Folksonomies

Synonyms

Social Tagging, Social Indexing, Social Classification, Collaborative Tagging, Tagging System

Glossary

- **Resource** – any kind of web content, e.g., documents, hyperlinks, images or videos, which is uniquely addressable
- **Tag** – arbitrary text label associated with a resource
- **Tagging**
  1. process of annotating resources with tags by users
  2. a single assignment of one tag to a specific resource by a single user
  3. *pragmatics of tagging* - the way how users tag, e.g., in a more descriptive or categorizing manner
  4. *tagging semantics* - the meaning of tags, often dependent on the tag context
- **Personomy** – a subset of a folksonomy, containing all tag assignments from a specific user to all her resources
- **Tagging system** – software system that uses tagging for indexing, establishing a folksonomy
- **Semantic web** – concept for the world wide web to enable machine processing of information generated by and for humans; the goal is the “Web of Data”
- **Gold standard** – Highly reliable work, often manually generated. Used for evaluating experimental results
Definition

A **folksonomy** develops from a system – a so-called *collaborative tagging system* – that allows users to collaboratively annotate any kind of content with freely chosen keywords – so-called *tags* – which is one of the major features of the Web 2.0 (Vander Wal, 2005). The term *folksonomy* is a portmanteau of the words *folk* and *taxonomy*. *Folk* (or *Folks*) refers to the people within such systems. *Taxonomy* denotes the connection of folksonomies to classical indexing within the world wide web to finally create taxonomies. In literature two basic interpretations of the term folksonomy can be found: (i) tripartite graph-structure and (ii) folk-generated hierarchical taxonomy (Strohmaier et al, 2012).

A folksonomy can be seen as a data structure that is implemented in a collaborative tagging system and can formally be written as a tuple $F := (U, T, R, Y)$ where $U$ (users), $T$ (tags), and $R$ (resources) are finite sets. Users of collaborative tagging systems can annotate a resource with an arbitrary number of tags, which results in a ternary relation $Y$ between the three modules of folksonomies (Hotho et al, 2006). The relations between users, tags and resources are depicted in figure 1.

Folksonomies can be seen as a lightweight knowledge representation and they follow a bottom-up approach – in contrast to more formalized *Semantic Web* approaches. Tagging systems expose a converging use of the same vocabulary by their users and this leads to a reliance on *emergent semantics* (Steels, 1998). Collaborative tagging systems try to hide any complex methods from the users and thus there should be no formal overhead for the high percentage of non-expert users (Hotho et al, 2006).

Most of the time social tagging systems allow users to view and browse bookmarks and assigned tags of all other users inside the corresponding system. In such *broad folksonomies* it is also possible to annotate all objects inside the system with the own vocabulary. The complement is the *narrow folksonomy*, where users only annotate
Fig. 1. Structure of a folksonomy including the three elements – user, tags and resources – and the tag assignment relation.

their own content. The collection of all tag assignments of a user (see figure 1) is called **personomy**.

**Historical Background**

The process of tagging documents in a collaborative manner started in the 1990s, when people could add tags to their documents and objects which they published in Compuserve forum libraries. Nevertheless, a system operator would add more relevant terms from a controlled vocabulary whilst keeping the user’s annotations (Vander Wal, 2005).

In 1998, the World Wide Web Consortium (W3C) started the Annotea project. In this project, the W3C wanted to establish a standard to enhance document-based collaboration by using shared document metadata which were based on tags, bookmarks and other annotations. Web documents are associated with text strings without
the need to change the document itself. To share these annotations, one has to connect to a specific annotation server to see contributions from other users (Koivunen, 2005).

When Delicious started its service in 2003 as one of the first application implementing the ideas and features of the Web 2.0, the process of free tagging of resources by ordinary users quickly gained widespread popularity. Soon, Flickr also included the possibility of tagging while being still in early product development. Today, Flickr and Delicious are two of the most known tagging systems still in use. Tagging helped in finding new resources, rather through exploration and serendipity than searching and intent (Vander Wal, 2005).

In 2004, Gene Smith asked on the Information Architecture Institute’s mailing list, if there was a name for this kind of informal social classification (Smith, 2007). After a few responses, Eric Scheid suggested to use the term *folk classification*. Directly after that, Thomas Vander Wal mentioned the term *folksonomy* for the first time and described it as a “user-created bottom-up categorical structure development with an emergent thesaurus” (Vander Wal, 2005), thus coining the term *folksonomy*.

Since then, many applications were built on the folksonomy model. There are some examples of such web applications in the following section.

**Key Applications**

The type of resource in a collaborative tagging system differs from platform to platform. Table 1 names some of many other examples of social tagging systems that utilize the structure of a folksonomy together with their targeted type of resource.

Figure 2 shows the user interface of the collaborative bookmarking system *Bibsonomy* and depicts a sample procedure of adding a publication to the personal library of a user. A user has the possibility of adding an arbitrary number of freely chosen tags to the bookmark. If logged in, the interface shows a *tag cloud* – a way of illustrating
Table 1. List of some collaborative tagging systems together with the corresponding URL and targeted type of resource

<table>
<thead>
<tr>
<th>Name of system</th>
<th>URL</th>
<th>Type of resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>BibSonomy</td>
<td><a href="http://www.bibsonomy.org">http://www.bibsonomy.org</a></td>
<td>Websites and publications</td>
</tr>
<tr>
<td>Delicious</td>
<td><a href="http://delicious.com/">http://delicious.com/</a></td>
<td>Websites</td>
</tr>
<tr>
<td>CiteULike</td>
<td><a href="http://www.citeulike.org/">http://www.citeulike.org/</a></td>
<td>Publications</td>
</tr>
<tr>
<td>Flickr</td>
<td><a href="http://www.flickr.com/">http://www.flickr.com/</a></td>
<td>Photos</td>
</tr>
<tr>
<td>43Things</td>
<td><a href="http://www.43things.com/">http://www.43things.com/</a></td>
<td>Goals and hopes</td>
</tr>
<tr>
<td>Youtube</td>
<td><a href="http://www.youtube.com">http://www.youtube.com</a></td>
<td>Videos</td>
</tr>
</tbody>
</table>

tags of a folksonomy – on the right side. In this case the tag cloud consists of tags present in the vocabulary of the personomy of a specific user.

![BibSonomy User Interface](image)

**Fig. 2.** User interface of the social bookmarking system BibSonomy, showing a sample procedure to add a publication to the library and annotating it with an arbitrary number of tags.
Aspects of Folksonomies

In contrast to the more traditional taxonomy, tagging in folksonomy systems is neither exclusive nor hierarchical. In a taxonomy a user always has to decide about the keywords and the hierarchical structure she uses and this may also lead to an uncomfortable discovery of distinct folders when a user wants to re-find content. On the other hand, a user of a collaborative tagging system can choose an arbitrary number of tags for a resource and does not need to think about the hierarchy of the system. This results in an easier way to re-find items based on a simple keyword based search that shows all assigned resources with the corresponding tag in the output (Golder and Huberman, 2005).

This free annotating of content without many restrictions leads to an often uncontrolled vocabulary within such tagging systems. (Golder and Huberman, 2005) were one of the first to analyze the structure of collaborative tagging systems and they speak about the following three problems:

- **Polysemy** – a word can have many senses
- **Synonymy** – multiple words can have the same meaning
- **Basic level variation** – terms associated to an item can vary in their specificity (very general to very specific)

In order to counteract these problems there should be a general agreement about the vocabulary between users of such collaborative tagging systems. This can be achieved when the users come to an agreement about which tags to use for a certain case and this results in a stable system.

One of the first application to support the user while searching for information in tagging systems was a graph based ranking approach introduced by Hotho et al (2006). The authors introduce the so-called *FolkRank* – an adoption of the *PageRank* – that
exploits the inherent folksonomy structure. The emergent semantic patterns play an important role for Semantic Web applications (Cattuto et al., 2007, 2008). It was shown, that over some time, the users in a folksonomy tend to agree on a common vocabulary, where semantics can be extracted. Recommendation systems make extensive use of these emergent semantics, such as supporting users in the tagging process or even mildly refraining them to a defined vocabulary (see Jäschke et al., 2007). A big problem for emergent semantics lies in spamming. Through spam, the development of the emergent semantic patterns is negatively affected because of massive noise (see Krause et al., 2008). Furthermore, emergent semantics can help to construct hierarchical structures from such user generated metadata (Strohmaier et al., 2012).

Evaluation

The evaluation of folksonomies and algorithms that induce hierarchical structure from unstructured tagging data is a very challenging task which is based on the lack of golden standards. Strohmaier et al. (2012) have shown that algorithms specifically designed for social tagging systems outperform traditional hierarchical clustering techniques for folksonomy induction, i.e., for producing folksonomies out of flat tagging data. Two approaches to evaluation are usually pursued: (i) in semantic evaluation, the semantic quality is measured, whereas in (ii) pragmatic evaluation, the usefulness of folksonomies for certain tasks, such as navigation, is assessed. Similar to this pragmatic approach, Helic and Strohmaier (2011) have shown that it is possible to integrate tag hierarchies into the user interface of collaborative tagging systems in order to enable users to efficiently navigate tag categories. It is also possible to improve searching and information retrieval through emergent semantics.
Cross-References

Tag Clouds
Analysis and Mining of Tags, (Micro-)Blogs, and Virtual Communities

Recommender Systems
Web Science
Ontology Matching: Social/Collaborative
Semantic Annotation

References


Recommended Reading


Acknowledgements

This work is in part funded by the FWF Austrian Science Fund Grant I677 and the Know-Center Graz as well as by the DFG over the PoSTS project.