Ambiguity in Tagging and the Community Effect in Researching Relevant Resources in Folksonomies

Samia Beldjoudi[1], Hassina Seridi[1], Catherine Faron-Zucker[2],

[1] Laboratory of Electronic Document Management LabGED, Badji Mokhtar University, Annaba, Algeria
[2] I3S, University of Nice Sophia Antipolis - CNRS, France
[1] {seridi,Beldjoudi}@labged.net
[2] catherine.faron-zucker@unice.fr

Abstract. In this paper we present an original approach to improve the search of web resources tagged with folksonomies. We argue that the semantics in folksonomies can be extracted from the force of the community effect and the social interactions between the members of a community. Our approach called SSF for Semantic Social Folksonomy tackle the problem of ambiguity between tags by considering user profiles and social interactions in folksonomies to reduce tags ambiguity.

Keywords: Folksonomies, Web 2.0, Semantic Web, Social Interactions, Tag Ambiguity.

1 Introduction

Before the revolution of Web 2.0, the creation of keywords or metadata associated to web resources was realized by the professionals of a domain in order to clearly specify its data. Therefore the cost in time and effort was too exaggerated. With the advent of the so called Social Web, users are allowed to create their own keywords which are known as tags and to tag web resources. The term *folksonomy* has appeared on the web to designate a new technology based on the triple (the *User*, the *Tag* and the *Resource*), *w*here the user is an actor in the system, the tag designates the keyword proposed by the user and the resource indicates the object that will be annotated by the user with the tag. According to Wikipedia a folksonomy is a system of classification derived from the practice and method of collaboratively creating and managing tags to annotate and categorize content. For instance, delicious[1] is one of the most famous such websites, based on folksonomies, which allow users to publish, tag and share resources. It represents an effective way to collaboratively manage bookmarks: it enables each user to add sites to his personal collection of links, to categorize those sites with tags, and to share them with other users. Folksonomies are used to tag any kind of multimedia web resources: Flickr[2] is a photo management and

---

[1] http://del.icio.us
[2] http://www.flickr.com

sharing web application; youtube[3] and dailymotion[4] are dedicated to the sharing of videos; myspace[5] and Odeo[6] to the sharing of music files.
These social practices continue every day to gain popularity. However, the use of folksonomies for finding information on the web poses some problems, among which the problem of tags ambiguity: a tag can designate several concepts, i.e. can have several meanings. For example when a user employs the tag "orange" to annotate a resource, the system will not understand whether he is talking about the "color" orange or the "fruit" orange. A related problem comes from the variations in writing the same concept. For instance, "cat" and "chat" both means the same concept (animal) in English and in French, but when a user searches resources annotated with the tag "cat", the system will not offer him those tagged with the word "chat" because it cannot understand that the tag "cat" is equivalent to the tag "chat". In addition, the tags that are freely chosen in these systems are likely to contain spelling errors and therefore make the retrieval of resources more doubtful than the metadata recovering from a lexicon examined by the professionals of information.

In this paper we present an original approach to improve the search of web resources tagged with folksonomies. We argue that the semantics in folksonomies can be extracted from the force of the community effect and the social interactions between the members of a community. Our approach called SSF for Semantic Social Folksonomy tackle the problem of tags ambiguity by considering user profiles and social interactions in folksonomies. More precisely, we intend to propose each user resources relevant to his profile in the folksonomy even when he uses ambiguous tags. The key of our approach is the similarities between the community's members.
Our paper is organized as follows: Section 2 presents an overview of the major contributions in this field. The design of our approach is presented in Section 3. In Section 4 the experimental phase is in order to measure the performance of our approach and also the obtained results are discussed. We conclude in Section 5.

2 Related Works

Despite the relative newness of folksonomies, there already are a lot of works in this domain; each one trying to resolve a specific problem. Most of these contributions are balanced between searching semantic relationships among folksonomy's terms, and helping users choosing appropriate terms when annotating their resources for increasing the weights associated to each tag. In this paper, we focus our state on the art on both the works relative to the reduction of the ambiguity of terms in folksonomies and those on the extraction of semantic links between terms in a folksonomy.

Mika [8] proposed to extend the traditional bipartite model of ontologies to a tripartite one: that of folksonomies, where instances are keywords used by the actors of the system in order to annotate web resources. In this article, Mika focuses on

[3] http://www.youtube.com
[4] http://www.dailymotion.com
[5] http://www.myspace.com
[6] http://odeo.com

social network analysis in order to extract lightweight ontologies and, the exploitation of the strength of these ontologies in order to explicit semantics between the terms used by the users (actors).

In another work, Gruber [4] argued that there is no contrast between ontologies and folksonomies, and therefore 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 by ontologies.

Specia and Motta [11] in their turn have preferred the use of ontologies to extract the semantics of tags. Their approach consists in building tags clusters, and then trying to identify the possible relationships between tags in each cluster. The authors have chosen to use ontologies available on the semantic web in order to express the correlations which can exist between tags. A more detailed attempt to mechanize this method is described in [1].

In the same trends; Buffa et al. [2] presented a semantic web application baptized SweetWiki reconciling two trends of the 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 ontologies and semantic web models to improve the notion of social tagging. According to the authors, tagging remains easy and becomes both motivating and unambiguous.

The niceTag project of Limpens et al. [5] is focused on this same principle: the use of ontologies to extract semantic links existing between tags in a system. In addition, this project has introduced the idea of exploiting interactions between users and the system to validate or invalidate automatic treatments carried out based on tags. The authors have proposed methods to build lightweight ontologies that can be used to suggest terms semantically close during the search of documents guided by tags.

Pan et al. [9] aimed at reducing the problem of ambiguity in tagging. They proposed to extend the search of tags in afolksonomy by using ontologies. They defended this principle of extension of the search in order to avoid bothering the users with the rigidity of ontologies. More precisely, they concatenated ambiguous terms with other ones in order to increase the precision of the results of a keyword-based search.

Scott and Hubermann [10] analyzed the structure of collaborative tagging systems as well as their dynamic aspects. Specifically, they discovered regularities in user activity, tag frequencies, kinds of tags used, bursts of popularity in bookmarking and a remarkable stability in the relative proportions of tags within a given URL. They also discussed why it is difficult to retrieve contents (which in our case play the role of resources) in a folksonomy. In particular they highlighted some problems like synonymy and polysemy.

Markines at al. [6] discussed how to extend and adapt traditional notions of similarity to folksonomies, and which measures are best suited for applications such as navigation support, semantic search, and ontology learning. The authors builded an evaluation framework to compare various general folksonomy-based similarity measures derived from established information-theoretic, statistical, and practical measures.

We also investigated the works relative to the recommendation of relevant resources. De Meo et al. [3] proposed to recommend a set of resources to enrich user profiles, a user profile being represented by the list of tags involved in his query. They expand queries to recommend resources to users performing a keyword-based search; in order to enhance their profiles. A user query is enriched with tags discovered through the exploration of the two graphs TRG (Tag Resource Graph) and TUG (Tag User Graph). According to the authors, this enrichment improves both the strategy of recommender systems and that of collaborative filtering and content-based filtering systems.

To sum up, most of the works relative to folksonomies aim to bring together ontologies and folksonomies as a solution to the tags' ambiguity problem and that of the lack of semantic links between tags. The approaches summerized in this section showed that the social nature of resources sharing is not in contradiction with the possibilities offered by ontology-based systems. But the rigidity which characterizes ontologies and the need of an expert who must control and organize links between terms as in [4] seems a little cumbersome and too much expensive. Even the structures extracted automatically as in [8] still suffer from the ambiguity of concepts. Regarding the work of [11], the use of semantic web ontologies for extracting relationships between terms is not sufficient, because the semantic web does not include enough specific domain ontologies and this will push the problem further. Also the expertise of users which was introduced in [5] is characterized by the complexity of its exploitation when we try as much as possible to avoid a cognitive overload, to limit the necessary effort for the formalization of this expertise.
Based on these observations, we propose an original approach which avoids the explicit use of ontologies.

3 The Semantic Social Folksonomy Approach

Our approach is to introduce both semantics and social aspects in forlksonomies in order to enable the user to retrieve relevant web resources tagged by the members of his community and close to his preferences. For this purpose, we propose to measure the similarity between the members of a community and to compare their preferences. Our goal is to provide a flexible method to reduce the problem of tags ambiguity and the lack of semantic relationships between terms in folksonomies, based on the similarities between users.

The innovative aspect of our approach is that it enables to surmount the lack of semantic relationships between tags and to suggest the user web resources close to his preferences when he performs a tag-based search. We assign to each suggested resource a degree of recommendation — *'strongly recommended'*, *'recommended'* or *'weakly recommended* ' — depending on the degree of similarity among users.

Let us note that our approach comes with a new point of view on folksonomies. The 'display's instability' of folksonomies is a corrolary to the fact that the results of a search procedure vary depending on the interests of each user. Moreover we answer the problem of ambiguity in tagging by retrieving resources depending on their similarity with user profiles.

Example 1. o better explain our approach, let us consider the following example (Table 1) where three users are each represented by a list of tags:

Table 1: A set of users with their tags

Users	*Tags*
U_1	computer, java, folksonomies, programming.
U_2	apple, fruit, strawberry, kitchen.
U_3	computer, apple, software.

Let us suppose that user U_1 wants to retrieve resources relative to the word (i.e. the tag) *'apple'*. In the current folksonomies, the output result will contain all the resources tagged with the *'apple'* i.e. those relative to food and computer, despite the fact that it is clear for a human reading U_1's tags, that his preferences are relative to computer and not to food. This can be summarized in the following table (Table 2):

Table 2: The key points of our approach: problem, assumption solution

Problem	Lack of semantic links between tags leading to a problem of ambiguity: a tag can refer to several concepts, i.e., a tag can have several meanings.
Hypothesis	Two resources tagged with the same term (tag) are similar, if they are used by users who share similar interests or similar tags[7]
Solution	Measure the similarity between users, to specify those who have similar preferences.

To make the system flexible, we propose to make it interact with the user to accept or reject the retrieved resources.

To avoid the "cold start" problem which generally results from a lack of data required by the system in order to make good quality recommendation, we propose to measure the similarity between resources when there are no similar users.

- **Example 2.** Let us now consider the following situation described in Table 3:

Table 3: A second set of users with their tags

Users	*Tags*
U_1	computer, java, folksonomies, programming.
U_2	apple, fruit, strawberry, kitchen.
U_3	computer, apple, software.
U_4	apple.

[7] Of course the degree of similarity between two resources will differ according to the number of the common tags between them.

If user U_1 searches resources tagged with 'apple'; our system will first propose him the resource corresponding to tag 'apple' which is used by the user U_3 with a 'very strong' level of recommendation because the two users U_1 and U_3 have similar preferences.

On the contrary the resources corresponding to the tag 'apple' which is used by user U_2 will be given to U_1 with a percentage 'low' level of recommendation because U_1 and U_3 do not share the same interests.

Now how should the system answer U_4 for whom it does not have much information about his interests? For such cases, we propose to measure the similarity between the resources corresponding to the tag 'apple' which is used by U_4 and the resources already proposed to U_1 with a high percentage i.e. those of $U_{3,}$. If the resources are similar, the system will propose them to U_1 with a 'very strong' level of recommendation, otherwise with a 'low' level of recommendation.

3.1 Formal description of the approach SSF

In this section we explain how the folksonomy is represented and how the community members, resources, and tags are exposed, how the relationships between these three elements are symbolized and how the system can recommend the best resources to each user according to his profile in the folksonomy.

Formally, a folksonomy is a tuple F = <U, T, R, A> where U, T, R represent respectively the set of users, tags and resources. A represents the relationship between the three preceding elements (i.e. $A \subseteq U \times T \times R$). In this approach a folksonomy is considered as a tripartite model where the instances are a web resources associated by user to a list of tags. So we can extract a tripartite hypergraph with three types of vertices which are tags, users and resources. However working with this kind of graphs is in general not easy to understand. Thus, we have preferred to extract three bipartite graphs with three social networks 'tag_user', 'tag_ressource' and 'user_ressource' from the folksonomy as in [7]. These three graphs are represented as follow:

-TU = <vertices, edges> noting that:

- Vertices = $T \cup U$ where $T = \{t_1, t_2, ..., t_n\}$ is the set of tags and $U = \{u_1, u_2, ..., u_m\}$ is the set of users.
- Edges = {t, u} represents the set of links between the different vertices.

-TR = <vertices, edges> noting that:

- Vertices = $T \cup R$ where $T = \{t_1, t_2, ..., t_n\}$ is the set of tags and $R = \{r_1, r_2, ..., r_k\}$ is the set of resources.
- Edges = {t, r} represents the set of links between the different vertices.

-UR = <vertices, edges> noting that:

- Vertices = $U \cup R$ where $U = \{u_1, u_2, ..., u_m\}$ is the set of users and $R = \{r_1, r_2, ..., r_k\}$ is the set of resources.
- Edges = {u, r} represents the set of links between the different vertices.

For better representation, the three graphs formalization is represented by a matrix form as follows:

-**TU= [X_{ij}]** where X_{ij} = { 1 if there is an edge between the tag i and the user j; 0 otherwise

-**TR= [Y_{ij}]** where Y_{ij} = { 1 if there is an edge between the tag i and the resource j; 0 otherwise

-**UR= [Z_{ij}]** where Zij = { 1 if there is an edge between the user i and the resource j; 0 otherwise

Three binary matrices have gotten and each one indicates possible links or correlations that can exist between every two elements selected from the folksonomy (Tags, Users); (Tags, Resources) or (Users, Resources).

- **Example 3.**The following three tables 4, 5 and 6 shown an example that will illustrate the relations between the different elements in a folksonomy:

Table 4: A matrix 'tag_user' (TU)

Users / **Tags**	U1	U2	U3	U4	U5
T1= computer	1	1	0	0	0
T2= kitchen	0	0	1	0	1
T3= programming	1	0	0	1	0
T4= apple	0	1	1	0	0

Table 5 : A matrix 'tag_resource' (TR)

Resources / **Tags**	R1	R2	R3	R4	R5
T1= computer	1	1	0	0	0
T2= kitchen	0	0	0	1	1
T3= programming	1	1	0	0	0
T4= apple	0	0	1	0	1

Table 6: A matrix 'user_resource' (UR)

Resources / **Users**	R1	R2	R3	R4	R5
U1	1	1	0	0	0
U2	0	1	1	0	0
U3	0	0	0	1	1
U4	1	0	0	0	0
U5	0	0	0	0	1

Noting that; the profile of each user has been represented by the list of his tags or his resources in the folksonomy.

- **Example 4.** From the Example 3 (Table 4), the following profiles are found:

Table 7: A table shows the users' profiles in the form of tags

Users Profiles	Tags
Profile $_{U1}$	computer, programming.
Profile $_{U2}$	computer, apple.
Profile $_{U3}$	cooking, apple.
Profile $_{U4}$	programming.
Profile $_{U5}$	kitchen.

3.2 The similarities between users

For the similarity between users computing, we suggest to use a measure that allows representing each user by a vector v_i designates a series of binary numbers defining the set of his tags or his resources. Thus, calculating the similarity between two users, for example u1 and u2, is using the cosines of the angle between their associated vectors v1 and v2 as shown in the formula (1):

$$cos(v_1, v_2) = \frac{v_1.v_2}{\|v_1\|^2 \|v_2\|^2} \quad (1)$$

The vectors corresponding to each user are computing by using two matrixes: **UR** x **RU** and **UT** x **TU** must be calculated. Where:

-RU is the transposed matrix of **UR.**

-UT is the transposed matrix of **TU.**

3.3 The similarities between resources

When users are not similar, we suggest measuring the similarity degree between resources in order to avoid "cold start" problem. And for that we suggest calculating the matrix **resource = RT x TR**. Where: **RT** is the transposed matrix of **TR**.

3.4 The search procedure

In the Figure 1 which is presented below, an activity diagram that will illustrate the search's procedure followed by our approach is introduced:

As it's shown in this Activity Diagram, the Levenshtein distance[8] is being used when the tag is not found in the folksonomy in order to measure the edit distance between

[8] By following the example of [10] and [5]

tags in the system, this allows our system to detect spelling variations and so it can offer to user presumably equivalent tags.

Example 5. In Table 8, there are some examples for the Levenshtein distance:

Table 8: Levenshtein Distance for some couples of tags

Tag1	Tag2	Distance of Levenshtein
Oracle	Orakle	0.83
User	Users	0.80
Stic2010	Stic-2010	0.88
Cat	Chat	0.75

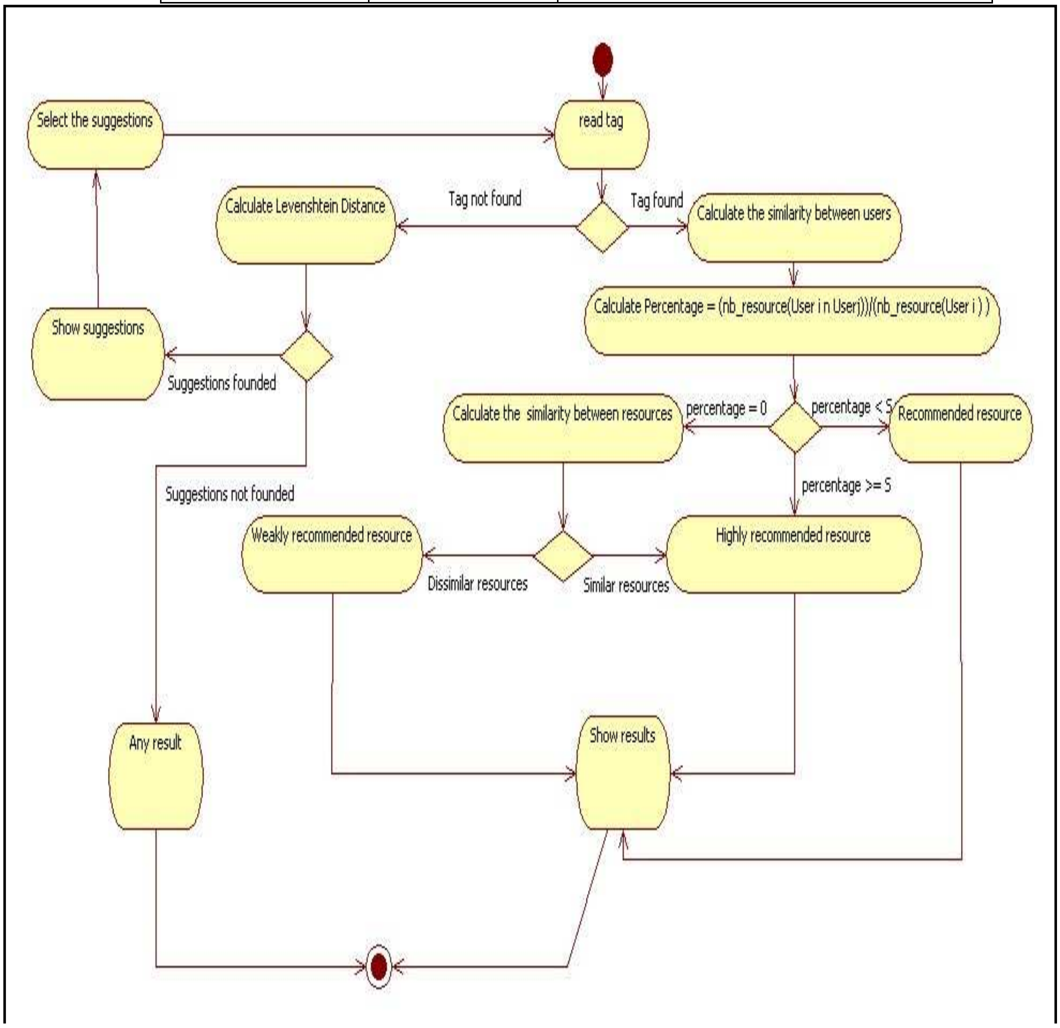

Fig 1: Diagram of Activity for the search procedure

For each resource recommended by the system a factor is proposed to indicate the percentage of its recommendation, i.e. the resource is 'highly recommended', 'recommended' or 'weakly recommended'. To achieve this classification, we have proposed to calculate the ratio between the number of resources used by the user himself (i.e. the one who does the research) and the number of shared resources between him and the others users. It must first select a threshold S ∈ [0, 1] to schedule the results. Where for each recommended resource, the following steps are performed:

1. Calculate $\frac{nb\_resource(User\ i \cap\ User\ j)}{nb\_resource(User\ i)}$ (2)

2. If $\frac{nb\_resource(User\ i \cap\ User\ j)}{nb\_resource(User\ i)} \begin{cases} >= S \text{ then the resource is highly recommended} \\ < S \text{ then the resource is recommended} \\ = 0 \text{ then calculate the similarity between resource} \end{cases}$

- **Example 6.** In table 9 the symmetric matrix User = UR x RU is represented:

Table 9: A symmetric matrix User =UR x RU

User \ User	U1	U2	U3	U4	U5
U1	2	1	0	0	0
U2	1	2	0	1	0
U3	0	0	2	0	1
U4	0	1	0	1	0
U5	0	0	1	0	1

In this matrix; for each user we have a vector as shown in Table 10:

Table 10: The correspondence between each user and his vector

User	Vector
U1	$v_{U1} = (2,1,0,0,0)$
U2	$v_{U2} = (1,2,0,1,0)$
U3	$v_{U3} = (0,0,2,0,1)$
U4	$v_{U4} = (0,1,0,1,0)$
U5	$v_{U5} = (0,0,1,0,1)$

For a threshold S = 0.5 for example. Supposing that after the similarity between users computing, the system provides the resource R_3 used by the user U_2 and which is annotated by the tag *'apple'* to user U_1. Noting in Table 8 that the resource number used by user U_1 is equal to 2, and the number of shared resources between the user U_1 and the user U_2 is equal to 1, so:

$$\frac{nb_{resource(u1 \cap u2)}}{nb_{resource(u1)}} = \frac{1}{2} = 0.5$$

Having $\frac{nb\_resource(u1 \cap u2)}{nb\_resource(u1)} = S$, and the resource is highly recommended.

4 Experiment

This section describes the used Dataset in the experiment and its treatment's method in order to fully understanding the given results. But first it should be noted that the experiment aim is to satisfy users when a research by keywords is done within folksonomies by offering them pertinent resources even in the presence of ambiguous tags. For this, we followed the steps described below:

4.1 Data Set

The dataset used in the test phase is described in this section followed by some analysis and discussion of the obtained results. The database of the website del.icio.us has been employed in this experiment representing the most used dataset for overall conducted experiments in folksonomies. Noting that we have selected a random set of data (i.e. a random sets of users, tags and resources) to well demonstrate the validity of our proposal. To fully shown the validity of SSF, two classes of users are randomly chosen: the first one contains the users who employed ambiguous tags and the second contains those who don't use these words but who can get them in the future. The users' number is equal to 21. For the set of tags used in this experiment we selected 97 different ones, among them there are some tags which are ambiguous. The used resources number is equal to 92; every resource can have multiple tags, and even multiple users. And in total we have 207 tag assignments. The table 11 summarizes the information about the data sample used in this experiment:

Table 11: The correspondent value for every element in the data sample

	Nb tags	**Nb users**	**Nb resources**	**Nb tag assignments**
Delicious dataset	97	21	92	207

4.2 Data treatment

The aim of this contribution is to generate a flexible technique, in other words, a technique which can be adapted to any situation. For this reason, we tried to automate it by using tools that can considerably reducing the site's administrator effort. Thus, after collecting all used data in the test phase, a tool of social network analysis called

"Pajek[9]" is used for extracting three networks 'Users_Tags', 'Users_Resources' and 'Tags_Resources'. The "Parek" usage greatly facilitating the three graphs generation and their corresponding matrixes, which makes this step automatically generated. The figures presented below (Figure 2, 3 and 4) shown the three generated networks from the data sample. Results of this step are used to calculate the similarities between users and between resources in order to detect the pertinent resources for each researcher (i.e. each user).

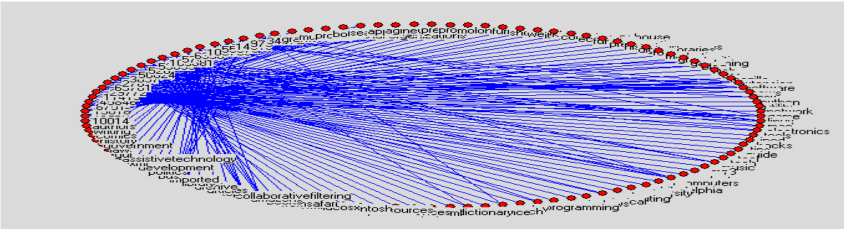

Fig 2: The network Users_Tags of our data sample with Pajek

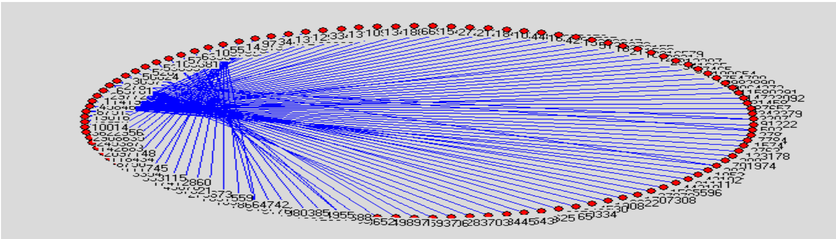

Fig 3: The network Users_Resources of our data sample with Pajek

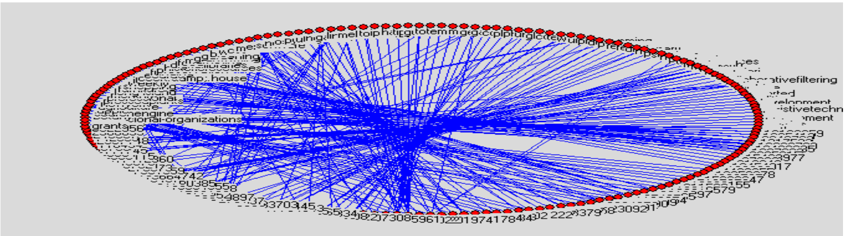

Fig 4: The network Tags_Resources de our data sample with Pajek

4.3 Results

Seeing that an information retrieval system allows giving to user documents that satisfy his need in information, and that a recommender system delivers documents to people basing on their profiles in long term. Sure enough the system that have been realized combines these two fields, because it allows at the same time proposing

[9] It's an analytical tool of social networks, used in [7].

resources that will help a user to satisfy his need for information, basing on its long-term profile. For this, a set of metrics are used in evaluating the proposal:

Precision: measures the system's ability to reject all not relevant resources to a query. It is given by the ratio of all relevant selected resources and the set of all selected resources.
Recall: measures the system's ability to retrieve all relevant resources inherent to a query. It's given by the relevant retrieved resources ratio and all relevant resources in the database.
Rates of precision and recall are given by formulations (3) and (4):

$$\textbf{Precision} = \frac{R+}{M} \quad (3) \qquad\qquad \textbf{Recall} = \frac{R+}{R} \quad (4)$$

Where *R*: is the number of relevant resources in the entire collection, *M*: is the number of resources selected by the system and *R+*: is the number of relevant resources selected by the system.

It should be noted that elements of each set (i.e. those of R), M and R+ are selected from our data sample as follow:

-The R number for each user is calculated according to its profile in the folksonomy. For example we took the case of user who is identified by this list of tags {java, computers, mac}, in the evaluation we supposed that preferences of this latter are similar to computing sciences field and not to the food when he did a search by the keyword *apple*. All resources that are close to the first domain are considered relevant to this user. Now in order to avoid the case when this user changes his opinion and likes to search resources related to second field (i.e. food) we give him in our proposal this resources at the end of list in order to avoid this kind of problems. In other words the approach SSF takes the probable changes in tastes of each user.
- The *M* number is calculated from resources number proposed by SSF approach.
- The number of *R+* is calculated by computing the number of R resources which are highly proposed to the user.

The metric F1: is a combination of the two previous metrics and is defined by the formula (5):

$$\text{F1} = \frac{2\,.\,Precision\,.Recall}{Precision + Recall} \qquad (5)$$

In order to evaluate this approach; the three metrics listed above are calculated for each user, and then the average of each metric in the system is calculated. The results are shown in Figure 5.

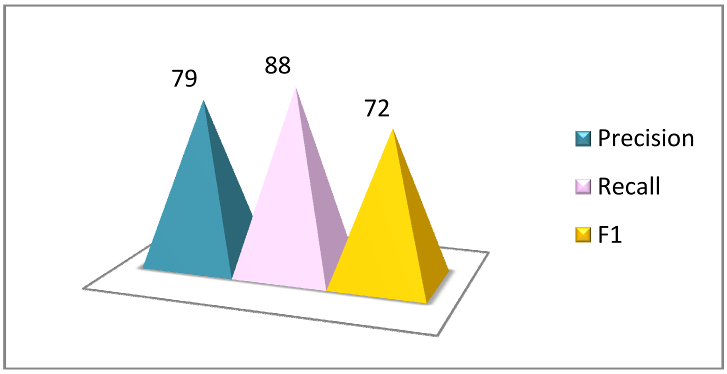

Fig 5: The average of the three metric: precision, recall and F1

4.5 Discussion

In SSF approach, we propose finding ways to equip in a "dynamic" manner the development of the sharing systems which trying to extract semantic in order to allow users to capture in the course of their daily tasks the social dimension of their use of some terms. It appears from this study highlights that can be summarized as follows: The consensus among users who have similar interests for using the same tags or the same resources plays an important role in eliminating the ambiguity problem. Also the increase in the weights of these terms or these resources has an influence on the emergence of semantic even when there are tags that can have several meanings. In fact, in the view of the obtained promising results in the experimental phase, it can be said that the use of analytical tools of social networks and also the collaborative filtering methods, even in the absence of ontologies, can help considerably to automate the nature of applications seeking to extract the semantic in the folksonomies. Indeed the obtained results the used data sample show that the technique SSF is succeeded in distinguishing between ambiguous tags.

Comparing SSF approach with others trying to discuss the problem of tags' ambiguity; for example the Pan's and al work [8], we can conclude that our results are very optimistic especially when we know that the proposed approach is flexible i.e. the result of the search's procedure will be changed according to interests and the profile of each user in contrary to other approach. Also concerning the works that aimed to recommending a set of resources for each user like in the approach cited in [3]. We find that the technique that has been designed doesn't take into account the semantic between terms, in particular it can't distinguish between the ambiguous tags and therefore it can provide to a user resources that can rejected by him because they are not close to his preferences.

5 Conclusion and Future Works

Our work presented in this paper contributes to answer the problem of tags ambiguity when searching for tagged resources. Our approach is based on taking into account social interactions between the members of a community to implicitly extract the semantics of the tags in a folksonomy. We therefore call it a Semantic Social Folksonomy. Our first experiments have shown the efficiency of our approach. However it is too early to conclude – the test phase should first cover the whole del.icio.us database. A short term issue will then be the validation of our approach on other large databases.

In the continuation of our work, we intend to enrich our approach of a Semantic Social Folksonomy by improving semantic network analysis (SNA) metrics.

References

1. Angeletou, S., Sabou, M., Specia, L., Motta, E.: Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report, Proc. Of ESWC workshop on Bridging the Gap between Semantic Web and Web (2007)
2. Buffa, M., Gandon, F., Ereteo, G., Sander, P. and Fabon, C.: SweetWiki: A semantic Wiki. J. Web Sem., 6(1), 84–97 (2008)
3. De Meo, P., Quattrone, G., Ursino, D.: A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy. User Model. User-Adapt. Interact. 20(1): 41-86 (2010)
4. Gruber, T.: Tag Ontology- a way to agree on the semantics of tagging data, http://tomgruber.org/ (2005)
5. Limpens, F., Gandon, F., Buffa, M. : Sémantique des folksonomies: structuration collaborative et assistée, IC (2009)
6. Markines, B., Cattuto, C,. Menczer, F., Benz, D., Hotho, A., Stumme, G.: Evaluating similarity measures for emergent semantics of social tagging. WWW 2009: 641-650. (2009)
7. Mika, P.: Ontologies are us: A unified model of social networks and semantics, In ISWC, volume 3729 of LNCS, p. 522–536: Springer (2005)
8. Pan, J., Taylor, S., Thomas, E.: Reducing Ambiguity in Tagging Systems with Folksonomy Search Expansion, In Proc. of the 17th International World Wide Web Conference (2009)
9. Specia, L. Motta, E.: Integrating Folksonomies with the Semantic Web, 4th European Semantic Web Conference (2007)
10. Scott A. Golder, Bernardo A. Huberman: Usage patterns of collaborative tagging systems. J.Information Science 32(2): 198-208 (2006)
11. Specia, L. Motta, E.: Integrating Folksonomies with the Semantic Web, 4th European Semantic Web Conference (2007)