# On Social Semantic Relations for Recommending Tags and Resources Using Folksonomies

A. Dattolo, F. Ferrara, and C. Tasso

Artificial Intelligence Lab, Department of Mathematics and Computer Science,
University of Udine, I-33100 Udine, Italy
{antonina.dattolo,felice.ferrara,carlo.tasso}@uniud.it

**Abstract.** Social tagging is an innovative and powerful mechanism introduced by social Web: it shifts the task of classifying resources from a reduced set of knowledge engineers to the wide set of Web users. However, due to the lack of rules for managing the tagging process and of predefined schemas or structures for inserting metadata and relationships among tags, current user generated classifications do not produce sound taxonomies. This is a strong limitation which prevents an effective and informed resource sharing; for this reason the most recent research in this area is dedicated to empower the social perspective applying semantic approaches in order to support tagging, browsing, searching, and adaptive personalization in innovative recommender systems. This paper proposes a survey on existing recommender systems, discussing how they extract social semantic relations (i.e. relations among users, resources and tags of a folksonomy), and how they utilize this knowledge for recommending tags and resources.

## 1 Introduction

Social Web applications provide users with a set of tools for creating, sharing, and promoting new content: users can easily leave the role of passive consumers of resources and become active producers (prosumers) of knowledge. This approach increases both the information on the Web and the number of available resources. Consequently, the growing number of resources prevents an effective access to them: a user needs to read the content of each resource for evaluating whether it is interesting for her.

An effective classification of the resources could greatly improve the access to knowledge. Although the manual process usually reaches high quality levels of classification for traditional document collections, it does not scale up to the enormous size of the Web, both in terms of cost, time, and expertise of the human personnel required [Dattolo et al. 2010].

In order to overcome this limitation, researchers proposed automatic classification tools based on ontologies, which add a semantic layer to the classification

Z.S. Hippe et al. (Eds.): Human - Computer Systems Interaction, AISC 98, Part I, pp. 311–326.
springerlink.com 

process. But, these tools are domain dependent due the obvious difficulties to build and maintain universal ontologies covering all possible information needs.

According to Mathes[1], a possible, cheap, and domain independent solution is provided by social tagging applications, which are not constrained to a specific informative domain and distribute the task of classifying document over the set of Web 2.0 users. While approaches based on ontologies use semantic information defined by knowledge engineers, in social tagging systems semantic relations emerge from the classification process exploited by Web 2.0 users, that tagging resources generate folksonomies. This means that on one hand people can freely choose tags in order to classify resources, on the other hand meaningful relations (socially defined) between pairs of tags can be extracted by analyzing the aggregated mass of tagged content.

The tagging activity does not require significant efforts since users can associate tags to resources without following specific rules: each user applies her personal classification which then can be used by others to find resources of interest. For this reason, social tagging applications have both private and public aspects [Golder and Huberman 2006]: users may apply tags for personal aims (typically they associate labels to resources in order to find them again), or they can enjoy/exploit the classification applied by other users and browse related documents.

However, due to the freedom of social tagging systems the classification process is not rigorous. This means that the classification proposed by a user may not be useful to other users and, for this reason, tools able to adapt and personalize the access to knowledge embedded in social tagging systems are fundamental to allow users to access information in a highly effective way.

In particular, it is assumed that in order to simplify the access to information in folksonomies, the following set of recommendation tasks should be addressed:

- *Users profiling*. Given a user, create a model to describe her interests according to her tagging activities. This is the basic task for being capable to provide personalized services.
- *Finding similar people*. Given a user, find a community of people with similar interests.
- *Finding similar resources*. Given a resource, find similar items with similar features (referring the same topic or informative context).
- *Finding domain experts*. Given a resource or a set of tags, find people who classify and share relevant information in a specific topic. They can help a user to locate resources related to her interests.
- *Supporting browsing*. Suggest tags for refining the search of contents according to a given information need.
- *Tag recommendation.* Given a resource, find a set of tags, which classifies the resource in a personalized or not personalized way.
- *Content recommendation*. Given a user, filter resources according to her user profile.

[1] http://www.adammathes.com/academic/computermediatedcommunication/folksonomies.html

The main aim of this paper is to present the state of the art related to tag and content recommendations. In order to face these tasks, the approaches proposed in literature basically exploit two phases: (a) mining social semantic relations (i.e. similarities among users, resources, and tags) analyzing socially annotated resources; (b) computing recommendations by means of social semantic relations.

So, this paper organizes the description of the state of the art describing first current techniques to extract social semantic relations from a folksonomy, and then, presenting methods to compute tag and content recommendations by means of social semantic relations.

More specifically, the rest of this paper is organized as follows: Section 2 introduces the reader to social tagging and recommender systems, while knowledge representation and data mining techniques for extracting social semantic relation from folksonomies are described in the Section 3; Section 4 and Section 5 deepen the discussion on the use of social semantic relations for recommending respectively resources and tags. Final considerations and a look to the future conclude the paper.

# 2 Background

In this section we present an overview of social tagging and recommender systems and describe how users apply tags, what are the limitations connected to the tagging process, and how recommender systems can be classified.

## 2.1 Social Tagging Systems

By using social tagging systems users share resources within a community, upload them, and mainly introduce personal classifications, applying on them, specific tags.

A *tag* is a term freely chosen by a user as significant for a resource; it represents a metadata describing the item; so it can be useful as a keyword to identify or to find again later a document. Tags are also the main mechanism used to browse and search new resources in social tagging systems. The collection of all the tag assignments performed by a user constitutes her *personomy*, while the collection of all personomies, present in a system, is called *folksonomy*.

Folksonomies [Dattolo et al. 2010] substitute traditional hierarchical taxonomies: while taxonomies are defined by a selected set of experts which categorize resources following a strict hierarchical predefined schema, folksonomies are flat spaces of keywords freely applied by communities of users. Thanks to the systematic work of experts, taxonomies are more rigorous than folksonomies because the classification is based on a well-defined vocabulary. On the other hand, users contributing to a folksonomy are free to add tags without using terms from a specific predefined vocabulary: this allows users to possibly use more than just one term for associating a same concept to a resource, providing in such a way a potentially very rich content to folksonomies.

Taxonomies are expensive because they require a systematic work by experts, which have to follow a well-defined set of procedures and rules. On the other hand, folksonomies are cheap because the work is distributed among Web 2.0 users.

However, the freedom associated to folksonomies causes some limitations, which may hinder an effective classification of resources:

- Due to the absence of guidelines, constraints, and control, users can exploit the same tag in different ways: for example, acronyms are a potential cause of *ambiguity,* or the same tag may be written using *different lexical forms* (e.g. 'photo', 'photos', 'web20', 'web_2', 'Web-2.0').
- It is frequent to find *synonymy*, i.e. different words which describe, more or less, the same concept, or *polysemy*, i.e. single words associated to various different meanings.
- Users classify documents using *different levels of expertise* and *specificity.* Since relations among tags are not defined, it is difficult to understand when distinct tags are referring the same concept.

Nevertheless, tags contain rich and potentially very useful, social/semantic information, and their nature can be understood by analyzing motivations/goals that usually lead a user to perform tagging [Dattolo et al. 2010; Golder and Huberman 2006]. Common purposes are:

- *Describe the content*. Tags may be used for summarizing the content of a resource.
- *Describe the type of the document.* Some users utilize tags for identifying the kind of document. A document may be classified according to its MIME type (as, for example, 'pdf' or 'doc') or taking into account the publication form (as, for example, 'article', 'blog', 'book', 'journal').
- *Describe features and qualities*. Adjectives (such as 'interesting', 'good', and so on) may be used for expressing opinions, emotions, or qualitative judges.
- *Associate people to documents*. Tags can report the authors of a document or people involved in a particular task or event. Moreover, tags such as 'my', 'my comments', 'mystuff', and so on are used to define a relationship between the resources and the tagger.
- Associate events to documents. Locations, dates, conferences acronyms are widely used for associating an event to a document.
- Associate tasks to documents. Some tags, such as 'mypaper', 'to read', 'jobsearch' reveal personal matters or engagements.

These possible motivations should be considered together with the following two further factors:

1. *Heterogeneity of users*. Taggers have different levels of expertise and goals. This has several consequences: classifications exploited by some user may be not understandable (or acceptable) for other users; different users may describe the content of a resource using distinct vocabularies; different users may have

different opinions about a topic; users may not have knowledge about people, events, or tasks associated to a resource by other users.
2. *Temporal changes*. Users' knowledge, motivations, and opinions may change over time. A tag used today for describing an item can be useless in the future: emotions and opinions of people may change; reputation of people evolves; a topic may be not any more interesting to the user.

Currently, tags are mainly used in social networks, social bookmarking applications, and Web 2.0 document sharing systems. Social networks (both general purpose ones, like Facebook or domain-specific ones, such as aNobii), allow users to apply tags for expressing opinions and for defining relationships among resources and people. Social bookmarking applications, such as Delicious, extend traditional bookmarking tools allowing users to upload, label, and access bookmarks from each computer connected on the Web, simplifying the process of content sharing among peers. Finally, Web 2.0 document sharing systems allow users to upload and share file with other peers. Remarkable examples of these systems are Flickr for photo sharing, YouTube for video sharing, Last.fm for music sharing, and BibSonomy for publication sharing. However, it is known that some authors propose a taxonomy of these applications, and classify applications according to *tagging rights* (who is allowed to tag), *tagging support* (what facilities are provided to simplify the tagging process), and *support to social interaction* among users.

### 2.2 Recommender Systems

The increasing volume of information on the Web is the main motivation for recommender systems: they support users during their interaction with large information spaces, and direct them toward the information they need; these systems model user interests, goals, knowledge, and tastes, by monitoring and modelling the feedback provided by the user. Such user feedback can be acquired by using appropriate ratings that quantify a relation between the user and an item: the ratings may be explicit, when they require the user evaluation, or implicit when they are automatically generated by the system in terms of measures, such as, for example, the time spent by a user on a Web page. By taking into consideration the ratings provided by a user, a recommender system defines a personalized order of importance for the set of available resources.

Several classifications of recommender systems have been proposed in the literature according, for instance, to the type of data that the user profile includes (e.g. demographic recommender systems) or the data structure used to represent the user profile (e.g. graph-based recommender systems). However, recommender systems can be classified into three classes of systems, on the basis of the algorithm utilized to produce recommendations: collaborative filtering, content-based, and hybrid recommender systems.

1. *Collaborative filtering recommender systems* filter resources using the opinions of other people; in turn, they may be differentiated in two approaches:

- *Model-based approaches*, which build a probabilistic model for predicting the future rating assignments of a user, on the basis of her personal history.
- *Memory-based approaches*, which use statistical techniques for identifying users with common behaviour (user-based approaches) or items evaluated in a similar way by the community (item-based approaches). In particular, user-based approaches look for people, called neighbours, similar to a given user, and then combine neighbours' feedbacks for generating a list of recommendations. On the other hand, item-based approaches look for resources similar to that the user liked, i.e. resources judged similarly by the community.

2. *Content-based recommender systems* analyze the past user activities looking for resources she liked; they model resources by extracting some features (for example, topics or relevant concepts) from documents. The user profile is then defined describing what features are interesting for the user. The relevance of a new resource for a user is computed by matching a representation of the resource to the user profile.
3. *Hybrid recommender systems* combine the results produced by collaborative and content-based recommender systems.

Two main recommendation tasks can be identified: (a) recommending content (i.e. suggesting documents, references, or URL's) and (b) recommending tags. In order to fulfil these tasks, the approaches proposed in the literature analyze and extract social semantic relations from folksonomies.

Next Section 3 provides a description of the techniques used to mine social semantic relations, while Sections 4 and 5 show how these relations are used to support, respectively, content and tag recommendation.

## 3 Mining Social Semantic Relations

Social tagging systems merge personal and social perspectives: the personal perspectives are embedded in personomies while social ones come from the union of all personomies; for this reason, personomies and folksonomies offer two distinct levels for mining social semantic relations.

### 3.1 Data Mining in a Folksonomy

A folksonomy is defined on a ternary relation which maps the tagging activities of all users: for each user, the ternary relation stores information about which tags have been applied on which resources. The ternary relation, which involves users, tags, and items, is the starting point to model knowledge, relationships and similarities in a folksonomy. However, mining similarities is not trivial because the ternary relation merges relations among objects of the same type as well among objects of different types. Two approaches have been proposed to handle this scenario:

- projecting the 3-dimensional space into lower dimensional ones [Dattolo et al. 2011];
- modelling the ternary relation by a 3-order tensor.

The projection of the ternary relation into two-ways relations (throwing away information about just one dimension) allows the system to extract the following three different matrices:

1. The *User-Resource* (*UR*) matrix. It describes the two-way relation between users and resources. Each row of this matrix is associated to a user which is described by a binary vector: if the user *u* tagged the resource *r* then the cell *UR(u,r)* is set to *1* (*0* otherwise).
2. The *Tag-Resource* (*TR*) matrix. It describes the two-way relation between tags and resources. Each row of the matrix, associated to a tag, is a vector, which counts how many times a tag has been applied on each resource.
3. The *User-Tag* (*UT*) matrix. It describes the two-way relation between users and tags. Each row of the matrix, associated to a user, is a vector, which counts how many times a user applied each tag.

These matrices describe relations among set of heterogeneous objects. Several notions of similarity between pairs of objects of the same type can be inferred by comparing two rows or two columns of the *UR*, *TR*, and *UT* matrices. The cosine and the Pearson similarities are commonly used to assess the similarity between two vectors. By means of this approach, given **a pair of users**, we can compute:

- *UR_user_sim*. Extracted from the *UR* matrix, this measure shows how much two users are similar according to the number of shared resources.
- *UT_user_sim*. Computed from the *UT* matrix, this measure specifies that two users are similar if they show a similar tagging behaviour.
  Given **a pair of resources** we can infer:
- *UR_resource_sim*. Computed from the *UR* matrix, it states that two resources are similar if they have been tagged by the same set of people;
- *TR_resource_sim*. Inferred from the *TR* matrix, it defines two resources as similar if they have been tagged in a similar way.
  Finally, given a **pair of tags** we can infer:
- *TR_tag_sim*. It is calculated from the *TR* matrix and states that two tags co-occurring frequently on the same resources share a common meaning;
- *UT_tag_sim*. We report this similarity just for the sake of completeness, as it is not really significant. It is computed from the *UT* matrix and states that two tags, used by the same user, share a common meaning. However, users may have several distinct interests and for this reason they may use tags which are not in any relation.

Unfortunately, the *UT*, *UR* and *TR* matrices used to discover similarities are sparse since each user labels only a small subset of all available resources and use only few tags. This sparsity can reduce the effectiveness of the methods developed to find social semantic relations from these matrices: for instance, *UR_re-source_sim* cannot be used to compare users who did not label the same resources.

Similarities inferred from the *UT*, *UR*, and *TR* matrices can be used to produce the *User-User* (*UU*) matrix, the *Resource-Resource* (*RR*) matrix and the *Tag-Tag* (*TT*) matrix in order to store respectively similarities between pairs of users, resources and tags. These matrices can be used to overcome the computational overhead needed to derive similarities in online scenarios and they represent also the starting point to develop graph-based mechanisms for extracting relevant information from a folksonomy. For instance, the *TT* matrix describes a graph where each node represents a tag and an edge connects two tags only if the similarity between them is greater than a certain threshold. For example, the PageRank algorithm and the HITS algorithm extract authoritative tags (i.e. tags semantically relevant) from this graph for a given set of input tags.

Similarly, the *RR* and the *UU* graphs can be built (using respectively similarities between resources and users) and then explored to discover new resources and new users for a given seed of resources or users.

The similarities among pairs of objects of the same type can be used to group together tags, users, and resources with similar properties. This task can be exploited, for instance, in order to create clusters of tags with a similar meaning, people with shared interests, or resources related to same topics or contexts.

Obviously, data mining techniques based on the projection of the 3-dimensional space into lower dimensional spaces lose some information. A different approach to model the ternary relation is to model the 3-dimensional space by a 3-order tensor. The HOSVD method, generalizes the SVD method to high dimensional spaces, and has been experimented to discover latent semantic association among users, tags, and resources.

### 3.2 Data Mining in a Personomy

A folksonomy collapses all users activities by combining all personomies, which include different personal interests and tagging strategies. On the other hand, a personomy contains information about just one user and can be analyzed to extract knowledge about the semantic relations that the user built during her tagging activities. More specifically, a personomy can be represented by a *Personal-Tag-Resource* (*PTR*) matrix, which stores information about how the user applied tags on resources. Starting form *PTR* matrix the *Personal-Tag-Tag* (*PTT*) matrix can be built and analyzed to find patterns in the user tagging activities. This matrix describes a co-occurrence graph where each node represents a tag and a weighted edge connects two tags only if the user applied these tags together. The weight associated to each edge is directly proportional to the number of times the two tags have been used together.

Graph clustering algorithms can be used to detect patterns in the user tagging strategy grouping sets of tags usually applied together to describe items: distinct group of tags can therefore reveal that the user is interested in different and disjoined topics.

## 4 Recommending Resources Using Tags

A system can recommend resources whenever it discovers some relevant ones (for example sending an email to the user) or whenever it receives a specific request by the *active user* (for example, a query). We call this approach *tag-aware recommendation.*

Given a query, the simplest approach is to give higher relevance to resources labelled by a large set of tags used by the active user or if the resource has been often associated to one or more tags applied by the active user.

In this way, popular resources become also the most relevant; however, although popularity is a good mean for assigning confidence to results, other parameters should also be considered, such as, for example, previous activities or habits of the user (for instance, how she usually apply tags or what resources she visited in the past).

In other approach there is suggestion that tags are a useful mean for understanding the relationship between a user and one or more resources. Following this idea, recently, several researchers proposed some attempts for providing personalized recommendations.

The following two subsections describe collaborative and content-based strategies to recommend resources using tags.

### 4.1 Tag-Aware Collaborative Recommender Systems

Tag-aware collaborative recommender systems extend collaborative filtering techniques using tags to model user interests and to produce personalized recommendation. In this context, tags have been used to achieve two possible goals:

1. Extending the classical collaborative filtering approach using tagging history for calculating similarities among users. Calculating user similarities by tags, the recommender system assumes that people with similar interests usually apply the same or similar tags.
2. Detecting adaptive neighbourhoods according to a specific topic or context defined by one or more tags. Each user may be interested in several topics and then she could tag resources referring distinct informative contexts. In order to obtain higher accuracy a recommender system can consider only users interested in a specific topic (i.e. users which used tags related to the specific topic) and resources associated only to the topic (i.e. resources labelled by a specific set of tags).

Both memory-based and model-based collaborative filtering approaches, aimed at comparing tagging histories to find similarities among users, have been proposed. For instance, Social Ranking is a memory-based recommender method that, given a user and a set of tags, computes a personalized ranking of resources. More specifically, Social Ranking extends the set of input tags including other similar tags by means of the *TR_tag_sim*: in this way, it discovers relevant tags for the user.

Then, it calculates a score for a resource according both to the relevance of tags associated it and to the *UT_user_sim* calculated between the active user and the other users who tagged the specific resource.

Alternatively, a model-based approach has been proposed in [Zhou et al. 2010], where the *PTT* matrix is considered to identify the distribution of user interests by clustering tags. Two distributions are then compared by means of the Kullback-Leibler divergence to assess the similarity among users.

TagiCoFi is another model-based recommender: it uses tags for facing the sparsity problem inferring some relationships among users and resources also if the users did not explicitly tagged the resource.

On the other hand, the idea of finding adaptive neighbourhood according to a topic or a context is exploited in two memory-based approaches described in [Nakamoto et al. 2007; Dattolo et al. 2009; Dattolo et al. 2011; Nakamoto et al. 2007], given a bookmark of a user, a *context* is defined by tags applied (by all users) on the specific resource. A context is used to filter documents and users: only users who applied tags in a given context and resources labelled by tags in the same context are considered to generate recommendations. In particular, the *UR_-user_sim* is used to evaluate the relevance of other users for a given context. Then, the relevance of a resource depends on the relevance of users who bookmarked it.

In [Dattolo et al. 2009] the authors use tags to distinguish different topics of interest for the active user: this task is performed by clustering tags with similar meanings, identified by using the *TR_tag_sim.* A cluster of tags allows the system to split resources tagged by the active user into different collections associated to distinct topics: a *topic of interest* is defined by a set of similar tags (applied by the active user) and the set of resources labelled by these tags. Given a topic of interest, the *UR_user_sim* and the *TR_tag_sim* are used to compute the relevance of new resources. In particular, resources labelled by tags, which are evaluated as more similar to the tags included in the topic, are considered more relevant than other resources as well as resources bookmarked by users more similar to the active user are more relevant than others.

Finally, in [Dattolo et al. 2011] the authors firstly operate on the disambiguation of tags and tag sets, by taking into account their synonyms, homonyms, and basic level variations; then they use the results of the disambiguation process to enhance both search and recommendation: in fact, tags sharing the same semantics are merged into one, while ambiguous ones are split according to their different contexts.

### 4.2 Tag-Aware Content-Based Recommender Systems

Generally speaking, tag-aware content-based recommender systems use tags in order to go deeper into a semantics-based approach. More specifically, they exploit tags for modelling interests, classifying documents, and comparing document representation to user profiles.

Meaningful examples of this trend have been described in [Shepitsen et al. 2008; De Gemmis et al. 2008; Shepitsen et al. 2008] the authors describe a recommender system representing users are on the rows of the *UT* matrix and resources on the columns of the *TR* matrix. Tag clustering is used to group tags with similar meanings. Each cluster of tags can be seen as a bridge between users and resources; in fact, looking at the user profile, it is possible to understand what tag cluster is relevant for the user and, on the other hand, the description of resources is used to detect resources relevant for a specific cluster. The recommendation algorithm uses as input a tag, a user profile and tag clusters, and produces an ordered set of items. In order to generate a personalized order of items, it computes, for each tag cluster, a score that is associated to both the cluster and the resources labelled by tags in the cluster. More specifically, the score assigned to the cluster depends on the number of times the active user applied the tags in the cluster (this step allows to personalize results), while the score of a resource depends on the number of times users associated to it tags which are in the specific cluster. By using this information, the relevance of a resource for a specific cluster is computed as the product of the score assigned to the cluster by the score assigned to the resource. Finally, given a resource, its relevance is computed by summing the relevance of the resource over all tag clusters. In [De Gemmis et al. 2008], the authors present a different approach where both the textual description of items and tags are used to build the user profile. This approach uses the synsets of Wordnet, structures defined as sets of words with a similar meaning and used for defining a semantic indexing of documents. A disambiguation strategy associates a synset to each word in the document looking at words that precede and follow it. Similarly, tags are also disambiguated using the textual content of the resource. In this way, a document is defined as a bag-of-synsets in opposition to the classical bag-of-words. Using this descriptive model, a Bayesian classifier considers the resources bookmarked by the user in order to learn about the synsets, which are relevant to her. Matching the synset representation of documents with the synsets in the user profile, the recommender system calculates a relevance value for each resource.

## 5 Recommending Tags

Tag recommendation is the second task we consider in this paper. This task can improve the usage of social tagging applications in several ways:

- Tag suggestions can *increase the probability* that people will assign many tags to resources. Users can just select one or more suggested tags instead of devising from scratch to meaningful tags.
- Tag suggestions can *promote a common vocabulary* among users. Proposing a well-defined set of tags, it become possible to reduce the problems connected to the absence of both guidelines and supervised methodologies for the tagging process.

The set of tags to recommend can be selected taking in account just metadata associated to the items (such as tags applied by other users, relevant keyphrases extracted from the text) or integrating the analysis of previous user tagging activities. Following these criteria, tag recommender systems can be divided into two classes:

1. *Not personalized tag recommender systems*. These systems select for each document a set of meaningful tags, ignoring the specific user's tagging habits. In this way, different users will receive the same suggestions for the same resource.
2. *Personalized tag recommender systems*. These systems suggest the set of the most relevant tags for a resource according to the specific user and her personal way to classify resources.

### 5.1 Not Personalized Tag Recommendations

Not personalized tag recommender systems do not follow the traditional organization of recommender systems because they do not build and maintain a user profile. Suggested tags can be extracted both from the content of specific resources and using tags applied by the whole community.

When tags are extracted from the textual content of a resource, well known techniques from information retrieval, natural language processing, and machine learning for classifying documents are applied. These approaches split the content of a textual resource into short textual slots, named *n*-grams (a sequence of *n* words), and then assess the relevance of each n-gram according to some criteria. Many example of this approach have been proposed in literature.

For example, the usage of the *tf*idf* metric to assess the relevance of *n*-grams. The same *tf*idf* metric is used also in KEA. It is based on a Bayesian classifier, to weight the terms, but KEA takes in account also their first occurrence. In order to filter the set of extracted keyphrases in an unsupervised and domain independent way, in [Pudota et al. 2010] the authors apply a POS (Part-Of-Speech) tagger. Then, the relevance of a keyphrase is computed according to the following set of features: frequency, first occurrence, last occurrence, and lifespan (the distance between the first and the last occurrence positions).

However, all these methods suggest only terms, which appear already in the document. For overcoming this limitation a semantic approach is needed. In [Baruzzo et al. 2009], the authors propose the use of ontologies. In this approach a set of keyphrases is extracted from the document and is used for browsing a domain ontology in order to find other, more abstract and conceptual terms. However, the performance of this approach depends on the quality of the available ontology.

User generated annotations can be also used to suggest tags. The simplest approach can suggest, for instance, the most popular tags for a resource. However, due to sparsity of social tagging systems there are resources tagged by only few people and for this reason more sophisticated methods have been proposed. AutoTag is a tag recommender system: it suggests tags for blog posts; this framework recommends tags following a three-step process: first, it selects resources similar to the starting document (according to the *tf*idf* measure) by retrieving the tags associated to these resources; then, it associates a weight to each tag according to the number of times the tag has been applied to the set of similar resources; and, finally, it suggests the top ranked tags. TagAssist[2] outperforms AutoTag thanks to a pre-processing phase, where the Porter's stemmer is used to compress the set of tags.

Other approaches consider that some users produce more meaningful and semantically rich classifications than others. FolkRank [Jäschke et al. 2006], for example, takes in account this feature by computing a ranking for users, resources, and tags through a PageRank-like algorithm. FolkRank models a folksonomy by a tripartite graph where tags, resources, and users are represented by three sets of nodes; edges link users to their tags and their resources, moreover, edges connect each resource to tags which have been used to classify the specific resource. The algorithm is based on the idea that a node of this graph is important if it is connected to many important nodes. So, the random surfer model of PageRank is used to spread weights over the tripartite graph in order to assign a weight for users, resources and tags.

### 5.3 Personalized Tag Recommendations

Personalized collaborative approaches evaluate the relevance of a tag considering the specific user tagging preferences.

Personalized collaborative strategies [Gemmell et al. 2009; Symeonidis et al. 2008] use people tagging strategies to detect the set of tags, which can be suggested to the active user.

In [Gemmell et al. 2009], the authors adapt the classical *K*-nearest neighbour algorithm to the task of generating a list of recommended tags: given a resource, a set of *K* neighbours is defined evaluating both the *UR_user_sim* and *UT_user_sim* over users which tagged the same resource. Tags assigned by similar users will be more relevant than others.

The ternary relation among tags, users, and items is modelled as a 3-order tensor in [Symeonidis et al. 2008]. Latent semantic analysis is performed on tensors to capture the latent association among users, resources, and tags. This approach builds a set of quadruplets *(u, r, t, likeliness)* where each quadruplet describes the probability that the user u will tag the resource *r* with the tag *t*.

Personalized content-based strategies analyze the relationship between the content of a resource and the tags applied by the active user in order to predict tags for new resources. Examples of this approach are provided in [Basile et al. 2007] and in [Musto et al. 2009].

The system proposed in [Basile et al. 2007] uses a Bayesian classifier for each tag employed by the user. Each classifier is trained using the textual content of documents tagged by the specific tag. In this way the text of a new document can be used for evaluating whether a tag can be suggested for that document.

STaR (Social Tag Recommender System) [Musto et al. 2009] is based on an approach similar to AutoTag (Section 4.1). The main difference is that STaR provides personalized tag suggestions. This framework collects two sets of documents similar to a starting resource: the set containing resources tagged by the active user and

[2] `http://infolab.northwestern.edu/media/papers/paper10163.pdf`

the set containing documents tagged by other users. Tags applied by the active user are weighted according to the similarity of the tagged resources to the starting one. In a similar way, a weight is assigned to tags applied by the other users. Finally, the two sets of tags are merged and a ranking of the tags is computed as a linear combination of the two scores associated to each tag.

## 6 Final Considerations and Future Work

In this paper, we analyzed current methods for finding social semantic relations (i.e. similarities among users, tags, and resources) in folksonomies and then we showed the ways in which these relations have been used to develop tag recommender systems and content recommender systems.

In order to extend collaborative and content-based approaches, social semantic relations have been introduced: in fact, on one hand, collaborative approaches can find similarities among users by looking at their tagging habits. On the other hand, content-based approaches use tags to model resources and build a user profile and then suggest resources, which appear relevant for the specific user profile. But, both the approaches have limitations mainly due to the ambiguity of tags and, for this reason, more semantics-based approaches are needed.

Some recommender systems use clustering for fighting redundancy of folksonomies [Shepitsen et al. 2008], but these approaches also show some open issue: during the time a user may apply the same tag for expressing different concepts; a tag may be in several clusters. Again, a deeper understanding of tags and resources can facilitate the disambiguation of tags.

Other works attempt to tackle the absence of semantic relationships among tags by associating a context to them [Nakamoto et al. 2007; Dattolo et al. 2009; De Gemmis et al. 2008; Dattolo et al. 2011]. However, it is hard to extract a context from some generic tag, such as, for example, ‘to read’, ‘my paper’, ‘job’.

A chance for reducing ambiguity is to support the manual tagging process using tag recommender systems: users do not need to define a concise description of the resource but they can just select among the suggested tags. Not personalized strategies ignore the heterogeneity of users suggesting the set of tags, which appear more representative for a resource according to a global measure. On the other hand, personalized approaches tailor the selection of recommended tags taking into account the user’s past tagging activities.

However, there is not evidence in the literature that personalized tag recommendation approaches outperform not personalized strategies, or vice versa, that not personalized techniques improve the user satisfaction. The evaluation of all these recommender systems is still an open challenge: results of different systems have been evaluated using different dataset and following different evaluation methodologies and procedures.

Other interesting lines of research on tags are actually ongoing. An interesting use of tags have been proposed in [Vig et al. 2009], where the authors introduce the concept of tagsplanations which are explanations based on community tags. Explaining the motivations at the basis of a recommendation improves the user satisfaction. Tagsplanations take into account two components: tag relevance and tag preference. Tag relevance refers to the representativeness of a tag for describing a resource, while tag preference defines the user's sentiment toward a tag. Another line of research is concerned with the extraction of basic semantic relations from folksonomies or the augmentation of social tagging with more ontology-like features [Baruzzo et al. 2009]. For example, Folk2onto [Sotomayor 2006] maps social tags (taken from delicious) to ontological categories (using a Dublin Core-based ontology) in order to classify and give a proper structure to the tagged resources. However, the task of associating semantic to tags, and extracting semantic relation among them is still far from a final solution.

## References

[Basile et al. 2007] Basile, P., Gendarmi, D., Lanubile, F., Semeraro, G.: Recommending smart tags in a social bookmarking system. In: Proc. of the Workshop on Bridging the Gap between Semantic Web and Web 2.0 the 4th European Semantic Web Conference, Innsbruck, Austria, pp. 22–29 (2007)

[Baruzzo et al. 2009] Baruzzo, A., Dattolo, A., Pudota, N., Tasso, C.: Recommending new tags using domain-ontologies. In: Proc of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Milan, Italy, pp. 409–412 (2009)

[Dattolo et al. 2009] Dattolo, A., Ferrara, F., Tasso, C.: Neighbor selection and recommendations in social bookmarking tools. In: Proc. of the 9th International Conference on Intelligent Systems Design and Applications, Pisa, Italy, pp. 267–272 (2009)

[Dattolo et al. 2010] Dattolo, A., Tomasi, F., Vitali, F.: Towards disambiguating social tagging systems. ch.20, vol. 1, pp. 349–369. IGI-Global (2010)

[Dattolo et al. 2011] Dattolo, A., Eynard, D., Mazzola, L.: An integrated approach to discover tag semantics. In: Proc. of the 26th Symposium on Applied Computing. Thungai University, Taiwan (2011)

[De Gemmis et al. 2008] De Gemmis, M., Lops, P., Semeraro, G., Basile, P.: Integrating tags in a semantic content-based recommender. In: Proc. of the 2nd ACM International Conference on Recommender Systems, Lausanne, Switzerland, pp. 163–170 (2008)

[Gemmell et al. 2009] Gemmell, J., Schimoler, T., Ramezani, M., Mobasher, B.: Adapting k-nearest neighbour for tag recommendation in folksonomies. In: Proc. of the 7th Workshop on Intelligent Techniques for Web Personalization and 21th International Joint Conference on Recommender Systems in Conjunction, Artificial Intelligence, Pasadena, California, USA, pp. 51–62 (2009)

[Golder and Huberman 2006] Golder, S., Huberman, A.: The structure of collaborative tagging systems. Journal of Information Science 32(2), 198–208 (2006)

[Jäschke et al. 2006] Jäschke, R., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Folkrank: A ranking algorithm for folksonomies. In: Proc. of FGIR 2006, Hildesheim, Germany, pp. 111–114 (2006)

[Musto et al. 2009] Musto, C., Narducci, F., De Gemmis, M., Lops, P., Semeraro, G.: STaR: a social tag recommender system. In: Proc. of the ECML/PKDD 2009 Discovery Challenge Workshop at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, CEUR Workshop Proceedings, CEUR-WS.org, Bled, Slovenia, vol. 497 (2009)

[Nakamoto et al. 2007] Nakamoto, R., Nakajima, S., Miyazaki, J., Uemura, S.: Tag-based contextual collaborative filtering. IAENG International Journal of Computer Science 34(2), 214–219 (2008)

[Pudota et al. 2010] Pudota, N., Dattolo, A., Baruzzo, A., Ferrara, F., Tasso, C.: Automatic keyphrase extraction and ontology mining for content-based tag recommendation. International Journal of Intelligent Systems, Special Issue on New Trends for Ontology-Based Knowledge Discovery 25(12), 1158–1186 (2010)

[Shepitsen et al. 2008] Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R.: Personalized recommendation in collaborative tagging systems using hierarchical clustering. In: Proc. of the 2nd ACM International Conference on Recommender Systems, Lausanne, Switzerland, pp. 259–266 (2008)

[Sotomayor 2006] Sotomayor, B.: Folk2onto: Mapping social tags into ontological categories (2006), `http://www.deli.deusto.es/Resources/Documents/folk2onto.pdf` (accessed October 29, 2010)

[Symeonidis et al. 2008] Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Tag recommendations based on tensor dimensionality reduction. In: Proc. of the 2nd ACM International Conference on Recommender Systems, Lausanne, Switzerland, pp. 43–50 (2008)

[Vig et al. 2009] Vig, J., Sen, S., Riedl, J.: Tagsplanations: explaining recommendations using tags. In: Proc. of the 13th ACM International Conference on Intelligent User Interfaces, New York, NY, USA, pp. 47–56 (2009)

[Zhou et al. 2010] Zhou, T., Ma, H., Lyu, M., King, I.: UserRec: A User Recommendation Framework in Social Tagging Systems. In: Proc. of the 24th AAAI Conference on Artificial Intelligence, Atlanta, Geogia, USA, pp. 1486–1491 (2010)