



# Improving Tag-based Resource Recommendation with Association Rules on Folksonomies

Samia Beldjoudi, Hassina Seridi, Catherine Faron Zucker

## ► To cite this version:

Samia Beldjoudi, Hassina Seridi, Catherine Faron Zucker. Improving Tag-based Resource Recommendation with Association Rules on Folksonomies. 2nd ISWC Workshop on Semantic Personalized Information Management: Retrieval and Recommendation, SPIM 2011, 2011, Bonn, Germany. hal-01201744

**HAL Id: hal-01201744**

**<https://inria.hal.science/hal-01201744>**

Submitted on 17 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Improving Tag-based Resource Recommendation with Association Rules on Folksonomies

Samia Beldjoudi<sup>1</sup>, Hassina Seridi<sup>1</sup>, and Catherine Faron Zucker<sup>2</sup>

<sup>1</sup> Laboratory of Electronic Document Management LabGED  
Badji Mokhtar University, Annaba, Algeria  
{seridi,beldjoudi}@labged.net

<sup>2</sup> I3S, Université Nice - Sophia Antipolis, CNRS  
930 route des Colles, BP 145, 06930 Sophia Antipolis cedex, France  
catherine.faron-zucker@unice.fr

**Abstract.** In this paper, we propose a method to analyze user profiles according to their tags in order to personalize the recommendation of resources. Our objective is to enrich the profiles of folksonomy users with pertinent resources. We argue that the automatic sharing of resources strengthens social links among actors and we exploit this idea to enrich user profiles by increasing the weights associated to web resources according to social relations. We base upon association rules which are a powerful method for discovering interesting relationships among a large set of data on the web. We extract association rules from folksonomies and use them to recommend supplementary resources associated to the tags involved in these rules. In this recommendation process, we reduce tag ambiguity by taking into account social similarities calculated on folksonomies.

**Keywords:** Folksonomies, Social Tagging, Association Rules, Tag-based Resource Recommendation, Tag Ambiguity.

## 1 Introduction

Web 2.0 technologies have created the conditions for new usages on the web which has become a social web. Users create, annotate, share and make public what they find interesting on the web and therefore are greatly involved in the evolution of the web. Folksonomies are one of the keystones of these new social practices: they are systems of classification derived from the practice and method of collaboratively creating and managing tags to annotate and categorize content. This practice is known as collaborative tagging or social tagging. Among the most popular social websites based on folksonomies let us cite Delicious which offers an effective way to conduct the collaborative management of social bookmarking, Flickr which is a photo management and sharing web application, Youtube and Dailymotion designed for sharing videos, Myspace and Odeo for sharing music files.

The basic principle of social tagging relies on three main elements: the user, the resource and the tag. The combination of these three elements enables the

development of semantic tools exploiting both folksonomies and annotations of web resources by users with tags. In this paper, we propose a method to analyze user profiles according to their tags in order to predict interesting personalized resources and recommend them. In other words, our objective is to enrich the profiles of folksonomy users with pertinent resources. We argue that the automatic sharing of resources strengthens social links among actors and we exploit this idea to reduce tag ambiguity in the recommendation process by increasing the weights associated to web resources according to social similarities. We base upon association rules which are a powerful method for discovering interesting relationships among a large set of data on the web.

We insist on the fact that our final aim is not to suggest tags to users: each time a resource is presented to a user, the tags already used to annotate this resource are indicated but the user is free to tag the resource by choosing a tag among them or using a new one. Our aim is to recommend resources which are annotated with tags suggested by association rules, in order to enrich user profiles with these resources (if they validate them). In other words, our aim is to enrich user profiles based on similarities between users and association rules and by doing so to increase the community effect when suggesting resources to a given user. Our approach comes from a new view on the community effect in folksonomies since it aims at automatically strengthening existing correlations between different members of online communities, without involving the user in this process. The fact of suggesting to each user some resources considered useful or interesting for him without him specifying explicit tags, this can significantly improve folksonomy-based recommendation systems, because the man-machine interaction and therefore the user effort are considerably reduced.

This paper is organized as follows: Section 2 is an overview of the main contributions related to our work. Section 3 is dedicated to the presentation of our approach. In section 4 we present and discuss the results of some experiences we conducted to measure the performance of our approach. Conclusion and future works are described in Section 6.

## 2 Related Works

### 2.1 Tag Recommendation

The general aim of tag recommendation systems is to help users choose the appropriate tags when annotating resources in order to increase the weights associated to each tag and so cross a step up to building a common vocabulary in these systems. Among the many works addressing this problem, let us cite that of Schmitz et al. [13] who showed how association rules can be adopted to analyze and structure folksonomies and how these folksonomies can be used for learning ontologies and supporting emergent semantics. Their approach consists in reducing the ternary dimension of a folksonomy by projecting it on a triadic context, and then in extracting association rules from this two-dimensional projection. An association rule  $A \Rightarrow B$  is interpreted in two ways: users assigning tags in

A to some resources often also assign tags in B to them or users labeling the resources in A with some tags often also assign these tags to the resources in B.

Another noticeable contribution is that of Jäschke et al. [7] who present a formal model and a new search algorithm called FolkRank, especially designed for folksonomies. It is also applied to find communities within a folksonomy and is used to structure search results. The authors have exploited the idea of the PageRank algorithm, which consists in considering a web page as important when there are several other pages connected to it. In FolkRank, a resource tagged by an important number of users with an important number of tags becomes important. The same type of relationships becomes true for tags and users. The idea is to create graphs, and to associate to each node of these graphs a weight indicating its importance.

Gemmell et al. [5] propose a tag-based recommendation method based on the adaptation of the K-nearest neighbor algorithm so that it accepts as input both a user  $U$  and a resource  $R$  and gives out a set of tags  $T$ . The interest of this approach is to orient users to use the same tags, and thus increase the chance of building a common vocabulary used by all members.

## 2.2 Resource Recommendation

The general aim of resource recommendation systems is to enrich the quantity and relevance of the recommended resources. Among the works addressing this problem, let us cite De Meo et al. [4] who propose an approach based on the principle of query expansion. The aim is to recommend resources to users searching by tags by enhancing their profiles represented by their tag-based queries. The principle of the approach is to enrich user profiles by additional tags discovered through the exploration of the two graphs TRG and TUG representing the relations respectively between tags and resources and between tags and users.

Let us note that, when compared to the works on tag recommendation, the principle is the same: extract the most appropriate tags. Most of the techniques performed in this process demonstrate their contributions for building a language more or less common between users of folksonomies. However the methods that are used to achieve this goal are different from one approach to another. Regarding the work of De Meo et al., we can say that the results obtained with their approach show that the idea of proposing a system of resource recommendation is pertinent: the rates of precision and recall are optimistic. However the fact of forcing the user to specify a list of tags in order to get resources can generate a cognitive overload and it obliges the system to focus on the participation of the user to perform its recommendation procedure. Moreover the technique that has been designed in this work does not take into account the semantics between tags, in particular it cannot distinguish between ambiguous tags and therefore it may recommend resources that will be rejected by the user because they are not close to his preferences.

### 2.3 Resolving Tag Ambiguity

According to Mathes [9], “the terms in a folksonomy have inherent ambiguity as different users apply terms to documents in different ways. There are no explicit systematic guidelines and no scope notes”. For this reason we are concerned by the problem of tag ambiguity in our approach of tag-based resource recommendation.

Among the most important contributions on resolving tag ambiguity or extracting the semantic links between tags in a folksonomy, we start with [10] where Mika has proposed to extend the traditional bipartite model of ontologies to a tripartite one where the instances are keywords used by the actors of the system in order to annotate web resources. In his paper, Mika focuses on social network analysis in order to extract lightweight ontologies, and therefore semantics between the terms used by the actors. In [6], Gruber stated that there is no contrast between ontologies and folksonomies, and so recommended to build an ontology of folksonomy. According to Gruber, the problem of the lack of semantic links between terms in folksonomies can be easily resolved by representing folksonomies with ontologies. Specia and Motta [14] in their turn have preferred the use of ontologies to extract the semantics of tags. The proposed method consists in building clusters of tags, and then trying to identify possible relationships between tags in the same cluster. The authors have chosen to reuse available ontologies on the semantic web in order to express correlations which can hold between tags. An attempt to automate this method has been done by Angeletou et al. [2].

Buffa et al. [3] present a semantic wiki reconciling two trends of the future web: a semantically augmented web and a web of social applications where every user is an active provider as well as a consumer of information. The goal here is to exploit the force of ontologies and semantic web standard languages in order to improve social tagging. According to the authors, with this approach, tagging remains easy and becomes both motivating and unambiguous. The niceTag project of Limpens et al. [8] is focused on the same principle: the use of ontologies in order to extract the semantics between tags in a system. In addition, the interactions among users and the system are used to validate or invalidate automatic treatments carried out on tags. The authors have proposed methods to build lightweight ontologies which can be used to suggest terms semantically close during a tag-based search of documents. Pan et al. [11] address the problem of tag ambiguity by expanding folksonomy search with ontologies. They proposed to expand folksonomies in order to avoid bothering users with the rigidity of ontologies. During a keyword-based search of resources, the set of ambiguous used terms is concatenated with other tags so as to increase the precision of the search results.

To sum up, most of the works aspire to bring together ontologies and folksonomies as a solution to resolve tag ambiguity and overcome the lack of semantic links between tags. Sure enough the approaches described in this section show that the social nature of resource sharing is not in contradiction with the possibilities offered by ontology-based systems. But the rigidity that characterizes

ontologies and the need for an expert who must control and organize the links between terms as in [6] seem a little cumbersome and too much expensive. Even the structures automatically extracted as in [10] still suffer from the ambiguity of concepts. Regarding the work of [14], we can say that the use of semantic web ontologies for extracting relationships between terms is not sufficient, because as the semantic web includes some specific domain ontology, that will push back the problem. Also the expertise of users which was introduced in [8] is characterized by the complexity of its exploitation. As a result we propose an approach of tag-based resource recommendation where we aim to resolve tag ambiguity without explicitly using ontologies.

### 3 Resource Recommendations based on Association Rules and Ontologies

#### 3.1 Association Rules: Basic Definitions

In data mining, learning association rules is a widely used method for discovering interesting relations among variables in large databases. [12] describes analyzing and presenting *strong rules* discovered in databases using different measures of interestingness. Based on this concept of *strong rules*, [1] introduces association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule  $\{onions, potatoes\} \Rightarrow burgers$  found in the sales data of a supermarket indicates that if a customer buys onions and potatoes together, he is likely to also buy burgers<sup>3</sup>. According to the original definition by [1], the problem of association rule mining is defined as follows:

**Definition 1.** Let  $I = \{i_1, \dots, i_n\}$  be a set of  $n$  binary attributes called items. A rule is an implication  $X \Rightarrow Y$  where  $X, Y \subseteq I$  and  $X \cap Y = \emptyset$ . The sets of items (itemsets)  $X$  and  $Y$  are called antecedent and consequent of the rule respectively.

To select interesting rules from the set of all possible rules in a database  $D = \{d_1, \dots, d_m\}$ , with each transaction in  $D$  containing a subset of items in  $I$ , two measures are commonly used: support and confidence.

**Definition 2.** The support  $supp(X)$  of an itemset  $X$  is the proportion of transactions in  $D$  which contain  $X$ .

**Definition 3.** The confidence  $conf(X \rightarrow Y)$  of a rule  $X \rightarrow Y$  measures the proportion of transactions in  $D$  that contain  $Y$  among those that contain  $X$ .  $conf(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$ .

Let us illustrate these notions on the following dataset.

<sup>3</sup> [http://en.wikipedia.org/wiki/Association\\_rules](http://en.wikipedia.org/wiki/Association_rules) [Retrieved 13 May 2011]

| Transaction ID | ItemSet             |
|----------------|---------------------|
| 1              | Bread, Cream, Water |
| 2              | Cream               |
| 3              | Bread, Cream, Milk  |
| 4              | Water               |
| 5              | Cream, Water        |

From this dataset, we can extract the rule  $Bread \Rightarrow Cream$  with a confidence  $conf(Bread \Rightarrow Cream) = \frac{supp(\{Cream, Bread\})}{supp(Bread)} = \frac{2/5}{4/5} = 1/2$ .

To be selected as significant and interesting, association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence. The process of generating association rules is usually split up into two separate steps: First, the minimum support constraint is applied to find all frequent itemsets in a database. Second, the minimum confidence constraint is applied among the rules involving these frequent itemsets. The quality of the extraction algorithm thus strongly depends on the values chosen by the user for the minimum support and minimum confidence, which adequacy is relative to the application.

### 3.2 Association Rules and Folksonomies

A folksonomy is a tuple  $F = \langle U, T, R, A \rangle$  where  $U$ ,  $T$  and  $R$  represent respectively a set of users, a set of tags and a set of resources, and  $A$  represents the relationships between the three preceding elements, i.e.  $A \subseteq U \times T \times R$  [10]. In our approach we consider a folksonomy as being a tripartite model where web resources are associated with a user to a list of tags. Therefore we have extracted three social networks from our folksonomy, which represent three different view-points on social interactions: one network relating tags and users, a second one relating tags and resources and a third one relating users and resources. We represent these social networks by three matrices  $TU$ ,  $TR$ ,  $UR$ :

- $TU = [X_{ij}]$  where  $X_{ij} = \begin{cases} 1 & \text{if } \exists r \in R, \langle u_j, t_i, r \rangle \in A \\ 0 & \text{otherwise} \end{cases}$
- $TR = [Y_{ij}]$  where  $Y_{ij} = \begin{cases} 1 & \text{if } \exists u \in U, \langle u, t_i, r_j \rangle \in A \\ 0 & \text{otherwise} \end{cases}$
- $UR = [Z_{ij}]$  where  $Z_{ij} = \begin{cases} 1 & \text{if } \exists t \in T, \langle u_i, t, r_j \rangle \in A \\ 0 & \text{otherwise} \end{cases}$
- $RU$ ,  $RT$  and  $UT$  are the transposed matrices of  $UR$ ,  $TR$  and  $TU$ .

This enables us to analyze the correlations captured from the different social interactions. We used Pajek, a tool which has already been used by Mika to analyze large networks [10].

To apply an association rule method to folksonomies, we represent each user in a folksonomy by a transaction ID and the tags he uses by the set of items which are in this transaction. The following table provides an illustrative example of a dataset of user tags.

| Transaction ID | Itemset               |
|----------------|-----------------------|
| $U_1$          | Computer, Programming |
| $U_2$          | Computer, Apple       |
| $U_3$          | Kitchen, Apple        |
| $U_4$          | Programming           |
| $U_5$          | Kitchen               |

Our goal is to find correlations between tags, i.e. to find tags which frequently appear together, in order to extract those which are not used by one particular user but which are often used by other users close to him in the social network. For example, let us consider a dataset in which it occurs that many users who use the tag *Software* also employ the tag *Java*. We aim at extracting a rule  $Software \Rightarrow Java$  so that we can enrich the profiles of users who employ the tag *Software* but not the tag *Java*, by the resources tagged with *Java*. Among the wide range of algorithms proposed to extract interesting association rules, we use the one known as *Apriori* [1].

Once the rules are extracted, our recommendation system proceeds as follows. For each extracted rule, we test whether the tags which are in the antecedent of the rule are used by the current user. If it is the case then the resources tagged with each tag found in the consequent of the rule are candidate to be recommended by the system. The effectiveness of the recommendation depends on the resolution of tag ambiguity, as explained in the next section.

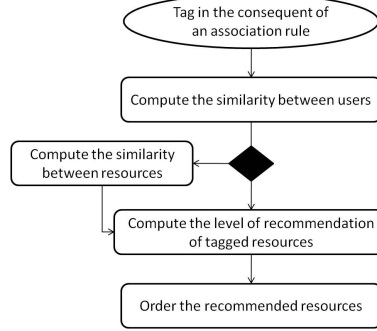
### 3.3 Resolving Tag Ambiguity in Recommendation

A tag can have several meanings, i.e. refer to several concepts. Therefore, a basic tag-based recommendation system would equally recommend resources relative to fruits or to computers for a user searching with the tag *apple*. The resolution of tag ambiguity is specially crucial in our approach where some tags which are used to recommend resources are not directly used by the user but deduced with association rules. To resolve the problem of tag ambiguity in recommendation, we propose to measure the similarity between users to identify those who have similar preferences and therefore adapt the recommendation to user profiles.

**Similarity between Users.** For each extracted association rule whose antecedent applies to the user searching for resources, we measure the similarities between this user and the users of his social network who use the tags occurring in the consequent of the rule. The resources associated to these tags are recommended to the user depending on these similarities.

To measure the similarity between two users  $u_1$  and  $u_2$ , we represent each of them by the vector of binary numbers representing all his tags (extracted from matrix  $UR$ ) and we calculate the cosines of the angle between the two vectors:

$$sim(u_1, u_2) = cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\|^2 \cdot \|v_2\|^2}$$



**Fig. 1.** Overview of the recommendation process

**Similarity between Resources.** To avoid the *cold start* problem which generally results from a lack of data required by the system in order to make a good recommendation, when the user of the recommendation system is not yet similar to other users, we also measure the similarity between the resources which would be recommended by the system (as related to a tag occurring in the consequent of an association rule) and those which are already recommended to the user.

To measure the similarity between two resources  $r_1$  and  $r_2$ , we represent each of them by the vector of binary numbers representing all its tags (extracted from matrix  $TR$ ) and we calculate the cosines of the angle between the two vectors.

**Levels of Recommendation.** Each resource recommended by the system is first associated an initial weight based on the similarities between users. Above a threshold fixed in  $[0, \dots 1]$ , we qualify the resource as *highly recommended*. Under this threshold, we consider the similarity between resources and we similarly highly recommend the resources which weights calculated on the product matrix  $RR = RT \times TR$  are above a given threshold. Otherwise, we compute the average ratio between the number of resources shared by the user of the recommendation system with his social network and the number of resources used by him. These numbers are given by the product matrix  $RR = RU \times UR$ . Above a threshold fixed in  $[0..1]$ , we qualify the resource as *highly recommended*; under this threshold, it is simply *recommended* or *weakly recommended* if the similarity is close to zero.

**Whole Process of Recommendation.** The activity diagram in Figure 1 gives an overview of the whole process of recommendation including the key steps described above to analyze existing interactions between the different elements of a folksonomy, especially those between users.

Let us note that our recommendation system is flexible, since the user can interact to accept or reject the recommended resources.

Let us consider the example of a folksonomy represented through the following three matrices  $TU$ ,  $TR$  and  $UR$ :

|             | $U_1$ | $U_2$ | $U_3$ | $U_4$ | $U_5$ |             | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ |       | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ |
|-------------|-------|-------|-------|-------|-------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| computer    | 1     | 1     | 0     | 0     | 0     | computer    | 1     | 1     | 0     | 0     | 0     | $U_1$ | 1     | 1     | 0     | 0     | 0     |
| kitchen     | 0     | 0     | 1     | 0     | 1     | kitchen     | 0     | 0     | 0     | 1     | 1     | $U_2$ | 0     | 1     | 1     | 0     | 0     |
| programming | 1     | 0     | 0     | 1     | 0     | programming | 1     | 1     | 0     | 0     | 0     | $U_3$ | 0     | 0     | 0     | 1     | 1     |
| apple       | 0     | 1     | 1     | 0     | 0     | apple       | 0     | 0     | 1     | 0     | 1     | $U_4$ | 1     | 0     | 0     | 0     | 0     |
|             |       |       |       |       |       |             |       |       |       |       |       | $U_5$ | 1     | 0     | 0     | 1     | 1     |

Let us now suppose that we have extracted the interesting association rule  $computer \Rightarrow apple$ . Matrix  $TU$  shows that tag *computer* is used by user  $U_1$ . Since *apple* is in the consequent of the rule, matrix  $TR$  shows that resources  $R_3$  and  $R_5$  are candidates for a recommendation to  $U_1$ . Matrix  $UT$  shows that *apple* is used by users  $U_2$  and  $U_3$ . Then we calculate the similarity between  $U_1$  and  $U_2$  and the similarity between  $U_1$  and  $U_5$ , based on matrix  $UU = UT \times TU$ :

|       | $U_1$ | $U_2$ | $U_3$ | $U_4$ | $U_5$ |
|-------|-------|-------|-------|-------|-------|
| $U_1$ | 2     | 1     | 0     | 1     | 0     |
| $U_2$ | 1     | 2     | 1     | 0     | 0     |
| $U_3$ | 0     | 1     | 2     | 0     | 1     |
| $U_4$ | 1     | 0     | 0     | 1     | 0     |
| $U_5$ | 0     | 0     | 1     | 0     | 1     |

$$sim(U_1, U_2) = cos(UU_1, UU_2) = \frac{(2 \ 1 \ 0 \ 1 \ 0) \times (1 \ 2 \ 1 \ 0 \ 0)}{\sqrt{4+1+0+1+0} \times \sqrt{1+4+1+0+0}} = \frac{4}{2 \times \sqrt{6}} = 0.66$$

$$sim(U_1, U_3) = cos(UU_1, UU_3) = \frac{(2 \ 1 \ 0 \ 1 \ 0) \times (0 \ 1 \ 2 \ 0 \ 1)}{\sqrt{4+1+0+1+0} \times \sqrt{0+1+4+0+1}} = \frac{1}{2 \times \sqrt{6}} = 0.16$$

$U_1$  and  $U_2$  show higher cosine similarity than  $U_1$  and  $U_3$ . Then, among the resources tagged with *apple*, namely  $R_3$  and  $R_5$  (see matrix  $TR$ ), those tagged by  $U_2$  are highly recommended to  $U_1$ : it is only the case of  $R_3$  (see matrix  $UR$ ).

$U_1$  and  $U_3$  are not similar. Then, among the resources tagged with *apple*, we compute the similarity of those tagged by  $U_3$ , namely  $R_5$ , with those already recommended by the system, namely  $R_3$ . It is based on matrix  $RR = RT \times TR$ :

$$sim(R_3, R_5) = cos(RR_3, RR_5)$$

$R_5$  and  $R_3$  are not similar. Then  $R_5$  is weakly recommended to  $U_1$ .

## 4 Experiments

In this section, we describe some experiments over two datasets and we analyze and discuss our results. We have developped a simple application with a convivial interface enabling the user to log in and get a personalized ordered list of recommended resources — depending on his tagging activity and social network.

#### 4.1 Experiment over a subset of the del.icio.us database

In order to validate our approach, we have conducted a first experiment with the del.icio.us database. Our test base comprises 207 tag assignments involving 21 users, 97 tags — some of which are ambiguous —, 92 resources — each having possibly several tags and several users. Our system has extracted a set of 17 association rules from the analysis of the dataset with a support equal to 0.5 and a confidence equal to 0.6. We have for example the rule *news*  $\Rightarrow$  *software*: 60% of the users using the tag *news* also use the tag *software*.

To demonstrate the validity of our approach, we have distinguished two classes of users: the first one contains the users who have employed ambiguous tags and the other one those who did not use those tags. This ambiguity of tags has been subjectively decided: for instance *apple* is ambiguous and *software* is not.

Not surprisingly, our experiment has showed that, by applying the extracted association rules, the resources associated to non ambiguous tags are highly recommended. It has also showed that, in the case of rules involving ambiguous tags, our system recommends to the user the resources which are close to his interests with a high level of recommendation and, on the contrary, those which are far from his interests with a low level of recommendation.

#### 4.2 Evaluation of our Recommendation System over an Experimental Dataset

To evaluate the quality of our recommendation system, we used the following three metrics: recall, precision and F1 metric. Precision measures the ability of the system to reject all the resources which are not relevant. It is given by the ratio between the number of the relevant resources recommended and that of the recommended resources. Recall measures the ability of the system to retrieve all the relevant resources. It is given by the ratio between the number of relevant resources recommended and that of all the relevant resources in the database. F1 is a combination of the two previous metrics; it is defined by the following formula:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Because the calculation of these metrics requires the knowledge of all relevant resources for each user in order to compare the results provided by our recommendation system and those which are actually preferred by each user, we have built a database by inviting 6 users to participate to an experiment.

We first made a prototype of a folksonomy in the form of a website. Then we asked the users to specify their preferred resources. Finally, we asked each user to tag a set of resources among 18 ones available on our website, by using free keywords. Based on this dataset, we extracted 10 association rules with a support equal to 0.5 and a confidence equal to 0.6. Afterwards we calculated the three metrics for each participant in our test, for each tag. The following table presents the average values of the metrics we obtained for our 6 users:

| <i>User</i>    | <i>Precision</i> | <i>Recall</i> | <i>F1</i> |
|----------------|------------------|---------------|-----------|
| $U_1$          | 0.66             | 0.57          | 0.61      |
| $U_2$          | 1.00             | 0.85          | 0.91      |
| $U_3$          | 0.60             | 1.00          | 0.75      |
| $U_4$          | 0.44             | 0.80          | 0.56      |
| $U_5$          | 0.66             | 0.66          | 0.66      |
| $U_6$          | 0.37             | 1.00          | 0.54      |
| <i>Average</i> | 0.62             | 0.81          | 0.67      |

These are quite encouraging results, showing that our approach of recommendation adapted to user profiles is truly able to help users when searching for resources.

### 4.3 Discussion

We have shown through our experiments that the use of data mining methods and tools has proved its effectiveness for folksonomy-based recommendation. The results of our data sample are optimistic and so we can say that the community effect which characterizes folksonomies has showed its power in users profiles enrichment. This enhancement can significantly help to improve recommendation systems. At the same time that our approach contributes to increase the weights associated to the relevant resources, it reduces tag ambiguity: every time when there are shared resources between two users, the system can avoid the trap of tag ambiguity in the research phase and it will test the similarity between resources when the users are not similar. The extraction of association rules is based on tags rather than on resources because we believe that tag popularity in folksonomies is greater than resource popularity and the meaning of tags in these systems is more significant than that of resources: the same resource can be used for many different purposes.

## 5 Conclusion and Future Works

In this paper, we have proposed a method to automatically enrich user profiles with a set of relevant resources, based on social networks and folksonomies. We exploit association rules extracted from the social relations in a folksonomy to recommend resources tagged with terms occurring in these rules by other users close in the social network. Our objective is to create a consensus among users of a same network in order to teach them how they can organize their web resources in a correct and optimal manner.

We have tested our approach on a small amount of data where we have obtained good results, but the validation of our approach still requires a larger sample set. In order to continue and improve our work, we aim to address the problem of scalability of our approach on larger databases. The measure of similarity we use is based on several products of matrices whose dimensions are the

numbers of resources and tags of a folksonomy. In real scenarios, these dimensions are usually too large. We are intending to explore matrix factorization and latent semantic analysis.

## References

1. R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. of the ACM SIGMOD Int. Conference on Management of Data, Washington, USA*, 1993.
2. S. Angeletou, M. Sabou, L. Specia, and E. Motta. Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report. In *Proc. of ESWC workshop on Bridging the Gap between Semantic Web and Web*, 2007.
3. M. Buffa, F. Gandon, G. Ereteo, P. Sander, and C. Faron. SweetWiki: A semantic Wiki. *Journal of Web Semantics*.
4. 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. *User Modeling and User-Adapted Interaction*, 20(1), 2010.
5. J. Gemmell, T. Schimoler, M. Ramezani, and B. Mobasher. Adapting k-nearest neighbor for tag recommendation in folksonomies. In *Proc. of 7th Workshop on Intelligent Techniques for Web Personalization & Recommender Systems (ITWP'09), Pasadena, California, USA, in conjunction with IJCAI 2009*, 2009.
6. T. Gruber. *TagOntology - a way to agree on the semantics of tagging data*. <http://tomgruber.org/writing/tagontology.htm>, 2005.
7. R. Jäschke, L.B. Marinho, Hotho A., L. Schmidt-Thieme, and G. Stumme. Tag recommendations in folksonomies. In *Proc. of 11th Eur. Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Warsaw, Poland*, volume 4702 of *LNCS*. Springer, 2007.
8. F. Limpens, F. Gandon, and M. Buffa. Collaborative semantic structuring of folksonomies. In *Proc. of IEEE/WIC/ACM Int. Conference on Web Intelligence (WI 2009), Milan, Italy*, 2009.
9. A. Mathes. *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*. <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>, 2004.
10. P. Mika. Ontologies are us: A unified model of social networks and semantics. In *Proc. of 4th Int. Semantic Web Conference (ISWC 2005), Galway, Ireland*, volume 3729 of *LNCS*. Springer, 2005.
11. J.Z. Pan, S. Taylor, and E. Thomas. Reducing ambiguity in tagging systems with folksonomy search expansion. In *Proc. of 6th Eur. Semantic Web Conference (ESWC 2009), Heraklion, Greece*, volume 5554 of *LNCS*. Springer, 2009.
12. G. Piatetsky-Shapiro. Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*. AAAI/MIT Press, 1991.
13. C. Schmitz, A. Hotho, R. Jäschke, and G. Stumme. Mining association rules in folksonomies. In *Proc. of IFCS 2006 Conference: Data Science and Classification, Ljubljana, Slovenia*. Springer, 2006.
14. L. Specia and E. Motta. Integrating folksonomies with the semantic web. In *Proc. of 4th Eur. Semantic Web Conference (ESWC 2007), Innsbruck, Austria*, volume 4519 of *Lecture Notes in Computer Science*. Springer, 2007.