t_i and t_j.


Friends
Alimentation and Classification
important friends
Consultation
Alimentation and Classification
...

***B-Sem-CEvent*: Sim_cevent (t- c_ev)**

	C_ev$_1$	C_ev$_2$	..
t_1	Sim (t1,ce1)	Sim (t1,ce2)	..
t_2	Sim (t2,ce1)	..	..
..	..	..	..
t_n	Sim (tn,ce1)	..	..

***B-Sem-PEvent*: Sim_pevent (t- p_ev)**

	P_ev$_1$	P_ev$_2$	..	P_ev$_k$
t_1	Sim (t1,pe1)	Sim (t1,pe2)	..	Sim (t1,pek)
t_2	Sim (t2,pe1)	..	..	..
..	..			..
t_n	Sim (tn,pe1)			Sim (tn,pek)

***B-Sem-Scl*: Sim_Scl (t-g)**

	g_1	g_2	..	g_k
t_1	Sim (t1,g1)	Sim (t1,g2)	..	Sim (t1,gk)
t_2	Sim (t2,g1)	..	..	..
..	..			..
t_n	Sim (tn,g1)			Sim (tn,gk)

***B-Sem-Them*: Sim_Them (t-t)**

	t_1	t_2	..	T_n
t_1	_	Sim (t1,g2)	..	Sim (t1,tn)
t_2	Sim (t2,t1)	_	..	..
..	..		_	..
t_n	Sim (tn,t1)	Sim(tn,t2)		_

Base of Research Situations (*B-RS*):

RS1

Location:
Nam_loc: ……..
Typ_loc : ………
Cty_loc : ………
Ctry_loc : ………

Search History :
D1, D2, D3, …

Relevant terms:
(t1,wt1), (t2,wt2), …

Tags freq :
(g1,wg1),(g2,wg2), …

…

RSn

Location:
Nam_loc: ……..
Typ_loc : ………
Cty_loc : ………
Ctry_loc : ………

Search History :
D1, D2, D3, …

Relevant terms :
(t1,wt1), (t2,wt2), …

Tags freq:
(g1,wg1),(g2,wg2), …

***B-Sem-Loc*: Sim_Loc (t-p_loc)**

	t_1	t2	..	T_j
Nam-loc$_1$	Sim(t1, nam-loc1)	Sim (t2, nam-loc1)	..	Sim (tj, nam-loc1)
Nam-loc$_2$	Sim (t1, nam-loc2)	Sim (t2, nam-loc2)	..	..
..	..	..	..	..
Typ-loc$_1$	Sim (t1, typ-loc1)	Sim (t2, typ-loc1)	..	Sim (tj, typ-loc1)
Typ-loc$_2$	Sim (t1, typ-loc2)	Sim (t2, typ-loc2)	..	Sim (tj, typ-loc2)
….	….	….	..	….
cty-loc$_1$	Sim (t1, cty-loc1)	Sim (t2, cty-loc1)	..	Sim (tj, cty-loc1)
….			..	…..
ctry-loc$_1$	Sim (t1, ctry-loc1)	Sim (t2, ctry-loc1)	…	Sim (tj,ctry-loc1)
..	….	….	….	….

Figure 3. Semantic Representation Space of the Search History

- | $d(t_i)$ |: Number of documents containing t_i.
- | $d(t_j) - d(t_i, t_j)$ |: Number of documents containing t_j and not containing t_i.

This thematic similarity is not bidirectional, ($sim\_them(t_i, t_j) \neq sim\_them(t_j, t_i)$). Indeed, it is possible that a term '*x*' appears in the majority of the documents containing the term '*y*', while '*y*' may appears only in a small part of the documents containing '*x*'. This reflects the fact that one term may be larger than another.

The formula (4) of thematic similarity favors, as terms more linked to a given term t_i, the terms satisfying the 3 criteria: (i) appearing in the maximum of documents containing t_i, (ii) absent in the minimum of documents containing t_i, and (iii) appearing in the minimum of documents not containing t_i.

4.2.2. Base of the Social Semantics (term-tag link)

The social semantics represent the link connecting each relevant term t_i of the user profile with each one of tags g_j most used by the user to annotate the documents containing t_i. We consider the frequent co-occurrence relationship between one term and one tag as semantic link. This social semantics between a term and a tag is stored in a base called (*B_Sem_Scl*), and is expressed by a similarity noted $sim\_scl(t_i, g_j)$, and is calculated as follows:

$$sim\_scl(t_i, g_j) = |d(t_i, g_j)| \,/\, [|d(t_i)|^2 + |d(g_j) - d(t_i, g_j)|^2] \quad (5)$$

With:

- | $d(t_i, g_j)$ |: Number of documents containing t_i and annotated by the tag g_j.
- | $d(t_i)$ |: Number of documents containing t_i.
- | $d(g_j) - d(t_i, g_j)$ |: Number of documents annotated by g_j and not containing t_i.

The formula (5) of social similarity favors, as tags more linked to a given term t_i, the tags satisfying the 3 criteria: (i) used to annotate the maximum of documents containing t_i, (ii) not used to annotate the minimum of documents containing t_i, and (i) iii) used to annotate the minimum of documents not containing t_i.

4.2.3. Base of the Location-Linked Semantics (Termlocation Link)

The Location-linked semantics represent the link between the relevant terms and the different locations from where the user has made information searches. We consider the frequent occurrence of one term in many documents returned in a same location as semantic link.

The Location-linked semantics is stored in a base called (*B_Sem_Loc*). We compute for each relevant term of the user profile, its connection degree with each one of the location parts among (name, type, city and country) of all the locations of the *B_RS*. So, for each relevant term t_i and each location part p_loc_j, we propose the similarity formula noted *Sim-loc* (t_i, p_loc_j), and calculated as following:

$$Sim\_Loc(t_i, p\_loc_j) = link(t_i, p\_loc_j) * Coef(dist\_Loc) \quad (6)$$

Where:

$$Link(t_i, p\_loc_j) = |RS(t_i, p\_loc_j)| \,/\, [|RS(t_i)|^2 + |RS(p\_loc_j) - RS(t_i, p\_loc_j)|]$$

And:

$$Coef(dist\_Loc) = |dist\_Loc\_RS| \,/\, |RS|^2$$

With:

- | $RS(t_i, p\_loc_j)$ |: Number of *RS* in which t_i is relevant and the location includes the concerned p_loc_j among (name, type, city or country of location).
- | $RS(t_i)$ |: Number of *RS* containing t_i.
- | RS |: Total number of RS in *B_RS*.
- | $RS(p\_loc_j) - RS(t_i, p\_loc_j)$ |: Number of the concerned p_loc_j in which t_i does not appear.
- | $dist\_Loc\_RS$ |: Total number of distinct locations in the B_RS.
- *Coef (dist_Loc):* Coefficient of distinct locations in user profile.

We notice that the formula (6) proposed to calculate the location-linked similarity, give importance to the number of distinct locations in the user profile, because if the user does not change place, we cannot judge that the relevance of a given term is related to the user location.

This formula favors, as terms more linked to a given part of location p_loc_j among (name, type, city and country of location), the terms satisfying the 3 criteria: (i) appearing in the maximum of *RS* whose location includes p_loc_j, (ii) absent in the minimum of *RS* whose location includes p_loc_j, and (iii) appearing in the minimum of RS whose location does not include p_loc_j.

4.2.4. Base of the Semantics Linked to Past Events (Term- past event Link)

The semantics linked to past events represent the link connecting each relevant term t_i of the user profile with each one of the important past events p_ev_j extracted from documents of the user search history. We consider the frequent co-occurrence relationship between one term and one important past event as semantic link.

For a given relevant term t_i, we calculate its link degree with all the past events. This link degree is represented by a similarity, noted $sim\_pevent\ (t_i, p\_ev_j)$, and calculated as follows:

$$Sim\_pevent\ (t_i,\ p\_ev_j)= |d(t_i,p\_ev_j)|\ /\ [|d(t_i)|^2 + |d(p\_ev_j)\text{-}d(t_i,\ p\_ev_j)|^2] \qquad (7)$$

With:

- $|\ d\ (t_i, p\_ev_j)\ |$: Number of documents containing t_i and the past event p_ev_j.
- $|\ d\ (t_i)\ |$: Number of documents containing t_i.
- $|\ d\ (g_j)\ \text{-}d\ (t_i,\ p\_ev_j)\ |$: Number of documents containing p_ev_j and not containing t_i.

The formula (7) favors, as past event more linked to a given term t_i, the events satisfying the 3 criteria: (i) appearing in the maximum of documents containing t_i, (ii) absent in the minimum of documents containing t_i, and (iii) appearing in the minimum of documents not containing t_i.

4.2.5. Base of the Semantics Linked to current Events (Term- current event Link)

New events may not appear in the user's search history. Nevertheless, the user can be interested by these new events since they have interested his friends on the social website. To help the user about current events, we add to his profile a base, called (*B_Sem_CEvent*), to store the most recent events and their attached terms extracted from the last search history of his most important friends.

For a given user U_i, The importance of another user U_j (friend) is based on the number of their common relevant past events relative to the number of relevant past events of U_i. The hypothesis here is that if they share the same interests regarding past events, then very likely that they will be interested by the same recent events.

This importance is represented by a similarity noted *Sim (ui, uj)*, and calculated as fellow:

$$Sim\ (Ui,\ Uj) = |B\_Sem\_PEvent_i) \cap B\_Sem\_PEvent_j||/\ |B\_Sem\_PEvent_i| \qquad (8)$$

Daily, a top-list of the user's most important friends is established, and the content of the base *B_Sem_CEvent* of the user profile is updated to containing the most recent event figured in the base *B_Sem_PEvent* of the profile of each friend. These inserted events are considered as current events that may be important for the user. To each inserted event in *B_Sem_CEvent* is associated its linked terms extracted from the profile of the concerned friend.

The semantic link between a term and an event is obtained by the combination of:

- Their semantic similarity stored in the base *B_Sem_PEvent* of the concerned friend's profile calculated withe the same formula (7).
- The importance of the friend himself according to the user (formula 8).

This semantic link is represented by a similarity noted $Sim\_cevent\ (t_i,\ c\_ev_j)$ and calculated as follow:

$$Sim\_cevent\ (t_i,\ c\_ev_j)= sim\_pevent_j(t_i,\ c\_ev_j)*\ sim(U_i,U_j) \ (9)$$

Where:

- $sim\_pevent_j\ (t_i,\ c\_ev_j)$: Is the similarity between t_i and *c_evj* stored in the base *B_Sem_PEvent* of the profile of concerned friend U_j calculated using the same formula (7).
- $sim(U_i,U_j)$: The importance of the friend U_j from its profile the event c_ev_j is obtained.

This formula (9) favors the most important events of the most important friends.

5. The Proposed Search Process

Our search process (Figure 4) follows the following scenario:

- The system receives the user query Q, and retrieves the current user location.
- From the 5 profile bases (*B_Sem_Them, B_Sem_Scl, B_Sem_Loc, B_Sem_PEvent* and *B_Sem_CEvent*), the most similar to the user query is selected as current context.
- The query *Q* is reformulated (enriched) according to the selected base. A reformulated query *Q'* is produced and sent to search.
- The system observes the reactions and the interactions of the user with the final returned results, to update his profile (identify and classify by contexts the new relevant terms, locations, tags, friends and events).

In the following subsections, we present the details of each step of our search process.

5.1. Selection of the Most Appropriate Profile Base for the User Query

To select from among the 5 profile's semantic bases (*B_Sem_Them, B_Sem_Scl, B_Sem_Loc, B_Sem_PEvent and B_Sem_CEvent*) the most appropriate for the user query *Q*, our system calculates a similarity score between the user query *Q* and each of these bases respectively as follows:

$$Sim\ (B\_Sem\_Them,Q)= \Sigma_{ti\in Q}\Sigma_{tj\in Q\ B\_Sem\_Them}\ sim\_them(t_i,t_j) \ (10)$$

$$Sim(B\_Sem\_Scl,Q)= \Sigma ti\in Q\Sigma gj\in\ B\_Sem\_Scl\ sim\_Scl(ti,gj) \qquad (11)$$

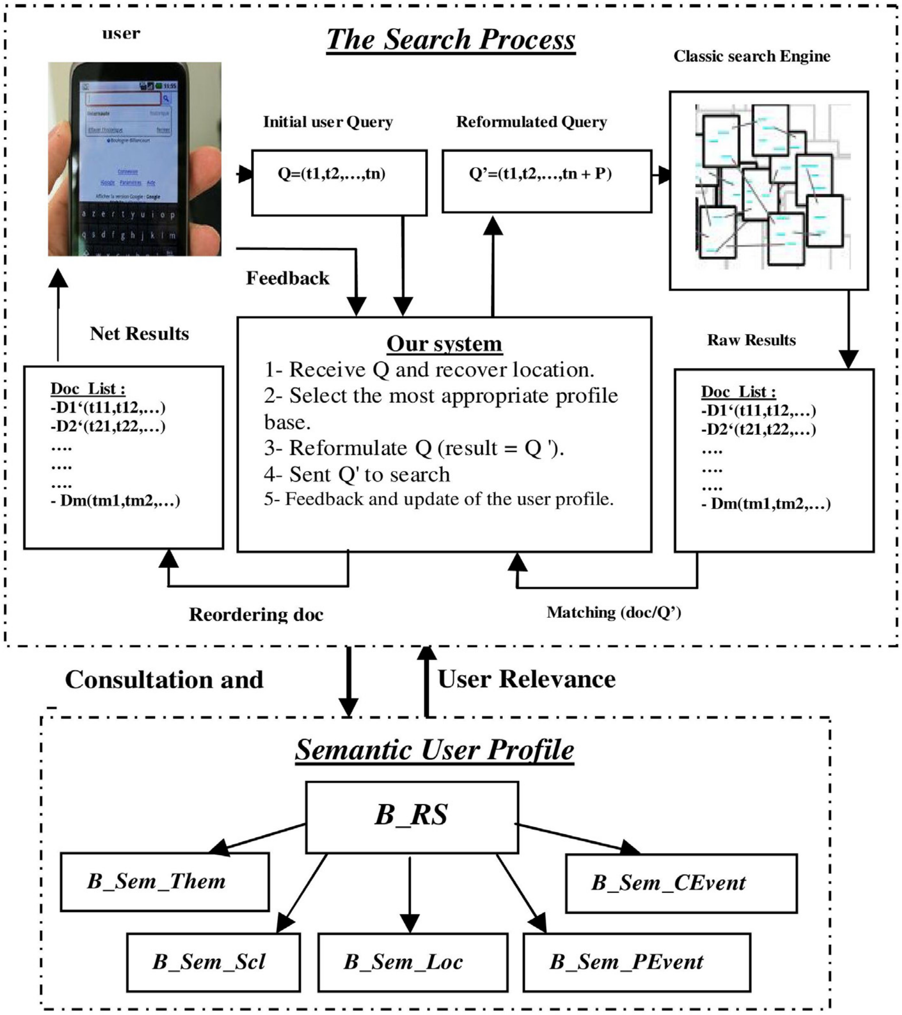


Figure 4. The Search Process Model

$$Sim(B\_Sem\_Loc, Q)= \sum_{ti\in Q}\sum_{locj\in current\ loc} sim\_Loc(t_i, loc_j) \quad (12)$$

$$Sim(B\_Sem\_PEvent, Q)= \sum_{ti\in Q}\sum_{p\_evj\in B\_Sem\_PEvent} Sim\_pevent\ (t_i, p\_ev_j) \quad (13)$$

$$Sim\ (B\_Sem\_CEvent, Q)= \sum_{ti\in Q}\sum_{c\_evj\in B\_Sem\_CEvent} sim\_cevent(t_i, c\_ev_j) \quad (14)$$

We retain as user's current search context, the base that maximize its similarity to the user query.

5.2. User Query Reformulation

The initial user query is reformulated to select more relevant results by widening the search field without going further than the user interests (Baziz, 2005). Most query reformulation techniques used in PIR approaches are based on adding terms issued from the user profile (Daoud et al., 2011). With respect to this principle, we reformulate the initial user query Q by the injection of some, terms, tags, events or location's information according to the selected profile base.

Thus, the new reformulated query Q' is obtained from the initial query Q as follows:

$$Q' = Q + P \quad (15)$$

Where:

- In case of B_Sem_Them: P is the term that maximizes the similarity (4) with the terms of Q.
- In case of B_Sem_Scl: P is the tag that maximizes the similarity (5) with the terms of Q.
- In case of B_Sem_Loc: P is the location part (name, type, city or country) that maximizes the similarity (6) with the terms of Q.
- In case of B_Sem_PEvent: P is the past event that maximizes the similarity (7) with the terms of Q.
- In case of B_Sem_CEvent: P is the current event that maximizes the similarity (9) with the terms of Q.

5.3. Updating of the User Profile

Concerning the update of the user profile, we must ensure on one side, its enrichment and its evolution over time by the addition of new research situations (RS) in the B_RS base, and the classification of their relevant terms in the different bases of the semantic representation space, and on the other side, we must optimize the space and the time occupation, by removing the old *RSs*.

5.3.1. Alimentation of user Profile

When the final results are transmitted to the user, our system observes the user's interaction with these results to identify the new relevant documents. The relevant terms are extracted and weighted using the formula (1), the tags used to annotate these documents are weighted using the formula (2) and also the most important past events are extracted using formula (3). A new research situation *RS* is instantiated with its 4 components (situational, thematic, social and temporal). Then, the new most weighted terms are inserted in the *B_Sem_Them* base, the new most weighted tags are inserted in the *B_Sem_Scl* base, the parts of the RS's location are inserted in the *B_Sem_Loc* base and the past events are also inserted in the *B_Sem_PEvent* base.

Exceptionally, the current events are selected among the most recent events of *B_Sem_PEvent* of the most important friends using the formulas (8 and 9).

After the enrichment of the 5 bases, the co-occurrencebased semantic links between terms themselves, terms and tags, terms and location parts, terms and past events, and between terms and current events are recalculated using respectively the formulas (4, 5, 6, 7 and 9).

5.3.2. Cleansing of user Profile

The formulas (4, 5, 6, 7 and 9) proposed for measuring the different semantic similarities are defined as a proportion of the number of *RSs* containing the concerned semantic relationship relative to the total number of *RSs* in the *B_RS*. Thus, these formulas allow growth and decay of these similarities.

For example, if a term appears frequently with a tag in many documents in the user last searches, the semantic link between this term and this tag will be reinforced, and even, if a term and a tag do not appear together for a long time their semantic link is diminished.

The evolutionary nature of the formulas quoted above allows us to adopt a very simple strategy to lighten the profile bases. This strategy consists of removing any *RS* that its intersection with all other *RSs*, in *B_RS* base, is zero for one year. So, the terms extracted only from a deleted *RS* are also deleted, and by transitivity, tags, locations and events associated only to a deleted term will be also deleted.

6. Evaluation of the Proposed Approach

6.1.The Challenge of the Evaluation in a PIR Context

According to (Toms et al., 2013; Saoud & Kechid, 2016), the relevance to be improved in the context of personalized information retrieval is very subjective, because this relevance depends particularly on the user. This user is the unique able to judge this relevance according to his viewpoint. This reality makes the evaluation of a personalized information retrieval approach a very problematic task.

An evaluation framework for a PIR approach must provide the following essential components (Bouhini et al., 2016): on one side, the user profile built using collected user data and the search history of their previous searches, and on the other side, a dataset (set of documents) and a set of new user queries associated with their user-centered relevance judgments used for testing the efficiency of the proposed search process.

Nevertheless, in the absence of an evaluation framework dedicated to a personalization context as already indicated and as is mentioned in several works, the authors of PIR approaches are forced to build their own evaluation frameworks.

6.2. Evaluation Strategy

Like several PIR approaches (Bouhini et al., 2016; Saoud & Kechid, 2016; Duong et al., 2013; Daoud et al., 2011; Bouidguaguen 2011; ...), we construct our proper evaluation framework as following:

User Profile Construction: We built 30 profiles for 30 different users from different professional domains to have different interests. We built each profile from the relevant documents of the user's previous searches. The user provides with these documents their locations and some tags used to annotate these documents. These documents are processed to extract relevant terms and relevant events.

Information Needs and Queries: We asked each user to define 5 new information needs, and to formulate 5 queries to express them. We insist, as much as possible, that the information needs must be specific to the user and that the queries that express them must be general and ambiguous as is the majority of user queries in the real world. We also took care to express some different information needs of different users with a same large sense query. In sum, we have 100 different queries to sent from 20 different locations (some queries and some locations are common).

Dataset and Search: With the absence of consensual dataset dedicated to the PIR context, we adopted the solution of dynamic dataset used by several PIR approaches. Then, we have used *Google* as search engine, and for each sent query we retain only first 20 returned documents. This allows us to construct a dataset of 3000 documents for testing the improvement made by our approach.

Relevance Judgments: To have a real feedback, we invite the same users, owner of used profiles, to manually examine and judge the relevance of the returned results.

Each user examines the documents returned for each query, and determines the relevant documents for him according to his information needs already defined. This operation of usercentered relevance post-judgment is repeated with each of the 4 search scenarios explained below.

Evaluation of Performances: The evaluation of the improvement of our proposed solution consists of calculating the two precision factors:

- *Precision @ x*: (*x* = 5, 10 and 15 returned documents) that represent the rate of relevant documents in the first 3 pages of results that can be visited by the user in the average.

- *R-precision*: for this measure, we supposed 3 values considered as total number of relevant documents for all the sent queries (*R* = 1, 2 and 3 relevant documents), and we calculated the *R-precision* according these 3 different points. Our goal is to test the capacity of our search process to rank relevant documents in top positions.

Baselines Approaches and Comparison of Performances: To compare the contribution of our approach with respect to non-semantic personalized search approaches, to non-personalized semantic approaches, and to personalized-semantic approaches, we will calculate the precision according to the following 4 scenarios: (A) query reformulated only by the location data without semantics according to the location-based personalized search approach proposed in (Bouidguaguen, 2011), (B) query reformulated using the concepts derived from the WordNet ontology without personalization according to the semantic search approach proposed in (Boubekeur and Azzoug, 2013),(C) the user query is reformulated with combination of semantics and personalization according to the approach proposed in (T.H. Duong et al. 2013), and (D) query reformulated using our profile.

6.3. Results and Discussion

6.3.1. Precision @ x Measure

The following table (table 1) presents the statistics of precision measure of search results obtained according to the 4 scenarios already explained.

	Scen (A)	Scen (B)	Scen (C)	Scen (D)
P @5	0.275	0.295	0.330	0.362
P @ 0	0.292	0.322	0.351	0.375
P @ 15	0.31	0.34	0.367	0.392
P-Avg	0.292	0.319	0.349	0.376

Table 1. Precision @x Results

The 4 columns of Table 1 (Scen A, Scen B, Scen C, Scen D) represent the average precisions for all the queries sent by all users according to one of the 4 search scenarios already explained, while the rows represent the average precision obtained at the different reference points (5, 10 and 15 documents). The last row represents the average precision obtained by each scenario.

We note that the results of the different scenarios are closes, but the use of our semantic personalized user profile has improved the precision factor relative other approaches. We note also that the results according to scenario (A) are less efficient than other scenarios because very few profiles are related to location.

In (Table 2) we present the improvement made by our solution relative to each other approach.

	D-A	D-B	D-C	Avg/P@x
P@5	0,087	0,067	0,032	0,062
P@10	0,083	0,053	0,024	0,053
P@15	0,082	0,052	0,025	0,053
P-Avg/scen	0,084	0,0573	0,027	0,0561

Table 2. Precision @ x Improvements

Each row of (Table 2) represents the improvements obtained with respect to one reference point among (5, 10 and 15 documents), and the last row represents the average improvement of our solution relative to the other 3 scenarios.

We note that the precision measure with our search process is better than other approaches, particularly the precision at the level of the top 5 documents (6.2 %) that are most likely to be examined by the user.

As overall remark, the global average improvement obtained by our search process is less than 10 percent (5.61 %). This is under than our expectations, but we find it reasonable considering the evaluation conditions that have been available to us.

6.3.2. R-Precision Measure

In the following table (Table 3), we present the statistics obtained for the R-precision factor calculated at R = 1, 2 and 3 first relevant documents returned by the different 4 search scenarios. The objective of R-precision measure is to show the capacity of our search process to rank the first R (1, 2 and 3) relevant documents better than other scenarios.

The different columns of Table 3 represent the precision values obtained respectively by each search scenario relative to 3 reference points (R = 1, 2 and 3 first relevant documents). While the rows represent the precision

	Scen (A)	**Scen (B)**	**Scen (C)**	**Scen (D)**
R=1	0,275	0,288	0,412	0,52
R=2	0,331	0,345	0,375	0,431
R=3	0,311	0,331	0,368	0,409
Avg R-pr	**0,305**	**0,321**	**0,385**	**0,453**

Table 3. R-Precision Results

obtained by the different scenarios respectively at the 1st, 2nd and 3rd returned relevant documents. The last row represents the average R-precision of the 4 scenarios.

We note from Table 3 that:

- The R-precision values obtained by our search process (scenario D) are better than other scenarios.
- The different scenarios provided close values of average R-precision.

The following table (Table 4) presents the improvements provided by the application of our solution (scenario D) relative to other scenarios concerning the R-precision measure.

	D-A	**D-B**	**D-C**	**Avg-Impr**
R=1	0,245	0,232	0,108	0,195
R=2	0,1 0,	086 0,	056	0,080
R=3	0,098	0,078	0,041	0,072
Avg-Impr	**0,1476**	**0,1320**	**0,0683**	**0,1160**

Table 4. R-Precision Improvements

We note from Table 4 that:

- Our proposed solution is more able than other scenarios to rank relevant documents in first.

- The different improvements made by our solution (scenario D) regarding other scenarios in a descending rank are as follow: D-A, D-B and D-C.

- The best improvement obtained by our solution (average R-precision = 19.50 %) is that concern the rank of the 1st relevant document. This improvement is also very important for the user.

As overall remark, the global average improvement obtained by the application of our search process concerning the R-precision measure exceeds 10 percent (11.60 %), which is a reasonable result for us.

The details of the evaluation phase show that our results are good in the case where the request is too short and the profile is very rich. And our results are not very interesting in the case of a new profile.

It should be noted that the use of an evaluation framework larger than ours in terms of the number of users, the size of the profiles, the number of queries and locations, may give results that better reflect the reality.

7. Conclusion and Perspectives

In this paper we have proposed a new semantic-based approach for a personalized information retrieval using a user profile. In this user profile, on one hand, we combine the user's interest indicators from different thematic, social, situational, and temporal contexts of the user's previous searches, and on the other hand, we present with a personal semantic vision the content of the user search history by valorizing the different co-occurrence relationships linking the relevant components (terms, tags, events and locations) of this user search history.

This approach is proposed in objective of improving the search results relevance by using the user profile to reformulate the initial user query to improve the search results relevance.

The evaluation of the proposed approach shows that more the user profile is richer, the user query is better reformulated and the search results are more relevant.

This work opens way to other research tracks. We plan to: (i) study the stability and the evolution over time of the user profile, to estimate how much this profile can be useful to infer the user's future information needs, and (ii) implement an online evaluation framework dedicated to the personalized information retrieval, with user accounts and their profiles, test collections, predefined queries with user's relevance judgments etc.

References

[1] Aimé, X., Fürst, Frédéric., Kuntz, Pascale., Trichet, Francky. (2010). Enrichissement sémantique de requêtes au moyen d'ontologies de domaine personnalisées. *In:* Actes de l'atelier Personnalisation du Web, 10ième Journées francophones d'Extraction et de Gestion de Connaissances (EGC'2010). Hammamet, Tunisie.

[2] Akermi, I., M. Boughanem, Faiz, R. (2015). Une approche de recommandation proactive dans un environnement mobile, *In:* INFORSID'15, France. p. 301-316.

[3] Aneja, N., Gambhir, S. (2014). Geo-Social Profile Matching Algorithm for Dynamic Interests in Ad-Hoc Social Network. *Social Networking,* 3, p. 240 - 247. http://dx.doi.org/10.4236/sn.2014.35029

[4] Anmol, G. P., Shete Devakinandan, S., Kahate, S. A. (2014). Personalized web search with user's profile in

reranking, *International Journal of Science, Engineering and Technology (IJSET)* 2 (8).

[5] Arnulphy Beatrice, (2012). Désignations nominales des événements : étude et extraction automatique dans les textes. *Phd Thesis. Univ Paris Sud - Paris XI. France.*

[6] Bao, J., Zheng, Yu., Mohamed, F., Mokbel. (2012). Location-based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data. *In:* ACM SIGSPATIAL GIS'12, Redondo Beach, CA, USA.

[7] Pierpaolo, Basile., Caputo, Annalina., Semeraro, Giovanni., Siciliani, Lucia. (2014). Extending an Information Retrieval System through Time Event Extraction, *In:* 8th International Workshop on Information Filtering and Retrieval, DART, Co-located with XIII AI*IA Symposium on Artificial Intelligence, AI*IA 2014 pp.36-47 CEUR Workshop Proceedings.

[8] Mustapha, Baziz. (2005). Indexation conceptuelle guidee par ontologie pour la recherche d'information, *Phd thesis, Univ Toulouse, France.*

[9] Bila, N., Cao, J., Dinoff, R., Ho, T., Hull, R., Kumar, B., Santos, P. (2008). Mobile User Profile Acquisition through Network Observables and Explicit User Queries, *In:* 9th Int'l conference on Mobile Data Management, p. 98-107.

[10] Fatiha, Boubekeur., Wassila, Azzoug. (2013). Conceptbased indexing in text information Retrieval, *International Journal of Computer Science & Information Technology* (IJCSIT), 5 (1) DOI: 10.5121/ijcsit.2013.5110. *p. 119-136.*

[11] Chahrazed, Bouhini., Géry, Mathias., Largeron, Christine. (2016). Personalized information retrieval models integrating the user's profile. *In:* 10th International Conference on Research Challenges in Information Science (IEEE RCIS'16), Grenoble, France. p.1-9, DOI: 10.1109/RCIS.2016.7549310.

[12] Ourdia, Buidguaguen. (2011). Accès contextuel à l'information dans un environnement mobile : approche basée sur l'utilisation d'un profil situationnel de l'utilisateur et d'un profil de localisation des requêtes. *Phd thesis, Univ Toulouse III - Paul Sabatier, France.*

[13] Cole, Charles. (2011). A Theory of Information Need for Information Retrieval That Connects Information to Knowledge, *Journal of the American Society for Information Science and Technology*, 62 (7)1216–1231, 2011, DOI: 10.1002/asi.21541.

[14] Mariam, Daoud., Tamine, Lynda., Boughanem, Mohand. (2011). A personalized search using a semantic distance measure in a graph-based ranking model, *Journal of Information Science (JIS)* 37(6) 614–636 DOI: 10.1177/0165551511420220.

[15] Duong Trong Hai, Mohammed Nazim Uddin., Cuong Duc Nguyen, (2013). Personalized Semantic Search Using ODP: A Study Case in Academic Domain. *In: ICCSA'13,* Part V, LNCS 7975 Springer, p. 607–619.

[16] Fouad K. M., Ahmed, R., Khalifa., Nagdy, M., Nagdy, Hany, M., Harb. (2012). Web-based Semantic and Personalized Information Retrieval, *IJCSI International Journal of Computer Science Issues,* 9 (3) 266-276.

[17] Jansen, B.J., Rieh, S. Y. (2010). The seventeen theoretical constructs of information search and information retrieval. *Journal of the American Society for Information Science and Technology*, 61 (8) 1517–1534.

[18] Jeon, H., Kim, T., Choi, J. (2010). Personalized information retrieval by using adaptive user profiling and collaborative filtering. *Advances in Information Sciences and Service Sciences,* 2 (4) 134–142.

[19] Kanhabua, N., Nørvag, K. (2012) "Learning to Rank Search Results for Time sensitive Queries. *In:* Proceedings of the 21st ACM International Conference on Information and Knowledge Management. p. 2463– 2466. CIKM '12.

[20] Ludovic Jean-Louis, Romaric Besançon and Olivier Ferret. (2013). Une méthode d'extraction d'information fondée sur les graphes pour le remplissage de formulaires. *In:* TAL. 54 (1) 139-170.

[21] Lin, C., Xie, R., Guan, X., Li, L., Li, T. (2014). Personalized news recommendation via implicit social experts, *Information Sciences.* 254, p.1–18.

[22] Nazim-Uddin Mohammed, Trong Hai Duong, Visal Sean, and Geun-Sik Jo, (2012). Construction of Semantic User Profile for Personalized Web Search*, In:* ICCCI'12, Part II, LNAI 7654 Springer, p. 99–108.

[23] Pustejovsky J., Castaño J., Ingria R., Saurí R., Gaizauskas R., Setzer A., Katz G., (2003). TimeML : Robust Specification of Event and Temporal Expressions in Text, *In:* IWCS-5, Fifth International Workshop on Computational Semantics., Tilburg University.

[24] Tahar, Rafa., Samir, Kechid. (2016). A geo-social user profile for a personalized information retrieval, *In:* ACMICIME' 16 Istanbul Turkey. p. 62-66, DOI: http://dx.doi.org/10.1145/3012258.3012270.

[25] Salton, G., Buckley, C. (1988). Term Weighting Approaches in Automatic Text Retrival, *Information Processing and Management.* p. 513-523.

[26] Zakaria, Saoud., Samir, Kechid. (2016). Integrating social profile to improve the source selection and the result merging process in distributed information retrieval. *Information Sciences 336,* p.115–128. http://dx.doi.org/10.1016/j.ins.2015.12.012.

[27] Serrano, Laurie., Charnois, Thierry., Brunessaux, Stephan., Grilheres, Bruno., Bouzid, Maroua. (2012). Combinaison d'approches pour l'extraction automatique d'évènements, *In:* Actes de la conférence conjointe JEPTALN- RECITAL'12, volume 2: TALN, p.423– 430, Grenoble, France.

[28] Mohamed, Soha.,Alia, I., Abdelmoty. (2017). Spatiosemantic user profiles in location-based social networks. *International Journal of Data Science and Analytics.* 4, p.127–142, DOI 10.1007/s41060-017-0059-9.

[29] Elaine, Toms., Agosti, Maristella., Fuhr, Norbert., Vakkari, Pertti. (2013). Evaluation Methodologies in Information Retrieval. *In:* Agstuhl Reports, 3 (10) 92–126.

[30] Vandenbussche, P. Y., Teissèdre, C., (2011). Events Retrieval Using Enhanced SemanticWeb Knowledge. *In:* Workshop DeRIVE 2011, in conjunction with 10th International Semantic Web Conference (ISWC'11).

[31] Zargayouna Haïfa, Catherine Roussey, Jean-Pierre Chevallet, (2015). Recherche d'information sémantique : état des lieux. *TAL.* 56 (3)*. 49-73.*