

## **Tag-Resource-User: A Review of Approaches in Studying Folksonomies**

**Jadranka Lasić-Lazić, Sonja Špiranec and Tomislav Ivanjko**

University of Zagreb, Faculty of Humanities and Social Sciences, Department of Information and Communication Sciences, Croatia

**Abstract:** This paper provides insight into main approaches to studying folksonomy based on its tripartite structure “tags-users-resources”. By conducting an exhaustive literature review in relevant scientific databases in LIS field, main approaches and methods in analysing folksonomies will be covered. The field of research is approached through three main focuses: (1) *tags* - covering approaches in analysing tag corpuses and structure; (2) *users* - carrying out studies for diverse communities of practice and (3) *resources* - covering research and methods dealing with the potential of folksonomy in providing new tools for information retrieval.

The area of researching folksonomies is still fairly new, so theoretical perspective and research methods are still being defined. In that light, this paper provides a review of the field of research, and a corresponding framework for the study of the field of folksonomy and social tagging systems.

**Keywords:** folksonomy, social tagging, collaborative tagging, literature survey

### **1. Introduction**

With the rise of Web 2.0, a new wave of user participation in creating and describing online resources instigated a new approach in knowledge representation – folksonomy. Folksonomy relies on the process of collaborative tagging, where many users add metadata in the form of keywords to shared content (Golder and Hubermann, 2006). The totality of these user-generated keywords (tags), gathered around any different platform or resource creates a folksonomy (Peters, 2009). When talking about formal models of folksonomy, the structure of folksonomy can be generally viewed through three different aspects: (1) *tags* – freely chosen user keywords that describe the resource; (2) *users* – those that perform the indexing, and (3) *resources* – items being described (Peters, 2009). Some authors add a fourth dimension thus defining a

folksonomy as a tuple including a set of users, set of resources and a set of tags including also ternary relations between those three sets (Mika, 2005; Hotho et al., 2006). Within this framework, different approaches are possible, where only one of the elements can be analyzed (for example, analyzing the linguistic characteristics of a chosen tag corpus) or, more often, the relationship between two elements is investigated (such as the relationship between tags and resources, identifying possible differences in tagging different types of resources). There are various approaches for studying social tagging and folksonomy presented in the literature where authors have tried to set the research framework. Here we will present two approaches relying on the tripartite structure *tags-users-resources*.

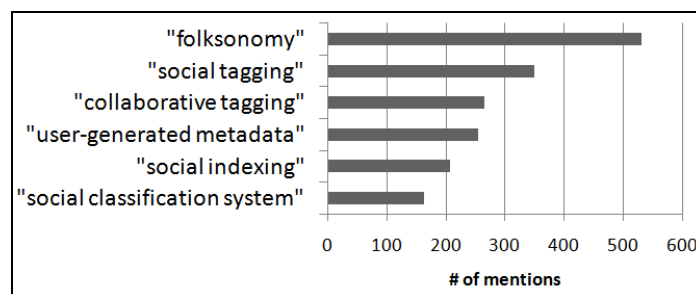
One of the first frameworks was set by Trant (2009) where the author proposes three complement fields of study: (1) studying tags – research focused on vocabulary control and evaluation, vocabulary analysis, finding structure in folksonomies and examining folksonomy as emergent ontologies; (2) studying tagging – includes studies on user tagging behaviour and motivation tags and (3) studying socio-technical systems – studies that describe systems and the inter-relationship of their parts, including the usage of tags as navigational/retrieval tools. When presenting the research on folksonomies in her book, Peters (2009) also uses the tri-partite disambiguation tags-users-resources, framing the overview in chapters regarding studying tags (tag distribution, tag categories,), studying users (cognitive skills, users' tagging behaviour, collective intelligence) and resource retrieval (tag recommender systems, traditional KOS vs. folksonomy).

Since both of these frameworks were set in the beginning of field development, there is a need to survey the literature and test these frameworks on empirical data, i.e. works published on the topic to see whether these approaches still cover main fields of study in social tagging and folksonomy and to propose possible refined or new categories.

## **2. Identifying key concepts**

Since the coining of the term folksonomy (Van der Wal, 2005) different authors proposed different terms for the concept. Peters (2009) list some of the most prominent ones found in the literature: "ethnoclassification", "communal categorization", "democratic indexing", "mob indexing", "social classification system", "social indexing", "user-generated metadata", "collaborative tagging", "social tagging" and "folksonomy".

In order to find out the most commonly used terms in the web space, a simple webometric analysis of the competing terms was conducted. By using the tool Webometric Analyst 2.0 (<http://lexiurl.wlv.ac.uk/>) as described by Thelwall (2013) a cross-domain analysis of web mentions was conducted and the most mentioned terms are presented in Figure 1.



**Figure 1 - Cross-domain web mentions of competing terms**

As we can see from the analysis the most widely used term is “folksonomy” with 531 cross-domain mentions, while terms not presented in Figure 1 had less than 50 cross-domain mentions. By examining the results in detail it became obvious that the term “social classification system” yielded such high results not because it relates to a concept found in the LIS literature but its origin derives from the field of sociology where it denotes a completely unrelated notion so it was clear it should be excluded from any literature search as it would generate a lot of false results. Also, the notion of “social indexing” is connected to the field of linguistics so those results included some false positives, but the majority of results corresponded to the concept researched. This analysis gave us a clearer picture on the most common terms used in the web space on which we could base further queries. Using these metrics we can conclude that the term “folksonomy” is the term being most commonly used to describe the concept in question with terms "social tagging", "collaborative tagging", "user-generated metadata" and “social indexing” also being significantly used.

### 3. Identifying key works in the field

Based on the results of a webometric study of key concepts a literature search with all the relevant terms related to the concept identified was undertaken. For the purpose of searching the relevant databases, a Boolean query (folksonom\* OR "social indexing" OR "social tagging" OR "user-generated metadata" OR "collaborative tagging") was created in order to include all the relevant concepts in the search. Four different sources were included in the search:

- Web of Science ([http://wokinfo.com/products\\_tools/multidisciplinary/webofscience/](http://wokinfo.com/products_tools/multidisciplinary/webofscience/)),
- SCOPUS (<http://www.info.sciverse.com/scopus>)
- LISTA (<http://www.ebscohost.com/academic/library-information-science-technology-abstracts-lista>)
- Google Scholar (<http://scholar.google.com>).

Table 1 shows the summarized data on search sources and search limits applied.

INTERFACE	DATABAS E	FIELDS	SUBJEC T AREA	RESUL T
SCOPUS	SCOPUS	TITLE- ABSTRACT- KEYWORDS	SOCIAL SCIENCE S	267
WOS	SCI- EXPANDE D, SSCI, AandHCI	TOPIC	LIS	119
EBSCOHO S T	LISTA	SUBJECT TERMS	LIS	71
POP 4 SOFTWARE	GOGLE SCHOLAR	ALL	ALL	1000+

**Table 1 - Sources included in the literature survey**

The first search through SCOPUS interface returned 275 articles because the results cannot be limited just to the field of LIS but only to the subject area of social sciences so the results also included works from other research fields within the social sciences, primarily linguistics, where the notion of “social indexing” is also used in different context so those articles were excluded from further analysis. In addition to searching the standard bibliographic databases in the field of LIS, Google Scholar was also included in order to provide a better insight into publications outside high impact journals, such as works published in conference proceedings, book and to include a wider journal base as suggested by Harzing (2008). The search of Google Scholar database was conducted using software Publish or Perish 4 (Harzing, 2007). Since the software is limited to processing the first 1000 results the total number of articles could not be calculated. Some studies have shown show that although Google Scholar’s ranking algorithm weighs heavily on articles’ citation counts (Beel and Gipp, 2009), top ranked articles are not necessarily those with the highest citation count. For that reason, the 20 top articles based on the Google Scholar ranking algorithm and top 20 articles based on the citation count were included in further analysis. When the entire duplicate articles (those appearing in multiple databases) as well as all the articles in languages other than English were removed, the final sample for literature survey was 345 articles from four different databases.

#### **4. Analysis result and framework construction**

The starting points for analysis were categories based on the tripartite *tags-users-resources* structure of folksonomies as presented in works of Trant (2008) and Peters (2009). All 345 articles were examined and assigned to a category. Since a large number of articles included multiple elements, an article could be

assigned to multiple categories. Based on the analysis and literature surveyed, initial categories were refined. Since there were a significant number of general introductory studies, a additional category called “General studies” was added and an article was assigned to that category when appropriate. In that case that was the single category it could be categorized into. That first category includes mainly works that are encompassing the field of folksonomy within the broader field of Web 2.0, where folksonomies are not the exclusive focus of research. The final framework for analysis and the number of papers in each category is presented in Table 2.

<b>CATEGORY</b>	<b>TOPICS</b>	<b>EXEMPLARY STUDIES</b>	<b>NO. OF WORKS</b>
<b>GENERAL STUDIES</b>	General Web 2.0 studies	Chua and Goh (2010); Quintarelli (2005); Warr (2008)	<b>40</b>
<b>EXPLORING TAGS</b>	Folksonomy models, structure, categorization, distribution, semantic and linguistic aspects	Golder and Hubermann (2006); Tonkin et al. (2006); Mathes (2004)	<b>113</b>
<b>STUDYING USERS</b>	Knowledge sharing and cooperation between experts Educational environment Access to learning objects New teaching approaches Users of different collections or services (e.g. Flickr, Delicious, etc.)	Kipp and Campbell (2006); Kamel Boulos, and Wheeler (2007); Morisson (2007)	<b>111</b>
<b>ENABLING ACCESS TO RESOURCES (INFORMATI</b>	Enhancing access Navigation, searching, personalizatio	Yi and Mai Chan (2009); Specia and Motta(2007);	<b>193</b>

<b>ON RETRIEVAL)</b>	n Enhancing description Complementi ng traditional KOS Extracting meaning Building ontologies	Mika (2005)	
--------------------------	--	-------------	--

**Table 2 - Approaches within the folksonomy framework**

When researching tags, studies were focused on examining the structure of folksonomies, tag categories and linguistic aspects (Golder and Hubermann, 2006; Spiteri, 2013) as well as tag distribution (Peters, 2008; Munk and Mork, 2007). By examining this element of the structure, authors produced models and frameworks for the field of research.

The next element of the tripartite structure included studies oriented to examining the value of folksonomies for different communities, from the educational environment (Vassileva, 2008; Kamel Boulos and Wheeler, 2007), expert collaboration (Lackes, Siepermann and Frank, 2009) or users of different services and systems (Rafferty and Hilderley, 2007).

The third element of the structure, the resources, encompassed studies carried out through the focus of information retrieval, focused on enhancing access and description of resources. A detailed analysis showed that within the field most of the studies are examining folksonomies as a new method of enhancing access to resources, by using tags to refine navigation interfaces (Bar Illan et al., 2012; Morisson, 2008) search results (McDonnell and Shiri, 2011), or as a basis for various recommender systems (Jaschke et al., 2009; Wetzker, Umbrath and Said, 2009). These studies are interested in exploring the ways in which user tags can improve the effectiveness of different systems.

The second main approach within the field of IR is examining the potential of user tags in enhancing resource description and complementing standard KOS methods. The main questions within this approach are focused on examining tags as viable alternatives to index terms assigned by professionals or to complement current indexing schemes by reflecting users' needs which are not found in existing indexing schemes (Yi and Chan, 2009; Rolla, 2009; Špiranec and Ivanjko, 2013). These studies are examining the role of traditional indexing tools and systems in the light of new user generated metadata.

The third main approach within the IR field is concerned with extracting meaning from folksonomies, connecting them with ontologies and Semantic Web technologies (Van Damme, Hepp and Siorpaes, 2007; Gruber, 2007; Specia and Motta, 2007). Within these approaches main questions explored are

connected with making explicit the semantics and meaningful relationships in social tagging systems, so they can be transformed to partial ontologies and used to represent knowledge in the Semantic Web environment.

## 5. Conclusions

The purpose of this study was to survey the literature and identify main concepts, topics and approaches in the study of social tagging and folksonomy in the field of Library and Information Science on the studies published in relevant databases.

It was shown that the term “folksonomy” is the term being most commonly used to describe the field of study with terms "social tagging", "collaborative tagging", "user-generated metadata" and “social indexing” also being significantly used. The literature search in the relevant databases has shown that the field of research is well established with a total of 457 articles found in WOS, SCOPUS and LISTA and over a 1000 articles in the Google Scholar database.

The categorization of approaches and research topics can be fitted in the tripartite *tags-users-resources* structure where each of the elements was extensively covered in the literature surveyed. Although these topics and patterns of research could be deduced from the literature survey to provide theoretical framework, it should be noted that the studies itself in most cases did not investigate only one element of the folksonomy structure but two or all three elements were included

When researching tags, studies were focused on examining the structure of folksonomies, tag categories and linguistic aspects as well as tag distribution. By examining this element of the folksonomy structure, authors produced models and frameworks for the field of research.

When examining the next element of the structure, the users, studies were oriented to examining the value of folksonomies in knowledge sharing and cooperation between experts, the educational environment, and the users of different collections or services.

It was shown that folksonomies are most commonly approached as a new method of knowledge representation, with the largest number of studies carried out through the focus of information retrieval, focused on enhancing access and description of resources and extracting meaning from social tagging systems.

Within this framework, studies were examining folksonomies as a new method of enhancing access to resources, by using tags to refine navigation, search result, or as a basis for various recommender systems. Other authors examined the potential of user tags in enhancing resource description and complementing standard KOS methods. The third main approach within the IR field is

concerned with extracting meaning from folksonomies, by making explicit the semantics and meaningful relationships in social tagging systems, so they can be transformed to partial ontologies and used to represent knowledge in the Semantic Web environment.

As far as different communities goes, the new user centred approach in organizing knowledge produced a number of studies from the field of libraries and museums, where folksonomies are examined as a tool to enhance access to digitized collections and library catalogues. On the other hand, lack of research connected with archives could be noted, where folksonomies weren't recognized as a viable approach.

This paper provided an exhaustive literature review and contributed to the field of folksonomy research by examining key terminology, concepts and research topics and by providing a theoretical framework based on empirical data. The field studying social tagging and folksonomies was shown as well established with a significant amount of publications on the subject published in most prominent LIS databases. Further research should build on these results and identify most influential papers and authors in the field and give further insight into the field of study.

#### References

- Chua, A.Y. K. and Goh, D. H., (2010). A study of Web 2.0 applications in library websites. *Library and Information Science Research*, Vol. 32, No. 3, 203–211.
- Bar-Ilan, J. et al., (2012). Tag-based retrieval of images through different interfaces: a user study. *Online Information Review*, Vol. 36, No. 5, 739-757.
- Beel, J. and Gipp, B., (2009). Google Scholar's Ranking Algorithm: The impact of articles' age: an empirical study. *Proceedings of the 6th International Conference on Information Technology: New Generations*, 160-164.
- Golder, S. A. and Huberman, B.A., (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, Vol. 32, No. 2, 198-208.
- Gruber, T., (2007). Ontology of folksonomy: a mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems*. Vol. 3, No. 1, 1-11.
- Harzing, A.W., (2007) Publish or Perish. [Online]. Retrieved on 12 April 2013 from <http://www.harzing.com/pop.htm>
- Harzing, A.W., (2008). Google Scholar: a new data source for citation analysis. [Online]. Retrieved on 12 April 2013 from [http://www.harzing.com/pop\\_gs.htm](http://www.harzing.com/pop_gs.htm)
- Hotho, A. et al., (2006). Information retrieval in folksonomies: search and ranking. *Proceedings of the 3rd European conference on The Semantic Web: research and applications*, 411-426.
- Jäschke, R. et al., (2009). Testing and evaluating tag recommenders in a live system. *Proceedings of the 3rd ACM conference on Recommender systems*, 369-372
- Kamel Boulos, M. N. and Wheeler, S., (2007). The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. *Health Information and Libraries Journal*, Vol. 24, No.1, 2-23.
- Kipp, M. E. I. and Campbell, D., (2006). Patterns and inconsistencies in collaborative tagging systems : an examination of tagging practices. *Annual General Meeting of the American Society for Information Science and Technology*. [Online]. Retrieved on 12 April 2013 from <http://hdl.handle.net/10760/8720>



Lackes, R., Siepermann, M., and Frank, E., (2009). Social networks as an approach to the enhancement of collaboration among scientists. *International Journal of Web Based Communities*, Vol. 5, No. 4, 577-592.

Lu, C., Park, J. R. and Hu, X., (2010). User tags versus expert-assigned subject terms: a comparison of LibraryThing tags and Library of Congress Subject Headings. *Journal of Information Science*, Vol. 36, No. 6, 763-779.

Mathes, A., (2004). Folksonomies – cooperative classification and communication through shared metadata. [Online]. Retrieved on 12 April 2013 from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>

McDonnell, M. and Shiri, A., (2011). Social search: a taxonomy of, and a user-centred approach to, social web search. *Program: electronic library and information systems*, Vol. 45, No. 1, 6-28.

Mika, P., (2005). Ontologies are us: a unified model of social networks and semantics. In *Proceedings of the 4th International Semantic Web Conference*, 522–536

Morrison, P. J., (2007). Folksonomies: why are they tagging, and why do we want them to? *Bulletin of the American society for information science and technology*, Vol. 34, No. 1, 12-15.

Morrison, P. J., (2008). Tagging and searching: search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing and Management*, Vol. 44, No. 4, 1562-1579.

Munk, T.B. and Mork, K., (2007). Folksonomy, the power law and the significance of the least effort. *Knowledge organization*, Vol. 34, No. 1, 16-33.

Peters, I. and Stock, W. G., (2010). “Power tags” in information retrieval. *Library Hi Tech*, Vol. 28, No. 1, 81-93.

Quintarelli, E. (2005). Folksonomies : power to the people. *ISKO Italy-UniMIB meeting*. [Online]. Retrieved on 12 April 2013 from <http://www.iskoi.org/doc/folksonomies.htm>

Rafferty, P. and Hilderley, R., (2007). Flickr and democratic indexing: dialogic approaches to indexing. *Aslib Proceedings*, Vol. 59, No. 4/5, 397-410.

Rolla, P. J., (2009). User tags versus subject headings. *Library Resources and Technical Services*, Vol. 53, No. 3, 174-184.

Specia, L., and Motta, E., (2007). Integrating folksonomies with the semantic web. *The semantic web: research and applications*. Springer, Berlin, 624-639.

Špiranec, S. and Ivanjko, T., (2013). Experts vs. novices tagging behavior: an exploratory analysis. *Procedia - Social and Behavioral Sciences*, Vol. 73, No. 27 456-459.

Spiteri, L. F., (2013). The structure and form of folksonomy tags: the road to the public library catalog. *Information technology and libraries* Vol. 26, No. 3, 13-25.

Thelwall, M., (2013.). Webometrics and Social Web research methods. [Online]. Retrieved on 12 April 2013 from <http://www.scit.wlv.ac.uk/~cm1993/papers/IntroductionToWebometricsAndSocialWebAnalysis.pdf>

Tonkin, E., et al., (2008). Collaborative and social tagging networks. *Ariadne*, 54.[Online]. Retrieved on 12 April 2013 from: <http://www.ariadne.ac.uk/issue54/tonkin-et-al>

Trant, J. (2009). Studying social tagging and folksonomy: a review and framework. *Journal of Digital Information*, Vol. 10, No. 1. [Online]. Retrieved on 12 April 2013 from: <http://journals.tdl.org/jodi/index.php/jodi/article/view/269>

Van Damme, C., Hepp, M. and Siorpaes, K., (2007). FolksOntology: an integrated approach for turning folksonomies into ontologies. Proceedings of the *ESWC workshop Bridging the gap between Semantic Web and Web 2.0*, 57-70.

Vander Wal, T., (2005). Explaining and showing broad and narrow folksonomies. [Online]. Retrieved on 12 April 2013 from <http://www.vanderwal.net/random/entrysel.php?blog=1635>.

Vassileva, J., (2008). Toward social learning environments. *IEEE Transactions on Learning Technologies*, Vol 1, No. 4, 199-214.

Warr, W. A., (2008). Social Software: fun and games, or business tools? *Journal of Information Science*, Vol. 34, No. 4, 591-604.

Wetzker, R., Umbrath, W. and Said, A., (2009). A hybrid approach to item recommendation in folksonomies. Proceedings of the *Workshop on Exploiting Semantic Annotations in Information Retrieval*, 25-29.

Yi, K. and Mai Chan, L., (2009). Linking folksonomy to Library of Congress Subject Headings : an exploratory study. *Journal of Documentation*, Vol. 65, No. 6, 872-900