A Personalized Recommender System Based on Users' Information In Folksonomies

Mohamed Nader Jelassi
Université de Tunis El-Manar, Faculté des Sciences de Tunis, LIPAH, 2092, Tunis, Tunisie
Clermont Université, Université Blaise Pascal, LIMOS, BP 10448, F-63000 Clermont Ferrand
nader.jelassi@isima.fr

Sadok Ben Yahia
Université de Tunis El-Manar, Faculté des Sciences de Tunis, LIPAH, 2092, Tunis, Tunisie
sadok.benyahia@fst.rnu.tn

Engelbert Mephu Nguifo
Clermont Université, Université Blaise Pascal, LIMOS, BP 10448, F-63000 Clermont Ferrand
engelbert.mephu_nguifo@univ-bpclermont.fr

ABSTRACT

Thanks to the high popularity and simplicity of *folksonomies*, many users tend to share objects (movies, songs, bookmarks, etc.) by annotating them with a set of tags of their own choice. Users represent the core of the system since they are both the contributors and the creators of the information. Yet, each user has its own profile and its own ideas making thereby the strength as well as the weakness of *folksonomies*. Indeed, it would be helpful to take account of users' profile when suggesting a list of tags and resources or even a list of friends, in order to make a more personal recommandation. The goal is to suggest tags (or resources) which may correspond to a user's vocabulary or interests rather than a list of most used and popular tags in *folksonomies*. In this paper, we consider users' profile as a new dimension of a *folksonomy* classically composed of three dimensions <users, tags, ressources> and we propose an approach to group users with equivalent profiles and equivalent interests as quadratic concepts. Then, we use quadratic concepts in order to propose our personalized recommendation system of users, tags and resources according to each user's profile. Carried out experiments on the large-scale real-world filmography dataset MovieLens highlight encouraging results in terms of precision.

Categories and Subject Descriptors

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

Keywords

Folksonomy, Quadratic Concepts, Profile, Recommender Systems, Users, Precision

1. INTRODUCTION

Complementing the Semantic Web effort, a new breed of so-called "Web 2.0" applications recently emerged on the Web. Indeed, social bookmarking systems have become the predominant form of content categorization of the Web 2.0 age. The main thrust of these Web 2.0 systems is their easy use that relies on simple, straightforward structures by allowing their users to label diverse resources with freely chosen keywords *aka* tags. The resulting structures are called *folksonomies*[1], that is, "taxonomies" created by the "folks", *i.e.,* people [6]. Formally, a *folksonomy* [27] consists of three sets $\mathcal{U}$, $\mathcal{T}$, $\mathcal{R}$ as well as a ternary relation between them, where $\mathcal{T}$ is a set of tags which are arbitrary strings and $\mathcal{R}$ denote resources which may be bookmarked websites[2], described movies[3] or personal shared videos[4] depending on the type of the considered *folksonomy* [15]. While, the third set $\mathcal{U}$ consists of the set of users of the *folksonomy* which are generally described by their user id (pseudonyme). Users are the main actors of such a system since they contribute to its content by adding resources and assigning tags to them; they are thus considered as the creators of the information. The massive participation of users to such systems is owe to the fact that no specific skills are needed for participating [17]; one is then able to contribute to the content without much effort giving full power to users, the main actors of the *folksonomy*.

However, it turns out that the choice of tags and resources shared by a user depends of its profile [25] [26]. Indeed, as he/she is a man or a woman, having less or more than 25 years and whether employee or a professor, the vocabulary changes from a user to another making the cultural diversity in *folksonomies* so fascinating [30]. Even if such a diversity can be seen as a strength, it may be also considered as a weakness. Indeed, *folksonomies* have to reflect the profile of each user during the recommendation of tags or resources, which is however not the case. This gap has led researchers to propose personalized recommender systems in order to suggest tags and resources the most personalized and appropriate to the user profile [33]. The research area of personnalisation attempts to provide solutions to help users share tags and resources among the huge amount of objects in *folksonomies*. For example, consider the user *Bob* (a male) which belongs to the age's category "14-25 years" and is a student. Our goal is to suggest to this user the resources

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

[1]http://wwww.vanderwal.net/folksonomy.html
[2]http://del.icio.us
[3]http://movielens.org
[4]http://youtube.com

or tags which are the most shared by users with same profile (students within the same category of age), *i.e.,* using the same vocabulary and interested in the same resources of the *folksonomy*. Otherwise, we could suggest him users (as friends) which have the same profile and equivalent interests, *i.e.,* interested in the same tags and resources of the *folksonomy*.

For such a task, we consider users' profile as a fourth dimension of a *folksonomy*, classically composed of three dimensions (users, tags and resources), and we propose an approach to group users with same profiles and interests into structures called *quadratic concepts* [19]. Such structures can therefore gather both information about the tagging history and the profile of users. For example, a possible quadratic concept can be: *"the* 18-25 *years old use the same tags to annotate both action and adventure movies"*. As a result, we can propose a personalized recommendation of tags and resources according to each user profile. In addition, we are able to suggest a list of users (friends) with same profiles and interests [13] [22]

The remainder of the paper is organized as follows. In the next section, we motivate the personnalisation of recommendations. Then, we thoroughly study the related work in Section 3. In Section 4, we provide a formal definition of a *folksonomy* and that of a quadri-concept. Then, in Section 5, we use quadri-concepts for introducing our personalized recommander system. In Section 6, carried out experiments about some examples of quadratic concepts and suggestion of users, tags and resources as well as an evaluation of our approach. Finally, we conclude the paper with a summary and we sketch some avenues for future works in Section 7.

2. MOTIVATION

The rise of *folksonomies*, thanks to the success of social bookmarking sharing systems (*e.g.,* FLICKR, BIBSONOMY, YOUTUBE, etc.) has attracted the interest of researchers for the recommendation [4]. Indeed, the lack of organization of shared resources and the (too) much freedom of choice of tags [3] led the present works to enhance the current recommendations in order to assist the user to choose the "good" resources (the most interesting among the thousands of available resources) to share and to use the "good" tags to annotate them [5]. This has resulted in the need to personalize these recommendations. But why has there so need to personalize? [1] [21] [26] The users of a *folksonomy* have different profiles and expectations that depend of their motivations. It is in efforts to comply with needs of each user that different works were interested in the personalization of recommendations. Moreover, personalization attempts to provide solutions to help users solve the information overload issue [22]. But why, in our case, do we further need to know the user's profile? [28] In order to try to fulfill the expectations of each user of the *folksonomy*, it is useful to have more information about him/her [8]. Indeed, implicit or explicit information as age, profession or location of users are information that may help us in the process of the personalization of recommendations. Take for example the case of four users (a rich man, a zoologist, a teenager and a 20-years-old student) looking for resources corresponding to the tag *Jaguar*. According to the profile of each one, a different recommendation (website, photos, ...) will be proposed. Then, the rich man will have as recommendations resources about the *Jaguar* cars; the zoologist will have recommended resources on *Jaguar*, the animal. Likewise, it is more interesting to recommend resources about the british band *Jaguar* for the teenager user. Conversely, we may suggest resources about the movie *The Jaguar* for the 20-years-old student.

On the one hand, if we can easily explore tags used by a single user on a single resource, it is obvious to note that the task quickly becomes unmanageable for a set of taggings involving several users and several resources. On the other hand, recommended tags (or resources) appear not to be very specific [18],*i.e.*, tags which are "hackneyed" words or waves resources which do not match the specific needs of the user. With quadratic concepts, we can solve these two problems. Indeed, quadratic concepts are structures bringing together tags and resources in common to a maximal set of users with same profiles. On the other hand, in a quadratic concept, tags and resources that have been used in combination are grouped resulting in a more specific result. Formally, a quadratic concept is a quadruple (U, T, R, P) which consists of a set U of users, a set T of tags, a set R of resources and a set P of profiles [7] [14]. In a quadratic concept, users with same profiles sharing tags and resources in common are then gathered together. Once extracted, such quadratic concepts are useful for our personalized recommendation system based on the following three areas:

- A user with a given profile sharing a given resource will have a personalized **suggestion of tags** that correspond to its vocabulary.
- The **recommendation of resources** that correspond to the profile of a targeted user.
- The **suggestion of users (friends)** with same profiles and interests to a targeted user.

Contribution

In this paper, we extend the *folksonomy* by considering the users' profile as a new dimension in order to have further information during the recommendation process. Our personalized recommendations are then made in order to avoid offering the same recommendation to any user regardless to its profile. The main originality of our approach is that, in addition to the users' profile, it relies on quadratic concepts (extracted from the extended *folksonomy*) which are structures that brings maximal shared sets of users, tags and resources. Hence, our personalized recommendation algorithm looks for both users' profile and tagging history before recommending the most specific and shared tags or resources to a target user. For instance, most researchers have argued that using shared tags improves resource discovery. While there may be some applications where such personalized tag recommendations would be useful. In this paper, we combine both shared tags/resources (thanks to quadri-concepts) and personalized tags/resources recommendations (thanks to users' profiles).

In the remainder, we will scrutinize the state-of-the-art approaches that propose to personalize the recommendations.

3. RELATED WORK

In an effort to improve the recommendations of users (*resp.,* tags and resources) in *folksonomies*, several works have been proposed in the literature. In [10], users' *personomies*, *i.e.,* relative tags, are used in order to recommend users which shared similar tags or similar resources. First, authors look at the most used tags *w.r.t* a given user, then, based on this tag-based profile, authors are able to recommend users (called *collaborators*) relying on a similarity measure between users called *cos_iuf*. Besides focusing on just one dimension (users), this approach suffer from an approximative measure. Indeed, *cos_iuf*, which relies only on the tags used by the users, does not provide full information about them. Generally speaking, the approaches of the literature [12] have a tendency to suggest friends for a given user according to the tags and resources that he shared in common with them: that is called the behavioral similarity between users. However, this information seems incomplete since it does not provide full information about the user. Thus, our approach differs from previous work by considering of such additional information about users in order to better personalize the proposition of friends: the result will be a list of users with equivalent interests and profile. Recently, in [16], tag recommendations were based on user's tagging history and the personalized preference learned from social contacts. Indeed, according to authors, the social contacts data can be used to provide more personalized recommendations of tags for a user when annotating resources. The limit of such approach is that it requires that a user must have social contacts to get tag recommendations. Hence, new users have not the possibility to get recommended tags. In [18], Hotho *et al.* propose tag recommendations in *folksonomies* based on most used tags. However, it turns out that such recommendation is not personalized at all since that the same tags are suggested to any user justifying even more the need of considering a new dimension in the *folksonomies* in our approach, *i.e.,* the users' profile. Lipczak proposed in [24] a three-step tag recommendation system: starting from tags assigned to resources, the authors add tags proposed by a lexicon based on co-occurrences of tags within resource's posts. Then, the system filters the tags already used by the user. While, the recommendation does not seem personalized despite filtering users' tags since it seeks for tags that co-occurred on other posts. The approach only comes back to remove tags previously used by a user from the list of recommended tags. In [21], the authors proposed an approach combining resource similarity and user similarity to recommend personalized tags. Through identifying a set of users similar to the active user and measuring their similarity, the approach is able to suggest personalized tags. Thus, two users are considered as similar if they assigned the same tags to the same resources. However, it is rare to find such a situation in *folksonomies* which tags used by two different users on the same resources are rarely the same.

To summarize, the information (tags and resources) considered by these works seems to be incomplete in order to personalize the recommendations. We believe necessary to use additional information about the user's profile in order to personalize the recommendations which would be suggested to the users that fits their profile in *folksonomies*. We will show that by combining such information with the history tagging of users, we can improve recommendations. Furthermore, the major problem of most of the approaches described is that the suggested tags are often those that are the most used in *folksonomies*. These tags, however, are not very specific[5]. Within quadratic concepts, we can tackle that issue by focusing not only on which tags have been used, but rather on which tags have been used in combination; the result is more specific. In the next section, we propose a definition of a quadratic concept which allow grouping both information about history tagging and profile's information of a user. Then, we introduce our personalized recommander system which is able to recommend users as well as tags and resources.

4. QUADRATIC CONCEPTS

In this section, we introduce the notion of a quadratic concept. But, we first introduce an extension of the notion of *folksonomy* [17] by the addition of a fourth dimension, *i.e.,* the users' profile.

DEFINITION 1. *(*P-FOLKSONOMY*) A p-folksonomy is a set of tuples* $\mathcal{F}_p = (\mathcal{U}, \mathcal{T}, \mathcal{R}, \mathcal{P}, \mathrm{Y})$ *where* $\mathcal{U}$, $\mathcal{T}$, $\mathcal{R}$ *and* $\mathcal{P}$ *are finite sets which elements are called users, tags, resources and profiles.* $\mathrm{Y} \subseteq \mathcal{U} \times \mathcal{T} \times \mathcal{R} \times \mathcal{P}$ *represents a quaternary relation where each* $y \subseteq \mathrm{Y}$ *can be represented by a quadruple :* $y = \{(u, t, r, p) \mid u \in \mathcal{U}, t \in \mathcal{T}, r \in \mathcal{R}, p \in \mathcal{P}\}$ *which means that the user* u *with the profile* p *has annotated the resource* r *using the tag* t.

REMARK 1. *Users are required to provide needed information to profile themselves, such as demographic information (e.g., gender, nationality, occupation), opinion information (e.g., reviews, comments and feedbacks). Moreover, explicit rating information is commonly used to profile users' interests or opinions. Such information are collected in the fourth dimension of the folksonomy, i.e., the users' profile. Later, in the recommendation process, we are going to study the* evolutive *aspect of the profile, i.e., tags and resources shared by users in the folksonomy.*

EXAMPLE 1. *Table 1 depicts an example of a p-folksonomy* $\mathcal{F}_p$ *with* $\mathcal{U} = \{u_1, u_2, u_3, u_4\}$, $\mathcal{T} = \{t_1, t_2, t_3\}$, $\mathcal{R} = \{r_1, r_2\}$ *and* $\mathcal{P} = \{p_1, p_2\}$. *Each cross within the quaternary relation indicates a tagging operation by a user from* $\mathcal{U}$ *with a profile from* $\mathcal{P}$, *a tag from* $\mathcal{T}$ *and a resource from* $\mathcal{R}$, *i.e., a user has tagged a particular resource with a particular tag at a date* d. *For example, the user* u_1 *with the profile* p_1 *and* p_2 *has tagged the resources* r_1, r_2 *and* r_3 *via the tags* t_2, t_3 *and* t_4. *In practice, we can imagine, for example, that the user* u_1 *which is a student (*p_1*) and* 25 *years old (*p_2*) has tagged three different websites via the tags sport, results and news. Meanwhile, we can also see that the users* u_2 *and* u_3 *have also tagged the same websites via the same tags.*

We now define a quadratic concept.

DEFINITION 2. *(*(FREQUENT) QUADRATIC CONCEPT*) A quadratic concept (or a quadri-concept for short) of a p-folksonomy* $\mathcal{F}_p = (\mathcal{U}, \mathcal{T}, \mathcal{R}, \mathcal{P}, \mathrm{Y})$ *is a quadruple (*U, T, R, P*) with* $U \subseteq \mathcal{U}$, $T \subseteq \mathcal{T}$, $R \subseteq \mathcal{R}$ *and* $P \subseteq \mathcal{P}$ *with* $U \times T \times R \times P \subseteq \mathrm{Y}$ *such that the quadruple (*U, T, R, P*) is maximal,* i.e., *none of these sets can be extended without shrinking one of the other three dimensions.*

For a *Quadri-Concept* $QC = (U, T, R, P)$, the U, R, T and P parts are respectively called ***Extent***, ***Intent***, ***Modus*** and ***Variable***.

[5]The same problem stands for resources.

$\mathcal{F}_p$	$\mathcal{R}$	r_1				r_2				r_3			
$\mathcal{P}$	$\mathcal{U}/\mathcal{T}$	t_1	t_2	t_3	t_4	t_1	t_2	t_3	t_4	t_1	t_2	t_3	t_4
	u_1		×	×	×		×	×	×		×	×	×
p_1	u_2		×	×	×	×	×	×	×	×	×	×	×
	u_3		×	×	×	×	×	×	×	×	×	×	×
	u_4					×			×	×			×
	u_1		×	×	×		×	×	×		×	×	×
p_2	u_2		×	×	×	×			×	×	×	×	×
	u_3		×	×	×	×	×	×	×	×	×		
	u_4					×			×	×			×

Table 1: An example of a *p-folksonomy*.

Remark 1. *Without support restrictions, we could have a huge number of quadri-concepts from a given p-folksonomy. In order to keep the interesting ones and have only the most significant ones, we can define minimum thresholds on each dimension of the p-folksonomy, i.e., $minsupp_u$, $minsupp_t$, $minsupp_r$ et $minsupp_p$. Doing so, it results in quadri-concepts which are called **frequent**.*

In order to mine all (frequent) quadri-concepts from a given *p-folksonomy*, *i.e.,* a quadratic context, we may use one of the reported algorithms in the literature: QuadriCons [20], an extension of the TriCons algorithm [31] or DataPeeler [7] which are able to extract all frequent quadri-concepts fulfilling user-defined minimum thresholds on each dimension of the *p-folksonomy*. Both algorithms take as input a *p-folksonomy* as well as four minimum thresholds and output the set $\mathcal{QC}$ of frequent quadri-concepts fulfilling these thresholds. Running this type of algorithm on the *p-folksonomy* illustrated by Table 1, with $minsupp_u = 2$, $minsupp_t = 2$, $minsupp_r = 2$ and $minsupp_p$=2 allows the extraction of the frequent quadri-concepts QC_1, QC_2 and QC_3 such as $QC_1 = \{(u_1,u_3), (t_2,t_3,t_4), (r_1,r_2), (p_1,p_2)\}$, $QC_2 = \{(u_1,u_2), (t_2,t_3,t_4), (r_1,r_3), (p_1,p_2)\}$ and $QC_3 = \{(u_2,u_4), (t_1,t_4), (r_2,r_3), (p_1,p_2)\}$.

In what follows, we use the quadri-concepts to introduce our personalized recommendation system of users, tags and resources.

5. PERSOREC: A PERSONALIZED RECOMMENDER SYSTEM FOR FOLKSONOMIES

In this section, we introduce our algorithm for a personalized recommendation in *folksonomies* oriented to a target user. The pseudo code of the PersoRec algorithm is sketched by Algorithm 1. PersoRec takes as input $\mathcal{QC}$, a set of frequent quadri-concepts, a user u with its profile p and (optionally) a resource r (to annotate) and outputs three sets: a set of proposed users, suggested tags and recommended resources.

PersoRec runs the set of frequent quadri-concepts in order to find those where the profile p of the user u belongs to the variable part (*i.e.,* the set of profiles). Then, depending of the type of application, PersoRec operates as follows:

User Proposition PersoRec seeks for quadri-concepts whose users have the same profile than u. If u already belongs to the quadri-concept, then the user (of qc) is skipped, otherwise, it is added to the set $\mathcal{PU}$ of proposed users (Line 7).

Algorithme 1 : PersoRec

Input :

1. $\mathcal{QC}$: the set of frequent quadri-concepts.
2. a user u with a profile p.
3. a resource r.

Output :

1. $\mathcal{PU}$: the set of proposed users.
2. $\mathcal{ST}$: the set of suggested tags.
3. $\mathcal{RR}$: the set of recommended resources.

1 **Begin**
2 **For each** *quadri-concept* $qc \in \mathcal{QC}$ **do**
3 **If** $p \in qc.Variable$ **then**
4 /**User proposition**/
5 **For each** *user of qc.extent* **do**
6 **If** $qc.extent \neq u$ **then**
7 $\mathcal{PU} = \mathcal{PU} \cup qc.extent$;
8 **End**
9 **end**
10 /**Tag Suggestion**/
11 **If** $qc.Intent = r$ **then**
12 $\mathcal{ST} = \mathcal{ST} \cup qc.modus$;
13 **End**
14 /**Resource Recommandation**/
15 $\mathcal{RR} = \mathcal{RR} \cup qc.Intent$;
16 **End**
17 **end**
18 **return** $(\mathcal{PU},\mathcal{ST},\mathcal{RR})$;
19 **End**

Tag Suggestion In this application, the aim is to suggest personalized tags to a target user which share a resource in the *p-folksonomy*. Thus, by looking at its profile, we could suggest him (her) tags that were affected by users with same profile to the same resource. The goal is to offer tags that are more in agreement with the user's profile. For this application, we need an additional information, *i.e.,* the resource to be annotated (r). Then, we display suggested tags that were affected to the same resource r by users with same profile than u into the set $\mathcal{ST}$ (Line 10).

Resource recommendation The goal is to propose a personalized list of resources to a targeted user that is susceptible to be in accordance with its interests. To do that, we look at its profile and displaying resources from quadri-concepts corresponding to the suitable profile. Hence, PersoRec is able to recommend resources that were shared by users with same profile than the user u. Hence, the set $\mathcal{RR}$ contains resources which are recommended to u (Line 15).

Remark 2. *It is worth of mention that the step of extraction of quadri-concepts is an off-line phase that is performed only once. Indeed, our* PersoRec *algorithm relies on quadri-concepts already extracted. Thus, our recommander system does not suffer from the cost of the extraction of quadri-concepts at each recommendation. This phase can thus be seen as a preprocessing phase of recommendation.*

In the following, we evaluate our approach on a real-world dataset: MovieLens. We give some interesting examples of quadri-concepts and real cases of recommandations. Then, we evaluate the quality of our recommendations, *i.e.,* their precision.

6. EXPERIMENTAL EVALUATION

In this section, we evaluate our approach on the real-world dataset MovieLens[6] by giving some examples of extracted quadri-concepts following different profiles of *folksonomy*' users. Then, we present some examples of personalized recommendations generated by the PersoRec algorithm. Finally, we evaluate our recommendations by computing their precision.

6.1 The MovieLens filmography dataset

MovieLens is a recommender system and virtual community website that allows users to share movies using tags. By September 1997, the website had over 50,000 users. In such a movie recommendation website, users are asked to annotate movies they like or dislike. The MovieLens dataset, used for our experiments, is freely downloadable[7]. This dataset contains 95580 tags applied to 10681 movies by 71567 users of the online movie recommender service MovieLens. Additional available information about users constitute the profile (the fourth dimension of the *p-folksonomy*) which can be: the gender (male or female), the profession (21 in total, which can be administrator, artist, doctor, educator, engineer, entertainment, executive, healthcare, homemaker, lawyer, librarian, marketing, programmer, retired, salesman, scientist, student, technician, writer, other or none) or the age (see, Remark 3).

Remark 3. *Contrariwise to the two first types of profile (gender and profession), first experiments show that it is very rare to find users with exactly the same age sharing same resources and same tags. Hence, to have exploitable results, we increase the granularity related to the age by partitioning the original data into five categories:*

- $7 - 18$ *years*
- $19 - 24$ *years*
- $25 - 35$ *years*
- $36 - 45$ *years*
- $46 - 73$ *years*

6.2 Examples of extracted quadri-concepts

In what follows, we present some interesting results of extracted quadri-concepts from the dataset MovieLens. For such an extraction, we ran both algorithms of the literature dedicated to this task, *i.e.,* QuadriCons and Data Peeler on a machine with a processor Intel Core *i*7 and a 4 Go memory. Several tests, carried out on the operating system Linux Ubuntu 10.10.1 (64 bits), have shown that the QuadriCons algorithm performs better than its competitor in terms of execution time (*e.g.,* between 1 and 50 seconds for QuadriCons *vs.* 300 seconds at least for Data Peeler). Given the large size of the considered dataset, we therefore opted for the QuadriCons algorithm[8] for generating the frequent quadri-concepts <users, tags, resources, profiles>.

In the first place, we defined the following thresholds values of supports: $minsupp_u = 2$, $minsupp_t = 2$, $minsupp_r = 2$ and $minsupp_p = 2$[9]. Obviously, it seems more interesting to set each minimum threshold to 2 (at least) in order to have quadri-concepts with an added value illustrating shared tags and resources by a group (of two users at minimum) with same profile. Hence, Table 2 depicts some examples of quadri-concepts among the 10627 frequent quadri-concepts fulfilling these aforementioned thresholds on each dimension. The first quadri-concept shows that the users *bernadette*, *bridget* and *margaret62*, three retired females aged between 46 and 73 years old, shared the movies *Star Wars*, *M.A.S.H* and *Rear Window* via the tags *classic*, *dialog* and *oscar*. While, in the second quadri-concept, we can see that the $25-35$-years-old males working in healthcare annotated the tags *cult* and *bestmovie* to the three mentioned movies. Finally, in the third quadri-concept, three writers aged between 36 and 45 years, *i.e., ross*, *anlucia* and *franela* opted for the tags *quotes*, *classic* and *oldmovie* when they shared the movies *Braveheart*, *Magician of OZ* and *Gone with the Wind*.

In the second place, we defined the following thresholds values of supports: $minsupp_u = 2$, $minsupp_t = 2$, $minsupp_r = 2$ and $minsupp_p = 1$[10]. Roughly speaking, we separate

[6]http://movielens.umn.edu/

[7]http://www.grouplens.org/node/73

[8]The code is available at this link *http://www.isima.fr/ mephu/FILES/QuadriCons/*

[9]in a frequent quadri-concept, at least, 2 users with same two profiles have assigned the same tags (2 at least) to same resources (2 at least).

[10]in a frequent quadri-concept, at least, 2 users with a same profile have assigned the same tags (2 at least) to same resources (2 at least).

quadri-concepts by profile in order to see how each category of age (*resp.,* gender, profession) see the *folksonomy*, *i.e.,* what they shared and using what tags. Hence, Tables 3, 4 and 5 depict some examples of the 17862 (frequent) quadri-concepts extracted from the MovieLens dataset for the aforementioned thresholds. We recall that quadratic concepts (U,T,R,P) have the property that each user in U, with a profile P, has tagged each resource in R with all tags from T, and that none of these sets can be extended without shrinking one of the other three dimensions. For example, for the quadri-concept {(*menez_jer*,*verra18*,*pasto*), (*oldmovie*,*award*,*classic*), (*Silence of the Lambs*,*Casablanca*,*Usual Suspects*), 36-45 years}, if we increase the set of users, we reduce inevitably the set of tags or/and resources. Inversely, if we increase the set of tags (or resources), the set of users will be automatically be reduced. This is due to the mathematical properties of quadri-concepts. For each profile described in section 6.1, we report the following results:

Gender Table 5 depicts quadri-concepts with the user's gender as a profile. For example, the first quadri-concept shows that the users *nichlea*, *chrisc* and *davidd* share the three movies *Silence of the Lambs*, *The Rock* and *Blade runner* through the tags *crime*, *thriller* and *action*. The Table shows that the tagging and shared resources differs between males and females. If females are more interested by romance movies with a love story like *Sweet November* or *Titanic* using tags like *romance* or *lovestory*, males are more interested in thriller and action movies such as *Blade Runner*. Thanks to quadri-concepts which group users with same profile, Table 3 clearly highlights two interests completely different that should be taken into account during the suggestion process.

Profession Table 4, which depicts quadri-concepts when the profile concerns the user's profession, highlights a difference of vocabulary and interests between users of different professions. For example, we have the users *ched50*, *slioua7* and *nina16*, which are students, that shared the movies *Appolo 13* and *Raiders of Lost Ark* via the tags *adventure* and *action*. We can see, for example, that students shared massively action movies within tags like *adventure* and *action*, while lawyers would rather share movies with a crime story like *The Fugitive* through tags like *detective* and *crime_story*. We note also that librarians users are interested by movies which are based on a book such as *Dead poets society* which corresponds better to their centers of interest.

Age Finally, it is interesting to analyze quadri-concepts with users from different **age**'s categories in order to view the different interests/points of view of each category. In Table 3, each quadri-concept corresponds to a particular age's category. Hence, the Table shows that users under 18 years prefer science fiction movies while users aged from 25 and 35 years are more interested by action and adventure ones. The older users have a particular interest for classic and old movies such as *Casablanca* and *Rear Window*. The most represented category concern users aged between 19 and 24 years which tend to share polar movies by using the tags *adventure* and *polar*.

6.3 Personalized recommendation

Quadri-concepts provides a new way for grouping under a same concept users with same profiles sharing tags and resources in common. There are several areas in which quadri-concepts can be useful for. Thanks to our PersoSec algorithm, we are able to illustrate three of them. In what follows, we give some examples of real cases for each application.

Tag suggestion Consider the user *davis* which wants to share the movie *The Fugitive*, then he will get a choice of tags that depends of his profile. If *davis* was a librarian, we would suggest him the tags *author* and *based_on_a_book*, while if he was a lawyer, he would be able to annotate the movie with the tags *detective* and *crime_story* (*cf.,* Table 4).

Resource recommendation Obviously, a young user will be more interested by comedy than by classical movies; likewise, we would better recommend romantic than war movies to a female user. Consider two new users *reyes* (22 years old) and *zlatan* (51 years old): our recommendation system will recommend the movies *Seven* and *Toy Story* for *reyes* while movies like *Rear Window* and *M.A.S.H* will be recommend to *zlatan*, which could correspond better to his interests. Since these two users are new in the *p-folksonomy*, we have no available information about their taggings. Hence, our recommendations for both users are only based on their profile (*i.e.,* their age in this case).

Friend proposition For example, consider the user *krycek* which is librarian (*cf.,* Table 4), users that we could recommend him as friends are *deepthr* and *mistx12* since they have same profiles and interests.

Example 2. *Consider the quadri-concepts illustrated by Table 2 and let u_1=Jacob (37 years, Writer) and u_2=Ilana (Female, 63 years) be two users of the p-folksonomy. Suppose that Jacob and Ilana both want to share the movie Rear Window. Thanks to our personalized recommender system, we are able to provide two kinds of suggestion following the profile of each user. Hence, Jacob will have the following suggested tags: classic, quotes and oldmovie while tags like classic, dialog and oscar will be suggested to Ilana in order to annotate that movie. Moreover, we may propose as friends the users rossy, anlucia and franela to Jacob since they shared the same interests and have the same profile. While, it seems more suitable to propose the users bernadette, bridget and margaret62 to Ilana since they all have about the same age and same interests.*

6.4 Evaluation of the recommendation

Training set/Test set

The MovieLens dataset was partitioned into two sub datasets: the first sub dataset, containing 80% of users, was used as **training set** while the second one, containing the remaining users (*i.e.,* 20% of users), was retained as the validation data for tests (*i.e.,* the **test set**). For each test user, a random 20% of its tagging is considered as the test/answer set and 80% as its training set. For each user of the test set, our recommender algorithm generate a list of items (tags, users or resources) based on the user's training set. If an item in

Users	Tags	Resources	Profile
bernadette	classic	Star Wars	
bridget	dialog	M.A.S.H	Female, Female, 46-73 years, retired
margaret62	oscar	Rear Window	
mulder	bestmovie	Usual Suspects	
scully	cult	Silence of the Lambs	Male, 25-35 years, healthcare
csmdavis		Sound of Music	
ross	classic	Rear Window	
anlucia	oldmovie	Magician of OZ	Male, 36-45 years, Writer
franela	quotes	Gone with the Wind	

Table 2: Examples of quadri-concepts extracted from the MovieLens dataset.

Users	Tags	Resources	Profile
steve,patr,mkl23	sciencefiction,cult	Star Wars,Independance Day,The Rock	7-18 years
nad16,mehz19,emma	adventure,polar	Toy Story,Seven,Star Wars,Braveheart	19-24 years
crisron,inies16,sirig,lionelz	thriller,action	Taxi Driver,Star Wars,Blade Runner,Die Hard	25-35 years
menez,verra18,pasto	award,oldmovie,classic	Silence of the Lambs,Casablanca,Usual Suspects	36-45 years
bernadette,bridget,robertn	classic,dialog,oscar	Star Wars,M.A.S.H,Rear Window	46-73 years

Table 3: Examples of quadri-concepts following the profile *age*.

Users	Tags	Resources	Profile
krycek,deepthr,mistx12	author,based_on_a_book	The Fugitive,Dead poets society	librarian
fox16,dana,cgbspender	crime_story,detective	Pulp Fiction,The Godfather,The Fugitive	lawyer
ross, anlucia, nabilawi	classic,oldmovie	Braveheart,Magician of OZ	retired
fran,chandy,joeytr,phoeb	adventure,history,thriller	Star Wars,Blade Runner,Monty Python	engineer
ched50,slioua7,nina16	adventure,action	Seven,Appolo 13,Raiders of Lost Ark	student

Table 4: Examples of quadri-concepts following the profile *profession*.

Users	Tags	Resources	Profile
nichlea,chrisc,davidd	crime,thriller,action	Silence of the Lambs,The Rock,Blade Runner	Male
regina,gilliana,laurieh	passion,romance,lovestory	Bridget Jones,Titanic,Sound of Music,Sweet November	Female

Table 5: Examples of quadri-concepts following the profile *gender*.

the recommendation list was also in the user's test set, then the item is considered as pertinent.

Assessing the efficiency of a recommendation algorithm is far from being trivial. First, because different algorithms can be better or worse according to the dataset on which they are applied. Second, the objectives of a recommendation system can be different and varied. A recommendation system can be implemented in order to accurately estimate the score that would give a user to an element whereas others will have as main objective not to propose incorrect recommendations. One can therefore legitimately wonder to what extent these different methods of recommendation are really effective. Nevertheless, to determine the efficiency of a system, the most common indicator in the literature is the *precision* (*cf.,* Equation 1), which represents the quality of the recommendation, *i.e.,* how much the proposed suggestions are in line (*i.e.,* pertinent) with the interests of the user.

$$Precision = \frac{number\ of\ pertinent\ recommendations}{number\ of\ recommendations} \quad (1)$$

Thus, precision determines the probability that a recommended element is relevant. Thus, the best measure of the efficiency of a recommendation algorithm and the relevance of the recommendations is to evaluate the precision of the prediction performed by the system by comparing the predictions with the choices that would have given the user in the real case [29]. In our experiments, we also varied the number of recommendations provided to the user. This is known as top-k query. With such requests, the user can specify the k answers (recommendations) the most relevant that the system shall return to him. This helps especially to avoid overwhelming the user with a large number of answers by returning only the number of the most relevant answers that he wishes [9]. In what follows, we focus on one of the three applications of our recommendation system, *i.e.,* the recommendation of resources and we assess the precision of our approach versus the closed approach of the literature to our works, *i.e.,* that of Liang *et al.* [23] which generates personal recommendations based on users' information. Thus, Figure 1 shows the different values of precision obtained by our recommendation algorithm *vs.* that of Liang *et al.* for different values of k going from 5 to 10.

Generally, our recommendations made to users of MovieLens reflects their expectations, *i.e.,* the recommendations are relevant in an average of 38% which outperforms the precision of the approach of Liang *et al.* which is between 24% and 30%. It suggests that the use of quadratic concepts improves the recommendations by suggesting the tags and resources the more specific to users' needs. Indeed, quadri-concepts offer to the users tags and resources that were be shared in common by a similar set of users to the targeted user. Results also show that the best performances were obtained with a value of k=5. This is because the first five recommendations correspond to the expectations of users and when the number of recommendations increased, this inevitably leads to a decrease in precision as the user selects fewer resources than those which are recommended [32].

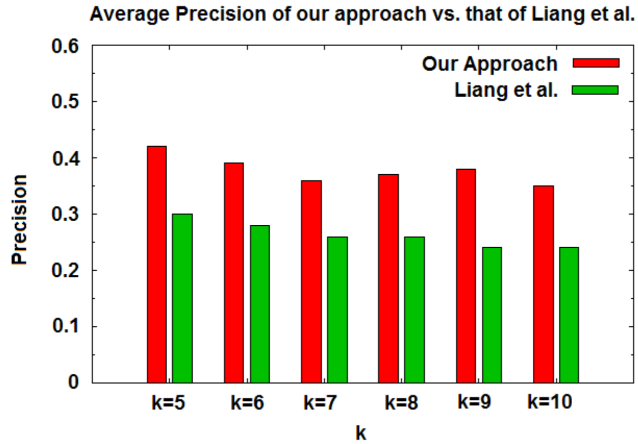

Figure 1: Average precision of our approach for the recommendation of resources.

7. CONCLUSION AND PERSPECTIVES

Because of its simplicity of use, *folksonomy* systems have rapidly attracted millions of users which share resources by affecting them freely chosen tags. However, it turns out that recommendation systems take little account of user's profile, resulting in brute and impersonal suggestions which may not correspond to user's desires. In this paper, we considered the users' profile as a new dimension in the *folksonomy* and we use the QuadriCons algorithm to mine quadratic concepts of users, tags, resources and profiles. Such concepts, bridging together into structures maximal sets of users sharing resources with the same tags, are then used by our introduced personalized algorithm in order to offer a personalized choice of tags and resources since suggestions are made following the users' profiles. Among perspectives of our works, we can investigate the feed back of users in order to show if our suggestions are better considered than the brute ones. When considering the users' profile, we may explore other implicit and explicit information as reviews, comments or browsing history and click streams in order to have the evolutionary aspect of the users' profiles to better purchase their interests. Moreover, we may also explore other application domains in which quadratic concepts are useful like query completion or mining sequential patterns by considering the time as an additional dimension/information in *folksonomies* [2] [11].

8. REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing Temporal Dynamics in Twitter Profiles for Personalized Recommendations in the Social Web. In *Proceedings of ACM WebSci '11, Koblenz, Germany*, 2011.

[2] R. Agrawal and R. Srikant. Mining sequential patterns. In *Proc. of the ICDE*, ICDE '95, pages 3–14, Washington, DC, USA, 1995. IEEE Computer Society.

[3] H. Alves and A. Santanchè. Abstract framework for social ontologies and folksonomized ontologies. In *Proceedings of the 4th International Workshop on Semantic Web Information Management*, SWIM '12, pages 10:1–10:4, New York, NY, USA, 2012. ACM.

[4] S. Amer-Yahia. Recommendation projects at yahoo! *IEEE Data Eng. Bull.*, 34:69–77, 2011.

[5] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu. Can all tags be used for search? In *Proc. of the 17th ACM CIKM 2008*, pages 193–202, Napa Valley, California, 2008. ACM Press.

[6] C. Cattuto, C. Schmitz, A. Baldassarri, A. Servedio, V. Loreto, A. Hotho, M. Grahl, and G. Stumme. Network properties of folksonomies. *In Proc. of AICSI on NANSE, Amsterdam, The Netherlands.*, pages 245–262, 2007.

[7] L. Cerf, J. Besson, C. Robardet, and J.-F. Boulicaut. Closed patterns meet n-ary relations. *ACM TKDD*, 3:3:1–3:36, March 2009.

[8] M. Das, S. Thirumuruganathan, S. Amer-Yahia, G. Das, and C. Yu. Who tags what? an analysis framework. *In Proceedings of PVLDB*, 5(11):1567–1578, 2012.

[9] W. Dedzoe. *Traitement de Requêtes Top-k dans les Communautés Virtuelles P2P de Partage de Données.* These, Université de Nantes, Nov. 2011.

[10] J. Diederich and T. Iofciu. Finding communities of practice from user profiles based on folksonomies. In *Proceedings of the 1st International Workshop on TEL-CoPs, Crete, Greece*, pages 288–297, 2006.

[11] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins. Visualizing tags over time. *ACM Trans. Web*, 1(2):193–202, 2007.

[12] L. Fang, A. Fabrikant, and K. LeFevre. Look who i found: understanding the effects of sharing curated friend groups. In *Proc. of the 3rd Annual ACM Web Science Conference*, WebSci '12, pages 95–104, New York, NY, USA, 2012.

[13] Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang. Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In *Proceedings of the 32nd international ACM SIGIR*, pages 540–547, New York, NY, USA, 2009. ACM.

[14] A. Hotho. Data mining on folksonomies. In G. Armano, M. de Gemmis, G. Semeraro, and E. Vargiu, editors, *Intelligent Information Access*, volume 301 of *Studies in Computational Intelligence*, pages 57–82. Springer, Berlin / Heidelberg, 2010.

[15] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *Proc. of ESWC, Budva, Montenegro*, volume 4011 of *LNCS*, pages 411–426. Springer, Heidelberg, 2006.

[16] J. Hu, B. Wang, and Z. Tao. Personalized tag recommendation using social contacts. In *Proceedings of Workshop SRS'11, in conjunction with CSCW*, 2011.

[17] R. Jäschke, A. Hotho, C. Schmitz, B. Ganter, and G. Stumme. Discovering shared conceptualizations in folksonomies. *Web Semantics.*, 6:38–53, 2008.

[18] R. Jäschke, L. Marinho, A. A. Hotho, S.-T. Lars, and G. Stum. Tag recommendations in folksonomies. In *Proc. of the 11th ECML PKDD, Warsaw, Poland*, pages 506–514, 2007.

[19] M. N. Jelassi. A quadratic approach for trend detection in folksonomies. In *Proceedings of the 6th RR, Vienna, Austria*, pages 278–283, 2012.

[20] M. N. Jelassi, S. Ben Yahia, and E. Mephu Nguifo. A scalable mining of frequent quadratic concepts in d-folksonomies. *ArXiv e-prints*, Dec. 2012.

[21] N. Landia and S. Anand. Personalised tag recommendation. *Recommender Systems & the Social Web, New York, NY, USA*, pages 83–86, 2009.

[22] H. Liang. *User profiling based on folksonomy information in Web 2.0 for personalized recommender systems.* PhD thesis, Queensland University of Technology, 2010.

[23] H. Liang, Y. Xu, Y. Li, and R. Nayak. Personalized recommender system based on item taxonomy and folksonomy. In *Proceedings of the 19th ACM CIKM'10*, pages 1641–1644, New York, NY, USA, 2010. ACM.

[24] M. Lipczak. Tag recommendation for folksonomies oriented towards individual users. In *Proc. of the ECML PKDD Discovery Challenge, Antwerp, Belgium*, 2008.

[25] E. Michlmayr. Learning user profiles from tagging data and leveraging them for personal(ized) information access. In *Proceedings of the 16th WWW*, 2007.

[26] E. Michlmayr and S. Cayzer. Learning user profiles from tagging data and leveraging them for personal(ized) information access. In *Proc. of the Workshop on Tagging and Metadata for Social Information Organization in the 16th WWW, Banff, Alberta, Canada*, 2007.

[27] P. Mika. Ontologies are us: A unified model of social networks and semantics. *Web Semantics.*, 5(1):5–15, 2007.

[28] M. Noll, G. Michael, and C. Meinel. Web search personalization via social bookmarking and tagging. In *Proc. of the 6th ISWC/ASWC, Busan, Korea*, pages 367–380, 2007.

[29] C. Penet, C.-H. Demarty, G. Gravier, and P. Gros. De la détection d'évènements sonores violents par SVM dans les films. In *Proc. of the 13th ORASIS - Congrès des jeunes chercheurs en vision par ordinateur*, Praz-sur-Arly, France, 2011.

[30] E. Tonkin and M. Guy. Folksonomies: Tidying up tags? *D-Lib*, volume 12(1), 2006.

[31] C. Trabelsi, N. Jelassi, and S. Ben Yahia. Scalable mining of frequent tri-concepts. In *Proc. of the 15th PAKDD, Kuala Lampur, Malaysia*, pages 231–242, 2012.

[32] C. Trabelsi, N. Jelassi, and S. B. Yahia. Auto-complétion de requêtes par une base générique de règles d'association triadiques. In G. Pasi and P. Bellot, editors, *CORIA*, pages 9–24. Éditions Universitaires d'Avignon, 2011.

[33] D. Vallet, I. Cantador, and J. M. Jose. Personalizing web search with folksonomy-based user and document profiles. In *Proceedings of the 32nd ECIR*, pages 420–431, Berlin, Heidelberg, 2010. Springer-Verlag.