Dynamic Enrichment of Social Users' Interests

Manel Mezghani, Corinne Amel Zayani, Ikram Amous, André Péninou, Florence Sèdes

Dynamic Enrichment of Social Users' Interests

Manel Mezghani*†, Corinne Amel Zayani*, Ikram Amous*, André Péninou† and Florence Sèdes†
* Sfax University, MIRACL Laboratory, Sfax, Tunisia
mezghani.manel@gmail.com, {corinne.zayani, ikram.amous}@isecs.rnu.tn
† Paul Sabatier University, IRIT Laboratory, Toulouse, France
{peninou, sedes}@irit.fr

***Abstract*—In a social context, the user is more and more an active contributor for producing social information. Then, he needs a tailored information reflecting his current needs and interests in every period of time. This aims to provide a better adaptation while accessing the information space by integrating users' interests dynamic. Indeed, users' interests may change and become "outdated" through time. So, an interest judged as relevant in a period of time may fluctuate in the next period of time. Moreover, analysing the classic user behaviour to deduce his current interests is a difficult task. In fact, his behaviour isn't always reflecting his real interests.**
In this paper, we propose a new approach for enriching the user profile in an evolutionary environment such as a social network. The enrichment takes into account: i) the social behaviour and more precisely the tagging behaviour (that reflects user's interests) and ii) the temporal information (that reflects the dynamic evolution of users' interests). Our approach focus on the concept of *temperature* that reflects the importance of a resource in each period of time. This concept is used to infer common interests of users tagging the same "important" resource. The originality of our approach relies on combining information tags, users and resources in a way that guarantees a better enrichment for the social user profile. Our approach has been tested and evaluated with the *Delicious* social database and shows interesting precision values.

Keywords*—*User profile; enrichment; social network; adaptation; tags; resources.

I. Introduction

Adaptation systems require an accurate representation of the user profile. This profile represents the user at a given time. It can be used in many purposes such as recommendation, customization, etc. and several applications such as detecting suspicious users (spammers, hackers, etc.) whose objectives is to harm other users, etc.
In our team, researches aim to adapt the information in a classical context e.g. [35] and in a social context e.g. [32]. To achieve a reliable adaptation, these studies have shown that the user profile is the main and essential entity. However, adaptation is performed on a set of static data and not through the evolution of these data.
In the social context, networks such as *Facebook*, *Flickr*, *Delicious*, etc. are characterized by their dense activity. In these changing environments, the user is more and more an active contributor for producing social information through: blogs, discussions, tags, etc. Then, he requires tailored information reflecting his current needs and interests in every period of time. The motivation behind this is to be up to date on what's happening in social networks. This could provide a better adaptation to the user while accessing the information space and while his interests are evolving. In practice this could be useful in online recommender system context, where the user needs relevant information to be recommended to him in every period of time.
However, analysing the classical user's behaviour has two main drawbacks. First, the user's classical behaviour isn't always reflecting his real interests. e.g. analysing the navigation behaviour of the user according to [20]: i) leads to analysing the anterior behaviour which may not reflect his current interests, ii) is not always an efficient indicator since the user may access a web page without having an interest on it. Second, the user's interests may change and become "outdated" through time [34].

To overcome these problems, we propose a new approach for enriching the user profile in an evolutionary environment (the social network). The enrichment takes into account (See figure 1): i) the tagging behaviour which is used to capture/infer the user's interests [12][17] and ii) the temporal information (which reflects the dynamic evolution of the users' interests). The variability of the interest implies its modulation over time. This modulation could be managed through the concept of "Temperature" proposed by [21]. The analysis of the temporal evolution is not widely used in social enrichment context. The enrichment of the user profile aims to add the interests judged pertinent in every period of time.

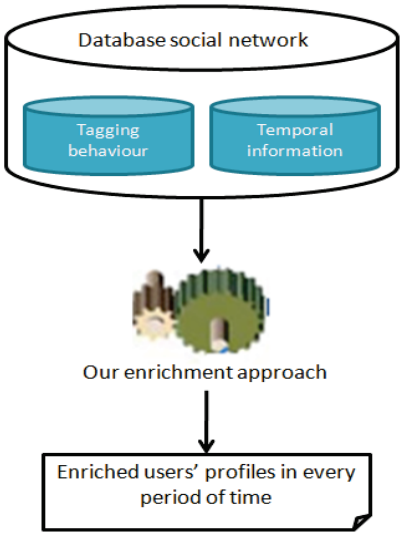

Fig. 1. An overview of the information used for our enrichment approach

In the rest of this paper, we first present some existing works on the user profile enrichment and some works that exploit the social behaviour (tagging behaviour) for an enrichment purpose. Second, we present our approach for enriching the user profile. Third, we experiment and validate our method. Finally, we conclude and discuss some future works.

II. Related Works

We present in this section, some existing works for updating (enriching and filtering) the user profile. Next, we present some works that exploit the social behaviour through the tag, the user and the resource and the benefit of this information in our context. Finally, we present a synthesis that shows the difference between our approach and the state of the art approaches.

A. Techniques For Updating The User Profile

In an environment characterized by constant change and density as the social web, a user profile cannot be considered as stable. The profile evolves as the data evolve. According to [10] the management of the evolution of the user profile is a complementary process to build a user profile and its adaptation refers to changes in the user's interests over time. The evolution of the profile is dependent to the time constraint. To manage this change, Daoud et al., [10] distinguish the short-term profile (representing the interests of current research sessions) and long-term profile (representing the persistent interests deduced from entire research history). Zayani et al., [35] distinguish permanent features that represent the personal data that are generally stable over time, and changing characteristics such as interests and preferences. The evolution is taken into account through: (a) the consideration of several variables of interest, (b) the importance of the variables of interest, (c) the existence of several interests for the same variable interest, (d) the importance of interest for a given variable of interest and finally (e) the existence of several evolutionary characteristics.

Updating the user profile can be done through learning techniques [29]. The basic principle is to study the behaviour of the user and the classification of its characteristics or researched objects. The advantage of this approach is the accuracy of the derived data. The disadvantage is the complexity of algorithms that are greedy in terms of time. Examples of such techniques are neural networks, classification methods (case-based reasoning, Bayesian classifiers, etc.), association rules, etc. [29].
Treating the evolution of user profile can be done by: i) enriching the profile with new information from different detection techniques of interests / preferences, or ii) simplifying (eliminating) information that is deemed irrelevant and its value decreases over time. Each of these techniques is detailed below:

- The enrichment of the user profile can be used to improve the quality of recommendation systems and therefore provide most appropriate information to the user. Enrichment is a technique that adds information to the user profile after a predefined treatment. In the work of [35] updating the user profile is only for changing attributes, such as the interests and preferences. Mechanisms have been used to increment the value of interest according to their frequency of use. The update of interest, do not extract the relationships between the user query and the condition included in this request. In addition, this technique does not take into account the semantic relationships between queries from the same user and/or similar interests.
  In a social context, user profile could be tag-based. A tag is defined as a keyword generated by the user himself. Researches on updating a tag-based profile are already studied in our previous works [22]. The enrichment is used in a recommendation and cross-system context [22]. For [1] the enrichment process is to consider the semantic of tags and the connection between users and also between *tweets* and articles. Kim et al., [17] enrich a user model with collaboration from other similar users. Beldjoudi et al., [3] enrich user profiles with relevant resources based on association rules extracted from social relationships.
- Simplification is to remove information deemed irrelevant to a given user. It reduces the amount of information contained in a profile to facilitate the treatment. It also allows the filtering of the old data that is no longer reflecting the needs of the user. It can be carried to the various attributes of the user profile. Interests are the attributes that vary the most in a profile in our context. Rebai et al., [29] use a learning technique to overcome the problem of the user profile overloading through a filtering approach in a distributed context.

B. Exploiting Social Information For Updating The User Profile

In this section, we focus on the three social information: user, tag and resources that constitute the tagging behaviour. We detail the researches done based on each element.

1) Exploiting the metadata of the resources: The evolution of digital documents has led to a classification of these documents into three categories [2]: The non structured documents (flat document), the structured documents (documents with an explicit structure defined and known a priori) and the semi-structured documents (documents with a flexible structure and an heterogeneous content). We focus on analysing semi-structured documents and more precisely their metadata. The metadata can provide a comprehensive information which may be used in retrieving or interpreting the data [2].
The resources on social networks are a powerful information which reflects the user's interests. In fact, the tagged resources reflects the interest of the user to the resource [11]. Also, the resources may be rated and this reflects the degree of interest of the user according to these resources [17]. The resources contain metadata which describes their content. The metadata could be used in an adaptation context such as recommendation [4] [36] [16], enrichment of the user profile [1], enrichment of the metadata [21] adaptation of the documents in a data warehouse context [2], etc. We detail each of these researches below.
Bogers et al., [4] use the metadata for a recommendation purpose. This approach uses the folksonomy and the item

metadata to boost the performance of traditional collaborative filtering algorithms.

Zitouni et al., [36] propose a method for recommendation of resource in E-Learning context. They use the metadata to recommend the new resources. First, they extract metadata from new resources. Then, they compare the new resource with the preferred collection of the user. In case of similarity, they send a notification to the user.

Joly et al., [16] propose a method of filtering and recommending resources. The proposed approach aggregates and interpret the context of the data on the user terminals in the form of weighted keywords (tags). They calculate the weight of the tag from the metadata of the web page according to the number of occurrences of each term in the title, in the keywords and in the text description.

Similar to our context we find Abel et al., [1] who exploit defined metadata (title, author, date of publication) to enrich the profile. The metadata are used to connect the *tweets* to the article news. The most related tweets are used to enrich the user profile.

Also, Manzat et al., [21] exploit the user's behaviour in order to enrich metadata. This enrichment is exploited for adaptation of the presentation. The metadata of the resources are weighted according to the user usage. This research introduces the concept of the *temperature*. The *temperature* corresponds to a metadata of usage, which reflects the popularity of a document or an element of metadata at a given time. The temperature of a metadata for a certain user group, at a certain moment, translated the interest of this group for the part of the document described by the metadata. If the resource is not consumed in a period of time, the weight of metadata decreases. The originality of this approach is that the metadata are always kept even if the weight is equal to zero. This is beneficial in the case of re-appearing of the resource, the calculation of the weight is easier.

Amous [2] proposes a data warehouse of documents in order to provide a local structure of the documents and also to organise and classify data. This approach aims to adapt dynamically the documents according to the user's needs.

2) Exploiting the tags: The word tag is becoming popular in social network like *Flickr* and *Delicious* [14]. A tag may be defined through different manners according to the web site or the social network. For example a tag on *Flickr* is associated with the photos uploaded by the user or by other users. Also a tag in *Delicious* is associated to the bookmarks shared by the users.

According to [14], several motivations are behind the use of a tag, such as: to contribute and share, to mark places for possible future research, to attract attention, to express its opinions, etc. According to [12], the use of a tag implicitly denotes the user's interests and is used to infer conceptual information about the user. Moreover, analysing the tag of the user is a powerful "knowledge management tool".

A tag is a way which allows to leave traces in the resources (photos, text, video, etc.). Thus, a tag may be defined as a social annotation. The action of annotating a resource by a specific user is called tagging behaviour. Many users can annotate the same resource by means of several tags which lead to collaborative tagging systems [11]. The collaborative tagging behaviour, leads to creating a folksonomy introduced by [25]. Unlike the ontology, this folksonomy is not structured.

Also, a tag may be a way to find information about the user according to his history of tagging [14]. In fact, [28] analyses the tagging behaviour of the users in order to understand their interests, preferences, etc. This could be useful to recommend information or to facilitate the access to the information. Also, [17] detect users' interests from the tagging behaviour of the users in order to recommend resources. This approach enriches the user profile from the neighbours' tags, based on the hypothesis that the user prefers the similar tags issued from his neighbours. So, the enrichment process is done according to the similar tags of the neighbours not present in the current user profile. This approach has proved the utility of the collaborative knowledge (of the neighbours) to improve the recommendation quality.

Cantador et al., [6] enrich the user profile with tags from the tagging history in order to improve recommender system. This approach associates the tags with ontologies in order to incorporate the tag which matches the concept of the ontology. This approach uses different sources of the tagging behaviour history extracted from popular social networking sites.

De Meo et al., [11] enrich the user profile with tags considered as important (for example tags having a high PageRank). This approach is graph-based, where we find two graphs: the tag resource graph called TRG and the tag user graph called TUG. These graphs are used in order to filter qualitative tags (e.g. funny, good, etc.), then generate a list of candidate tags by means of the IDDS (Iterative Deeping Depth First Search), and finally merging these candidate lists of tags by the Borda count technique. Finally, the user profile is enriched by tags from these candidate lists. This method has shown that the tags are automatically filtered and ranked at the same time through the Borda count technique. However, it does not consider the semantics of tags and the context of the user in the recommendation process. There is a risk of having no precise information through the Borda count technique even if it's simple to use and fast.

3) Exploiting the user (neighbours): The neighbour is the social relationship of the user with other users. This relation could be explicit (e.g. friend relationship) or implicit (e.g. users who interact with the same resource). This social relationship has been recently and well detailed in Musial et al., [27]. The neighbour of the user in social context is described through ties: "*a tie between two users aggregates all types of the relations that exist between these two persons*" [27].

Some studies analyse the social relationships in order to detect neighbours (users considered close to the user, in term of interests). Neighbours are detected by several metrics such as cosine similarity [17], "X-compass" [11], etc. Other studies detect neighbours through observations like Kim et al., [17], which enrich the user profile with tags of user's friends not included in the user profile based on the observation that two people who share common tags are considered close and may well have interests in common. Also Zhao et al., [33] assume that two users are similar if they share a large number of tags that are strongly related.

Other researchers have tried to combine different parameters in order to detect the similarity between users. In bibliographic domain, Cabanac et al., [5] calculate the similarity between authors by analysing their proximity, their connectivity and the number of paper in common. Guy et al., [15] calculate the score of proximity through different criteria: i) more people and/or tags within the user profile related to the item, ii) the stronger relationships of these people and/or tags to the user, iii) the stronger relationships of these people and/or tags to the item, and iv) the freshness to the item. Roth et al., [30] detect the implicit relationships between users through their mail exchange. They calculate the proximity through the frequency of interaction between users, the freshness of the interaction and the direction of the interaction.

Liang et al., [18] consider the social relation prediction as a link problem, where different techniques may be used such as neighbor-based methods (e.g. Common Neighbors and Adamic-Adar) which are inexpensive in computation compared to path-based methods. In a graph-based context and more precisely in FOAF ontology, social connections are described through the element $< foaf : knows >$ which describes user's friends. But there is no specification of the nature of this relationship. The neighbours are also detected in graph-based context, where Tchuente et al., [32] analyse the egocentric networks to derive relevant users profiles.

C. Synthesis

From the state of the art detailed above, we compare our approach with the researches done in the context of enriching the social user profile. The difference could be summarized as two main points.

First, we noticed that unlike [30] (in a context of finding relationships through the mail exchange), [15] (in a context of social recommendation) and [21] (in a context of enriching multimedia metadata), no researches are done to enrich the user profile in a social context with consideration of the temporal aspect. In fact, the enrichment is done according to the analysis of the data in a specific period and not through the evolution of the user's interests. So, we will consider in our approach the temporal aspect for enriching the social user profile.

Second, the researchers which enrich the user profile focus on : i) the information of the tag e.g. [11], which may be ambiguous, or on ii) analysing the neighbours e.g. [6], which could be spammers or do not provide the needed information or on iii) analysing the metadata of the resources e.g. [1], which may not have associated metadata in the case of multimedia resources. Also, the researches which combine the tag information and the neighbours e.g. [17], may overcome the associated shortcoming of the tags but they do not consider the document analysis. In fact, analysing common tagged resources may reflect common interests described differently through tags. So, we will use these three information: tag, neighbours and resources in order to take advantage from these valuable data to enrich the social user profile.

Regarding the temporal aspect, updating the user profile is an important criterion to follow the changing needs of users. Updating classic user's interests is to increase the weight of interests which are more attractive to the user and to decrease those which do not. In this way, an interest may disappear from the user profile if its weight is equal to 0. This technique is interesting in the way that we will not have a lot of information to process and so we can avoid the computation time issue. However, there may be an interest that reappears and so we will need to detect it once again. In fact, in a social network context, there are so-called *Buzz* which is a technique to make noise around an event. This technique engender that several users will be interested in this event at a time *t*, but a *Buzz* is temporary and may disappears at a time *t+1* and therefore users will no longer be interested. However, it may reappear and become an interest in time *t+2*. Based on this suppositions, we can assume that interest even if its low weight can be interesting for some time. This variation leads us to introduce a variable that we call "temperature", which will follow the variation of the popularity of interest over time. This concept is already used in [21], which defines the temperature as a variable associated to each descriptor metadata indicating the popularity of multimedia metadata of a resource.

In order, to take advantage of the information provided by the social users, our approach could be summarized through these points:

- We adopt the concept of the temperature and we associated it to the resource in order to reflect its importance at each period of time.
- We investigate in the elements of the tagging behaviour relation (tag, user and resource) to calculate the temperature of the resource.
- We consider a neighbour of a given user as the user tagging the same resources in the same period of time. This choice is based on the hypothesis that people tagging the same resource have similar interests.
- We focus on semi-structured resources and their associated metadata since they provide a comprehensive information about the resource content. In fact, Mezghani et al., [24] have demonstrated that the more the tag describes the content of the resource the more the tag reflects really the user's interests.
- We attribute a weight to the tag that reflects its correspondence to the associated resource. This weight aims to filter the insignificant tags and to avoid the ambiguity associated with these social annotations. Then, to enrich the user profile with comprehensible interests.

To summarize, the enrichment approach analyses the tagging behaviour of each user in order to detect the most significant interests to enrich the user profile. Also, the enrichment approach consider the relevance of the tag to the associated resource in order to try to remove the ambiguity associated to these social annotations. Then we enrich the user profile with comprehensible tags. The enrichment is based on temporal constraint that gives more importance to the recent and popular tags.

III. The Enrichment Approach

In this section, we present and develop the module associated with the architecture proposed in [23]. Then, we propose our approach of enriching the user profile. The approach uses the social information such as the tag information, the neighbours and the metadata of the tagged resources. We show how this information provides a solution for capturing the user's interests over time and then contribute for a better enrichment.

A. Architecture Of Social Adaptation

User profile enrichment is a part of the *user modelling* module and more precisely the *update* sub module extracted from the architecture of adaptation of social navigation proposed in [23] (see figure 2). Adaptation is reached through a technique of recommendation, which needs pertinent information about users' interests over time.

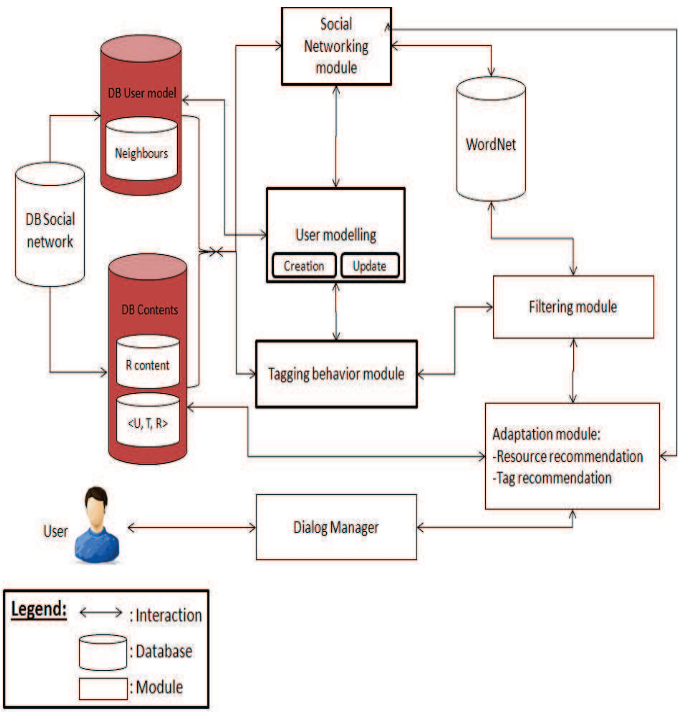

Fig. 2. Architecture of social adaptation

We develop our approach by using the databases: the *DB social network*, the *DB user model* and the *DB content model*. The used databases are explained hereafter:
The DB social network contains information about the objects in the social network including the information about the resources and the users. The data exploited in this database are extracted from a specific social network (e.g. *Delicious*, *CiteUlike*, *Last.fm*, *movieLens*, etc). The social information is adapted depending on the social network (e.g. bookmarks in *Delicious*, scientific articles in *CiteUlike*, music in *Last.fm* and video in *movieLens*).
The DB user model uses information from the *DB social network*. This module specifies information about users and networks of users (interests, preferences, friends, professional relationships, etc.).
The DB Contents uses also information from the DB social network. This module stores information about the resources of the social network (type of resource, tags associated by each users, metadata, etc.).

We also use the two modules: the *social networking module* and the *tagging behaviour module* since they are essential to detect user's interests. The used modules are explained as follows:
Social networking module exploits the user modelling by analysing the similarity between users to build networks of similar users (using same tags) and accesses the users' profiles to build networks of friends. This module is able to identify similar users with a similar tagging behaviour. Based on social relations, it is able to send information such as most popular users, friends, etc. for the adaptation module. So, from the social networking module, we extract users neighbours.
Tagging behaviour module contains information about the users who tags the resources of various types (e.g. photos, videos, scientific papers, etc.). Generally this activity is represented in a tripartite model which describes the users $U = \{u_1, \ldots, u_n\}$, the resources being tagged $R = \{r_1, \ldots, r_m\}$ and the tags $T = \{t_1, \ldots, t_h\}$, Where n is the number of users, m is the number of resources h and is the number of tags. A tagging behaviour is a triplet of the form:

$$Tagging\_relation :< U, T, R > \tag{1}$$

So, from this module we extract the tagging behaviour information. This behaviour should be associated to time information. This latter, allows us to follow the user's tagging behaviour in each period of time.

So, the enrichment approach is developed through these databases and modules. The *user modelling module* is the module that will use the information provided by the other modules to achieve the enrichment process. (the development the *update* sub module). We have already developed in previous work [24] the *creation* sub-module.

B. Description Of The Enrichment Approach

In this section, we detail our approach for enriching users profiles. This approach is part of the *user modelling module* and more precisely the *update* sub-module. The dynamic evolution of the user profile is treated by enriching users' interests with tags deemed relevant for each period of time. In fact, in social environment, the user consults the documents stored in the network, communicates and interacts with other users to find the information he needs. Enrichment in this context is done by analysing the environment of the user to detect relevant interests (relevant tags). The relevance of an interest is usually calculated from the frequency of use of the tag at a given time. Frequency varies periodically. This change has already been treated by [21], through the concept of "temperature". This notion is interesting since it models the popularity of a term over time.
In order to explain the used data, we present our model of the social user profile inspired from [32] in figure 3.

The social user is described through his profile (*Profile* class). A user may have neighbours (neighbours association) described also through their profiles. Each profile contains attributes (*Attribute* class) which are static attributes (which never change, e.g. name) and dynamic attributes (which change through time, e.g. interests). These latter are the tags (*Tag* class) applied by the user on a resource (*Resource* class) characterised by its metadata suach as title, keywords and description (*Metadata* class). The *EgoCentric Network* class

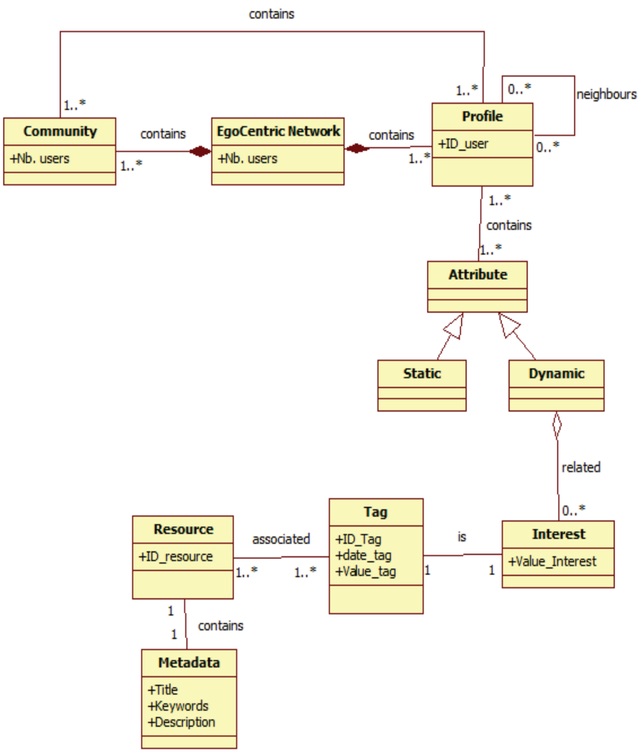

Fig. 3. The social user profile model

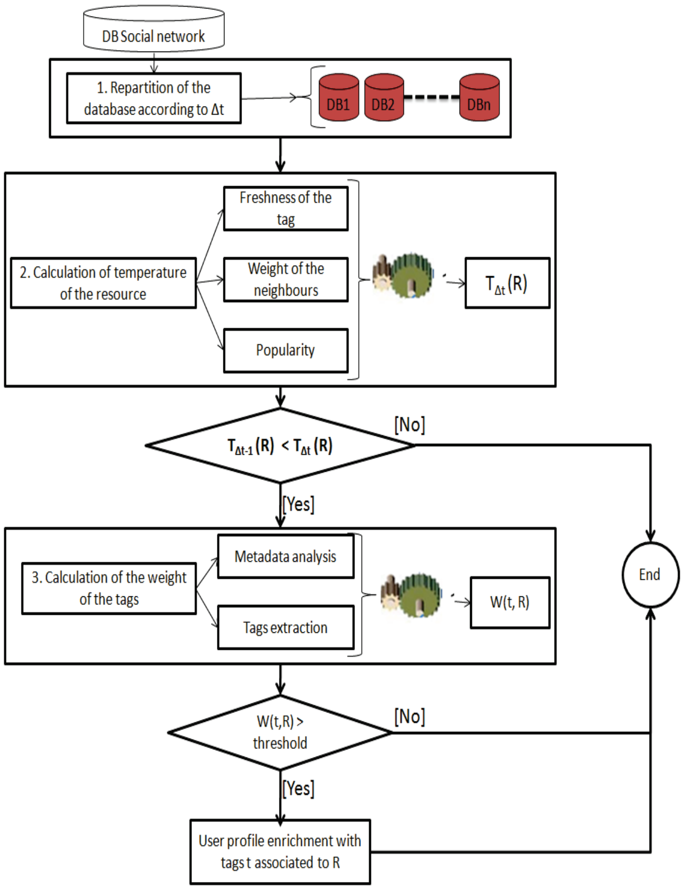

Fig. 4. The process of enriching the social user profile for a Δt.

contains the users connected explicitly through friend relationship. The *Community* class contains the users having same behaviour or same interests.
The user profile is constructed in an implicit way, using the list of tags assigned by the user. The user profile is enriched with tags (considered as his interests) in each period of time (Δt) in order to reflect the current interests of the user. This enrichment allows us to capture current user's interests in Δt and then we could used it in further purposes to provide temporal-based recommendation for example.

The process of updating the user profile through an enrichment approach is detailed in Figure 4.

The first step, consists in dividing the database in each period Δt. The choice of the Δt is important in our context. In fact, this period allows us to detect the evolution of the user's interests between two successive periods. This period should be coherent with the quantity of data contained in the social network. The value of the period Δt will be fixed in the experimentation.
The second step, consists in calculating the temperature of each resource at a given Δt. The concept of the temperature is inspired from [21]. Unlike [21] who associated it with the metadata in order to enrich the multimedia metadata, we attribute the temperature to the resource in order to figure out the most interesting resources to a user. The temperature of a resource reflects the popularity of this resource in a period of time. In order to calculate this attribute, we propose a formula which takes into consideration several parameters (see formula 2) :

1) the freshness of a tag associated to the resource; in fact, the more the tag is recent the more it is interesting for the user [34].
2) the similarity of the users whose tagged the resource; in fact, if two users have tagged the same resource with similar tags, that reflect their similarity in terms of interests. So, the neighbours in our context, are users sharing common behaviour with a specific user. We choose the cosine similarity, since it's a popular metric, to calculate the similarity between two users.
3) the number of tags associated with the resource. In fact, the popularity of a resource reflects that it is interesting [19].

Also, [19] proved that combining the freshness and the popularity reflect more the real interest rather than considering each parameter alone.
The temperature T of the resource r in a specific Δt is calculated through this formula:

$$T_{\Delta t}(r) = \alpha * \frac{\sum_{i=1}^{n} \frac{1}{p1(tag_i)}}{n} + \beta * \frac{\sum_{u,v=1 u \neq v}^{m} p2(u,v)}{2} + \gamma * p3 \quad (2)$$

where:

- Δt is a defined period of time.

- r is the resource.
- n is the number of tags associated with the resource *r*.
- m is the number of users tagging the resource *r*.
- p1(tag_i) is the freshness of the tag tag_i. Where $tag_i \in \{tag_1, \ldots, tag_n\}$. It's the result of the difference between the time of the tag and the current time. So, the more the difference is bigger, the more the freshness is smaller. That's why we calculate the inverse ($\frac{1}{p1(tag_i)}$).
- p2(u, v) is the weight between two users (u and v) who tagged the same resource. The weight is calculated through the cosine similarity between the vector of tags of each user.
$$sim(u,v) = cos(u,v) = \frac{u \cdot v}{\parallel u \parallel \cdot \parallel v \parallel} \tag{3}$$
We consider that sim(u, v)=sim(v, u). So, we divide by 2.
- p3 is the popularity. Its equal to n.
- α, β, γ are constants. They will be fixed in the experimentation. These constants reflect the degree of the influence of each parameter.

We develop the algorithm for calculating the temperature of the resources in table 1. This algorithm provides a temperature of all resources in a specific period of time Δt.

TABLE I. ALGORITHM FOR CALCULATING THE TEMPERATURE OF THE RESOURCES IN Δt

```
INPUT: ID_resource[]:int[], ID_tag[]:int[], ID_user[]:int[], date_tag:Date, m:int, cosine:float, α:float, β:float, γ:float, Δt: float, system_date:Date, fresh:float.
OUTPUT: T_Δt(R)[] :float[] //Temperature of all resources R
BEGIN
for each ID_resource do
    for each ID_tag do
        // Calculating the popularity
        for each ID_tag ∈ ID_resource do
            nb_tags++;
        end for
        // Calculating the freshness
        fresh=0; //Initialisation
        for int i=0; i<nb_tags; i++ do
            fresh=fresh+(system_date - date_tag);
        end for
        fresh=fresh/nb_tags;
    end for
    // Calculating the cosine similarity between two users
    for int k=0,h=0; k<m, h<m; k++,h++ do
        //m is the number of users
        cosine=cosineSimilarity(ID_user[k], ID_user[h]); // A predefined function
    end for
    float T_Δt(r)= α * fresh + β * cosine + γ * nb_tags; // Temperature for a specific resource r (identified through its ID_resource)
    T_Δt(R)[]=T_Δt(R)[]+T_Δt(r); //store the value of temperature of the resource r in the list of temperatrure of all resources (R)
end for
return T_Δt(R)[]
END
```

The temperature of the resource varies through time. It may increase or decrease for each Δt. We consider that the resource is interesting if its temperature increases.

The third step consists in detecting the resources that their temperature increases over time. After calculating the temperature of each resource, we consider only the resource which their temperature value is increasing between two periods of times. In fact, the increasing of the temperature reflects the interest of the user with this resource. However, in social networks which are characterized by the amount of the resources, we can have a lot of resources which their temperature is increasing and then their treatment can be complex. So, in order to overcome such a problem, we should keep only the most relevant resources to the user. That's why, we analyse the content of the resources and more precisely their metadata (we consider that the resources are semi structured data). In fact, in [24], we have demonstrated that the more the tag describes the content of the resource the more the tag reflects really the user's interests. The metadata have been used in many researches as mentioned before in section B.1. In our work, we use the metadata in order to filter the most relevant tagged resources. We attribute a weight for the tags associated with the resources. This weight is calculated according to the degree of correspondence of the tags with the metadata of the associated resource. From metadata, we consider the title, the keywords and the description, since they are elements that reflect the content of the resource. After extracting the metadata, we calculate the weight of the tags related to the resources which their temperature increases. We use the weight proposed by [16]:

$$\begin{aligned} W(tag,r) = \alpha' * |tag \in Tr| + \beta' * |tag \in Kr| \\ + \gamma' * |tag \in Dr| \end{aligned} \tag{4}$$

This function counts the number of occurrences of each term *tag* in a resource r by applying the coefficients α', β' and γ' depending on the location of the term in the metadata of the page. $|tag \in Tr|$ is the number of occurrences of term *tag* in the title element of the page, $|tag \in Kr|$ is the number of occurrences of term *tag* in the keywords element of the page, $|tag \in Dr|$ is the number of occurrences of term *tag* in the description element of the page. The coefficients are constant and will be fixed in the experimentation. We develop the algorithm for calculating the weight of the tag in table 2. This algorithm provides a weight for each tag associated with a resource which its temperature has increased.

The fourth step consists in enriching the user profile with the tags associated with the resources. After calculating the weight of the tags associated with the most interesting resources, we enrich in this step the user profile with the tags that reflect more the user's interests. In fact the more the tag has a higher weight, the more it reflects the content of the resource and then, the more it reflects the user's interests. So, we choose from the result of the previous algorithm, the tags that are more interesting to the user. A tag is stated as a potential interest if it has a weight > *threshold*. The *threshold* will be fixed in the experiment.
So, in every period of time Δt, we enrich the users' profiles of the users who tagged one of these interesting resources (which its temperature has increased) according to the threshold constraint. We develop the algorithm for enriching the user profile (described in the algorithm as the array $tag\_user_{\Delta t}[]$) in table 3.

From this algorithm, we obtain enriched profiles in every period of time. The enrichment process takes into consideration

TABLE II. ALGORITHM OF CALCULATING THE WEIGHT OF THE TAG

```
INPUT: ID_tag[]:int[], ID_Resource[]:int[], title:String, description:String, keywords:String, α':float, β':float, γ':float, T_Δt-1(r) :float, T_Δt(r) :float, Value_tag :String.
OUTPUT: Weight(ID_tag, ID_Resource)[]:float[] /* the weight of all tags associated to the resource .*/
BEGIN
// Initialisation
int nbOccurrence_description=0;
int nbOccurrence_keywords=0;
int nboccurrence_title=0;
for each ID_Resource do
    for each T_Δt-1(r)[] and T_Δt(r)[] do
        if T_Δt-1(r)[] < T_Δt(r)[i] then
            for each Value_tag do
                for each description do
                    if Value_tag ∃ description then
                        nbOccurrence_description ++ ;
                    end if
                end for
                for each keywords do
                    if Value_tag ∃ keywords then
                        nbOccurrence_keywords ++ ;
                    end if
                end for
                for each title do
                    if Value_tag ∃ title then
                        nbOccurrence_title ++ ;
                    end if
                end for
                float Weight (ID_tag, ID_Resource)= α' * nbOccurrence_title + β' * nbOccurrence_keywords + γ' * nbOccurrence_description;
                Weight (ID_tag, ID_Resource)[]= Weight(ID_tag, ID_Resource)[]+ Weight(ID_tag, ID_Resource); //store the value of the weight of the tag in the list of the weight of all tags
            end for
        end if
    end for
end for
return Weight(ID_tag, ID_Resource)[];
END
```

TABLE III. ALGORITHM OF ENRICHING THE USER PROFILE

```
INPUT: ID_Resource:int, Weight (ID_tag, ID_Resource)[]:float[], tag_user_Δt[]:String[], Value_tag:String
OUTPUT: Enriched_tag_user_Δt[]:String[] // enriched user profile at a given Δt
BEGIN
Enriched_tag_user_Δt[]=tag_user_Δt[]; //Initialisation
for each Δt do
    for each ID_Resource do
        for each ID_user who tagged ID_Resource do
            for each Weight(ID_tag, ID_Resource)[] do
                if Weight(ID_tag, ID_Resource) ≥ threshold then
                    Enriched_tag_user_Δt[]=Enriched_tag_user_Δt[]+Value_tag;
                end if
            end for
        end for
    end for
end for
END
```

the social behaviour of the users tagging the same resources. This technique analyses the social "intelligence" to provide potential interests for each user. The enrichment is not a cumulation of enriched interests of previous enrichment in previous period of time. In fact, the user profile is defined in each Δt (the result of the first step which is the repartition of the database) and defined as $tag\_user_{\Delta t}[]$ in the algorithm. This aspect allows us to analyse better the most interesting tags according to a specific period and not through the whole evolution of the user's interests. In fact, user's interests may evolve and become "outdated" and then irrelevant to analyse. So, our approach focus only on the tags in a specific period of time that describe better current user's interests.

IV. EVALUATION

We have evaluated our approach on the *Delicious* database. The *Delicious* database contains social networking, bookmarking, and tagging information. It provides information about the user's friend relationships and the tagging relation information $< U, T, R >$. The users U are described through their ID (e.g. userID=8). The resources R are described through their ID, title and URL (e.g. 1 IFLA - The official website of the International Federation of Library Associations and Institutions http://www.ifla.org/). The tags T are described through their ID and value (e.g. 1 collection_development). This dataset is extracted from [7]. We present some statistics of the data present in this dataset: 1867 users, 69226 URLs and 53388 tags. Also, the tagging behaviour is provided according to the time. An example of temporal tagging behaviour is shown in table 4.

The value of the period of time Δt, is fixed to 1 day. This choice is due to the dense activity in the social networks and our objective to capture the constant evolution of the user's interests. From the database, we obtain 1645 different periods (Δt). The other values of the constants are fixed as follows:

- $\alpha = \beta = \gamma = 0.5$, this choice aims to assign the same influence of each term of the temperature formula.
- $\alpha' = \beta' = \gamma' = 0.5$, this choice aims to assign the same influence of each term of the weight formula.

A. Precision Analysis

In order to validate the results obtained from the enrichment process, we compare if the enriched tags (found by our approach) exist in the neighbours profile. In fact, the neighbours reflects the user's interests [32]. In our validation, we consider the neighbours as the egocentric network like [32]. The egocentric network is the explicit friend relationship an it is given in the *Delicious* database. The egocentric network is detected according to each Δt (as we have the database according to each Δt). A tag is stated accurate if it exists in the neighbours profile. After enriching the users profile in each Δt, we calculate the precision also in each Δt. We calculate the precision $P_{\Delta t}(u)$ of the user u in a given Δt as the percentage of the relevant tags Rel_T (issued from our approach) from the known tags Ret_T (issued from the tags of the egocentric network). We calculate the precision as follows:

$$P_{\Delta t}(u) = \frac{|Rel\_T|}{|Ret\_T|} \tag{5}$$

The overall precision is calculated from the precision $P_{\Delta t}(u)$ for all users and for all Δt as the formula 6 where m=number of the users and l is the number of Δt.

$$Overall\_Precision = \frac{\sum_{\Delta t=0}^{l} \sum_{u=0}^{m} P_{\Delta t}(u)}{m * l} \tag{6}$$

We test the influence of the threshold of the tags in the precision values through table 5. In this table, we first show the overall precision obtained according to each threshold in the interval of]0..1] with 0.1 step. Then, we show the number of enriched profiles. We notice that the results vary from a user to another. Also, we can found for some users better results

TABLE IV. AN EXAMPLE OF THE TEMPORAL TAGGING BEHAVIOUR

userID	bookmarkID	tagID	day	month	year	hour	minute	second
8	1	1	8	11	2010	23	29	22

TABLE V. THE OVERALL PRECISION ACCORDING TO DIFFERENT THRESHOLD VALUES

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Overall_Precision	0.33	0.39	0.43	0.44	0,58	0,50	0.48	0.42	0.36	0.23
Number of enriched profiles	85	87	87	75	150	47	57	52	49	53

with the threshold 0.5 and for other user better results with the threshold 0.6.

We notice that the precision is higher for the threshold=0.5. We detail the cases of the thresholds 0.5 and 0.6 (since they provide better results) to better understand the precision values. Figure 5, details the result for 20 enriched users chosen randomly.

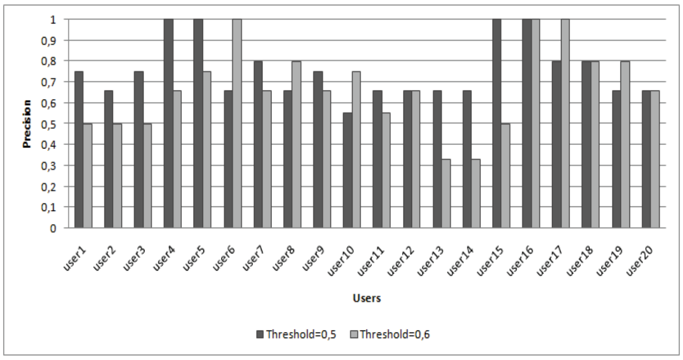

Fig. 5. The precision values of the enrichment process according to a set of 20 users and for the thresholds 0.5 and 0.6.

In order to understand better why we have this variety of results, we calculate the number of neighbours for each user according to the obtained precision. See figure 6.

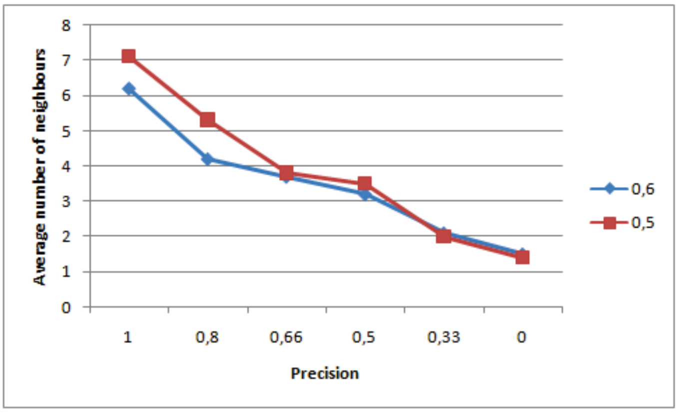

Fig. 6. Comparison between the precision values of the enrichment process and the average number of the neighbours according to the thresholds 0.5 and 0.6.

We notice that the higher values of precision are associated to the active users having a higher number of friends. So, the active users are better entities that reflect their interests through time.

B. Tags Ambiguity Analysis

The motivation of such an analysis is to discover the effectiveness of our approach to provide a comprehensible results. These latter could be used in further works for a recommendation purpose for example.

We analyse the resulting tags and we calculate if the enriched tags are comprehensible or ambiguous. This characteristic is relative to the knowledge of each user. For example the abbreviations (e.g. html5, apps, etc.) or/and commercial products (e.g. facebook, ipad, bbc, etc.) are not always known for all users. For this reason, we choose WordNet (http://wordnet.princeton.edu/) as a natural language processing tool that allows us to know if the tag is comprehensible or ambiguous without taking into consideration abbreviations, commercial products, etc. Such tags, even that they are real words, will not be stated as comprehensible.

Table 6 shows the percentage of ambiguous and comprehensible tags according to each threshold. From this table, we clearly notice that the percentage of comprehensible tags are always higher than the percentage of ambiguous tags. So, our approach enrich the user profile with tags that reflect real interests that could be used and exploited in other purposes such as recommendation. We highlight that the ambiguous tags are almost abbreviations, commercial products etc. that could be considered as comprehensible in other context such as specific social networks (professional, scientific, etc.).

V. CONCLUSION

In this paper, we have detailed techniques for enriching the user profile. We have proposed our approach of enriching the social user profile by analysing the social behaviour and especially by analysing the metadata of the resources, the tags assigned to the resources and the users' neighbours. Moreover, our approach takes into consideration the temporal aspect in order to capture the new information over time.

The combination of the three information, is in our opinion, a powerful and promising approach to provide flexible enrichment in an evolutionary environment.
The enrichment of the profile could be used for further purposes such as recommendation, customization since it provides an information which reflects the user's interests in every period of time.

We have experimented our approach through the *Delicious* social database. We have calculated the overall precision for the enriched user profile. Also, we have detailed the influence of the number of the neighbours in the precision values. Our approach provides better results for the active users (having a

TABLE VI. THE TAGS AMBIGUITY ACCORDING TO DIFFERENT THRESHOLD VALUES

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Percentage of ambiguous tags	13.20	20.00	20.00	22.60	19.23	35.08	26.04	22.61	24.35	22.5
Percentage of comprehensible tags	86.79	80.00	80.00	77.39	80.76	64.91	73.95	77.37	75.64	77.5

lot of a friends).
We have also analysed the tags ambiguity in the results of the enrichment process. We have found that our approach provides a good range of comprehensible tags. This is an advantage of our approach since we have not develop any tool to treat this ambiguity. Tags ambiguity is more present in some cases. This is most often due to the non consideration of abbreviations and commercial products, etc.

In future works, we intend to test our approach by varying the parameters of the temperature calculation in order to show how this variation influences the precision values. We will consider the semantic treatment in the calculation of the temperature of the resources and the weight of the tags. We will also focus on the case of the ”less” active users namely the cold-start users. So, we will try to over this limitation of our approach by proposing a solution for this case.

ACKNOWLEDGMENT

The authors would like to thank Souha Bouhali for participating to this work.

REFERENCES

[1] F. Abel, Q. Gao, G. Houben, and K. Tao, ”Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web”, in Proceedings of ESWC (2), pp.375-389, 2011.

[2] I. Amous-Ben Amor, ”Méthodologies de conception d’applications hypermdia : extension pour la ringnierie des sites web”. Doctoral thesis, Paul Sabatier University, Toulouse, France, 2002.

[3] S. Beldjoudi, H. Seridi, F. Zucker Catherine, ”Improving Tag-based Resource Recommendation with Association Rules on Folksonomies”, in 2nd Workshop on Semantic Personalized Information Management (SPIM).Volume 781 of CEUR Workshop Proceedings, CEUR-WS.org, pp. 26-37, 2011.

[4] T. Bogers and A. van den Bosch. ”Collaborative and Content-based Filtering for Item Recommendation on Social Bookmarking Websites”, in ACM RecSys09 Workshop on Recommender Systems and the Social Web, pp. 6-9, October 2009.

[5] G. Cabanac, 2011. ”Accuracy of inter-researcher similarity measures based on topical and social clues”, in Scientometrics Vol. 87, Issue 3. pp. 597-620, May 2011.

[6] I. Cantador, M. Szomszor, H. Alani, M. Fernandez and P. Castells, ”Enriching Ontological User Profiles with Tagging History for Multi-Domain Recommendations”, in 1st International Workshop on Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb), Tenerife, Spain, 2008.

[7] I. Cantador, P. Brusilovsky and T. Kuflik, ”2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems” (HetRec 2011), in Proceedings of the 5th ACM conference on Recommender systems, RecSys, ACM, New York, NY, USA, 2011.

[8] Y. Cai and Q. Li, ”Personalized search by tag-based user profile and resource profile in collaborative tagging systems”, in Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 969978, 2010.

[9] R. Cazabet , F. Amblard and C. Hanachi, ”Detection of Overlapping Communities in Dynamical Social Networks’, in Social Computing (SocialCom), 2010 IEEE Second International Conference, August 2010, pp.309-314.

[10] M. Daoud, L.-T. Lechani and M. Boughanem. ”Towards a graph-based user profile modeling for a session-based personalized search”, in Knowledge and Information Systems, December 2009, Vol. 21, Issue 3, pp. 365-398.

[11] P. De Meo, G. Quattrone and D. Ursino, ”A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy”, in User Modeling and User-Adapted Interaction. 2010, Vol. 20, Issue 1, pp. 4186.

[12] P. De Meo, E. Ferrara, F. Abel, L. Aroyo and GJ. Houben, ”Analyzing user behavior across social sharing environments”, in ACM Transactions on Intelligent Systems and Technology (ACM TIST) Vol.5, Issue 1, 2014, 31 pages.

[13] D. Godoy and A. Amand, ”Hybrid content and tag-based profiles for recommendation in collaborative tagging systems”, in Latin American Web Congress (LA-WEB), 2008, pp. 5865.

[14] M. Gupta, R. Li, Z. Yin, J. Han, ”Survey on social tagging techniques”, in SIGKDD Explorations, Vol. 12, Issue 1, 2010, pp. 58-72.

[15] D. Guy, N. Zwerdling, I. Ronen, D. Carmel and E. Uziel, ”Social media recommendation based on people and tags”, in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (SIGIR ’10). ACM, New York, NY, USA, 2010, pp. 194-201.

[16] A. Joly, P. Maret, J. Daigremont, ”Contextual Recommendation of Social Updates, a Tag-Based Framework”, in Active Media Technology Lecture Notes in Computer Science Volume 6335, 2010, pp. 436-447.

[17] H.-N. Kim, A. Alkhaldi, A.-E. Saddik and G.-S Joi, ”Collaborative user modeling with user-generated tags for social recommender systems”, in Expert Systems with Applications, 2011, pp. 84888496.

[18] Yicong Liang, Qing Li. ”Incorporating Interest Preference and Social Proximity into Collaborative Filtering for Folk Recommandation”, In SWSM11, July 28, 2011, Beijing, China. ACM 978-1-4503-0757-4/11/07.

[19] M. Lipczak, ”Hybrid Tag Recommendation In Collaborative Tagging Systems”, Doctoral thesis, May 2012. Dalhousie University Halifax, Nova Scotia.

[20] Y. Ma, Y. Zeng, X. Ren and N. Zhong, ”User Interests Modeling Based on Multi-source Personal Information Fusion and Semantic Reasoning, in Active Media Technology (AMT) 2011, pp. 195-205.

[21] A. Manzat, R. Grigoras, and F. Sèdes, ”Towards a User-aware Enrichment of Multimedia Metadata”, in Workshop on Semantic Multimedia Database Technologies (SMDT 2010), Saarbrcken, Germany, Vol. 680, CEUR Workshop Proceedings, 2010, pp. 30-41.

[22] M. Mezghani, C. A. Zayani, I. Amous and F. Gargouri, ”A user profile modelling using social annotations: a survey”, in Proceedings of the 21st international conference companion on World Wide Web (WWW ’12 Companion). ACM, New York, NY, USA, 2012, pp. 969-976.

[23] M. Mezghani, C. A. Zayani, I. Amous and F. Gargouri, ”An Extended Architecture for Adaptation of Social Navigation”. in International Conference on Web Information Systems and Technologies (WEBIST), 2012, pp. 540-545.

[24] M. Mezghani, A. Péninou, C. A. Zayani, I. Amous and F. Sèdes, ”Analyzing Tagged Resources for Social Interests Detection”, International Conference on Enterprise Information Systems (ICEIS), 2014, pp.340-345.

[25] P. Mika, ”Ontologies Are Us: A Unified Model of Social Networks and Semantics”, in The Semantic Web ISWC 2005 Lecture Notes in Computer Science Vol. 3729, 2005, pp. 522-536.

[26] A. K. Milicevic, A. Nanopoulos and M. Ivanovic, ”Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions”, in Artif. Intell. Rev. Vol. 33 no. 3, 2010, pp. 187-209.

[27] K. Musial, P. Kazienko, ”Social networks on the Internet”, in World Wide Web, January 2013, Vol. 16, Issue 1, pp. 31-72.

[28] A. Nauerz, S. Pietschmann and R. Pietzsch, ”Social Recommendation and Adaptation in Web Portals”, in Proceedings of the Workshop on

Adaptation for the Social Web, ASW. Nejdl, Wolfgang and Kay, Judy and Pu, Pearl and Herder, EelcoEditors. Springer. 2008.

[29] R.Z. Rebai, C. A. Zayani and I. Amous, ”An Adaptive Method for User Profile Learning”, in Advances in Databases and Information Systems ADBIS, 2013, pp. 126-134.

[30] M. Roth, A. Ben-David, D. Deutscher, G. Flysher, I. Horn, A. Leichtberg, N. Leiser, Y. Matias and R. Merom, ”Suggesting friends using the implicit social graph”, in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD’10), ACM, New York, NY, USA, 2010, pp. 233-242.

[31] Y. Song, L. Zhang and C. L. Giles, ”Automatic tag recommendation algorithms for social recommender systems”, in ACM Trans. Web, Vol. 5, Issue. 1, Article 4 (February 2011), 31 pages.

[32] D. Tchuente, M-F. Canut, N. Jessel, A. Péninou, F. Sèdes, ”A community-based algorithm for deriving users’ profiles from egocentrics networks: experiment on Facebook and DBLP”, in Social Netw. Analys. Mining, Vol. 3, Issue 3, 2013, pp. 667-683.

[33] S. Zhao, N. Du, A. Nauerz, X. Zhang, Q. Yuan, R. Fu, ”Improved recommendation based on collaborative tagging behaviors”, in Proceedings of the International Conference on Intelligent User Interfaces (IUI08), 2008, pp. 413416. ACM Press, Gran Canaria.

[34] N. Zheng, Q. Li, ”A recommender system based on tags and time information for social tagging systems”, in Expert systems with Applications,Pergamon Press, pp. 4575-4587.

[35] C. A. Zayani, A. Péninou, M-F. Canut and F. Sèdes, ”Towards an Adaptation of Semi-structured Document Querying”, in Proceedings of the CIR’07 Workshop on Context-Based Information Retrieval CIR, 2007, CEUR-WS.org, Vol. 326.

[36] H. Zitouni, L. Berkani, O. Nouali, 2012. ”Recommendation of Learning Resources and Users Using an Aggregation-Based Approach” in Advanced Information Systems for Enterprises (IWAISE), 2012 Second International Workshop on, pp.57-63, Novembre 2012 .