The Knowledge Engineering Review, Vol. 30:5, 521–544.
doi:10.1017/S0269888915000120

Quality-protected folksonomy maintenance approaches: a brief survey

FOUZIA JABEEN and SHAH KHUSRO

Department of Computer Science, University of Peshawar, Peshawar 25120, Pakistan;
e-mail: fouzia.jabeenphd@upesh.edu.pk, khusro@upesh.edu.pk

Abstract

Folksonomy gives liberty to its users to freely assign chosen keywords as tags, and this is the main reason behind its popularity. Apart from freedom, this system also reflects the collective intelligence of the crowd. However, this freedom and liberty can degrade quality of the folksonomy. It is required that quality of the folksonomy must remain consistently excellent and does not degrade with the passage of time. This is a survey paper, in which we present a brief survey of the research efforts intended to maintain a quality-protected folksonomy. We have organized our paper by looking at the problem from four aspects namely selection of quality tags, tag management features provided by folksonomy applications, folksonomy cleaning and interoperability of tags across platforms. We conclude our review with some of the interesting research topics, which need to be explored further. Our conclusion will be relevant and beneficial for engineers and designers who aim to design and maintain a quality-protected folksonomy.

1 Introduction

Folksonomy is the other name for user-generated metadata. From implementation point of view, folksonomy is the complete group or set of tags. Tag consists of one or two keywords that users of a shared content management system assign to individual pieces of content for grouping and categorizing them, and for long-term retrieval of those contents. The public nature of these tags is a key factor; as it enables the users to find out how others had used the same tags in categorizing their own content. And in addition the users also view the terms others have added in the folksonomy (Sturtz, 2004).

As users contribute towards it, it allows gathering of crowd intelligence. But if we look at the negative side, the users can also give incorrect tags, which will make folksonomy a mess constituting of erroneous words and also degrades its usefulness. Table 1 gives an overview of the types of erroneous tags encountered in the folksonomy identified by Solskinnsbakk and Gulla (2001), Guy and Tonkin (2006), Nandipati (2011). Cattuto *et al*. (2007) showed that how fast the growth of vocabulary takes place with the passage of time in Del.icio.us. By closely examining the growth, question arises regarding the quality of the vocabulary and about the worthiness of the folksonomy.

The limitations and drawbacks of folksonomies have been well documented; what is required is the in-depth analysis of efforts for keeping consistently well-maintained folksonomy in terms of quality. In this paper, we have investigated the research efforts to solve the said problem from four aspects namely selection of quality tags, tag management features provided by folksonomy applications, folksonomy cleaning and interoperability across the platforms. Existing related work by Peters and Weller (2008) covers two aspects while one aspect is covered in Dotsika (2009). To the best of our knowledge none of the reviews cover all four aspects, making our paper coverage more as compared with others.

Table 1 Erroneous tags in folksonomies

Tags	Description	Examples
Inflectional	Single plural use of same word	Horse,horses.flower,flowers
Orthographic	Spelling variation	Color,Colour. Semantic_web, Semweb
Synonyms	Tags with same meaning	USA,US. Hardisk,hardrive. Mac, Macintosh,Apple
Abstraction	Generalize, specialize and aggregation	Ui,gui. Programming,Javascript, perl
Association	Generic semantic relation	Agriculture,permaculture
Misspelled tags	Spell mistakes	libray,libary
Tags with no spaces/compound/ word grouping	Badly encoded tags, such as unlikely compound word groupings	TimBernersLee. 17thjuly. April11
Personalized tags	Personal tags that are without meaning to the wider community	Mydog
Single-user tag	Single-user tags that appear only once in the database	Billybobsdog
Symbols in tags	Characters other than alphabets	# in tag
Nonsense tags	Not understandable by humans and computers	$$$$Tsk,16s
Sentence tags	Enclosed in double quotes and if quotes are removed Flickr, for example, interprets each word as separate tag	"The best scene view,must see"

The paper is structured as follows: in Section 2 the significance of the selected aspects has been justified. Analysis of the major approaches along with its pros, cons and challenges in the selected aspects are reviewed in Sections 3–6, respectively. Section 7 presents the summary and finally conclusion and future research directions are outlined in Section 8.

2 Significance of selected aspects

In this section, we will justify that how the selected aspects are important and play role in solving problem of keeping consistently well-maintained folksonomy in terms of quality. Let us have a look at each aspect.

2.1 Quality tag selection

Folksonomies do not limit its users to use a set of keywords already chosen by a group of experts for interpersonal information retrieval. However, selection of good tags not only enhances an individual's analysis and view of the tags, it also encourages them to develop better tags in the future (Farhan & Sanderson, 2008). Conformity theory observed in folksonomies indicates that the tags that users observe from other users will influence the tags that they will choose for annotation of a resource (Asch & Solomon, 1956). The proof of this is that once a tag becomes popular it remains popular (Sen *et al*., 2006). So good tag selection by the users will influence other users tag selection.

A choice of low-quality tags in the user-generated folksonomy have negative effect on the quality of the tag-based retrieved search results and user's experience (Li *et al*., 2012). Further pointed out by Geldart and Cummins (2010), low-quality tag creates problems, especially when looking at the subjects that have commercial significance. Once we have a collection of worthy tags we can utilize them in various ways, for example, displaying the good tags on tag cloud to enhance search and navigation. Tag suggestions when labelling a resource is improved if we are able to automatically find out quality tags.

Quality of indexing provided by the users is a key to measure the usefulness of the folksonomy (Sturtz, 2004). For example, Layne (1994) computed depth or thoroughness of folks indexing of tagged images. Moreover, depth indexing increases precision in retrieval (Magnuson, 2008).

2.2 Tag management facilitation provided by folksonomy applications

Information design of social networking applications influences the quality of the metadata collected. Floeck *et al*. (2011) research focussed on this perspective and they found from experimentation on the social bookmarking website of BobrDobr.ru that if users were allowed to view the most popular tags, the overall variation of tags assigned to websites by all the users decreases.

At the time of resource submission if provided, for example, a set of helpful checklist of questions that could be applied to the object being tagged, or suggest tags, will promote quality tag selection. Type-ahead functionality and editing functions for the tags like changing or modifying it (Guy & Tonkin, 2006), renaming tags (e.g. typos), deleting tags, merging tags (e.g. single and plural words) helps to avoid spell mistakes and ambiguity. Mandatory facets, meaning a user must categorize each item with at least a single tag from a facet. This keeps the folksonomy well organized and improves search and retrieval.

The precision of the retrieved search results can also be improved by the facilitation provided by the folksonomy application, when searching a resource, for example, query substitution improves performance. Use of ontology in the query expansion solves the problem, like, a user searching with a high-level concept such as plant retrieves resources, which are not normally tagged with such a high-level concept (Pan *et al*., 2009). To address the same problem, in Flickr the user may choose types of relations (e.g. synonyms, hypernyms, hyponyms, holonyms or meronyms) for expanding the query. Thus, if a user made a query for 'shoes' the query could be expanded with the hyponyms 'slippers' and 'trainers' to retrieve all the pictures tagged with these subtypes of shoes from the Flickr collection. Accurate search results are one of the proofs of well-maintained quality folksonomy.

2.3 Folksonomy cleaning

Tags assigned to the resources often lack precision and completeness (Liu *et al*., 2009), in addition to the fact that some tags do not remain applicable all the time or in other words are expired tags. Tag weeding (the term used by Peters and Weller (2008)) is required to remove spam, invalid and expired tags, which will keep the folksonomy clean and also would result in precise search results.

2.4 Interoperability across platforms

It is often the case that a user has profiles on more than one folksonomy applications (e.g. Del.icio.us and Flickr) at a time. Interoperability allows utilizing a same tag set uniformly in different platforms. Furthermore, tags from different sources are compared in a way, that there must be an explicit agreement on the understanding of the promising patterns of the tags usage (Veres, 2006). This will keep folksonomy consistent and non-ambiguous.

Concluding remarks: the above discussed four aspects are very important and interrelated for keeping quality-maintained folksonomy. The first reasons is that the knowledge of quality tag metrics help users on one hand to choose quality tag(s) and the designers of folksonomy-based systems on the other to design interfaces (e.g. facets, tag suggestions, query expansion) that ensure only good tags selection when entering and searching for a resource. Second, quality tag metrics enable the folksonomy cleansing mechanisms to distinguish between good and bad tags and to remove bad tags. Third, the tag management aspect of a folksonomy-based system enables, for example, the removal of expired tags by utilizing the time stamps associated with the tags. Lastly, the first three aspects ensure a quality folksonomy development and maintenance and the fourth aspect makes it possible to shift the same quality folks vocabulary across different platforms.

In Table 2 we have summarized the evaluation criteria to measure quality of folksonomies and mentioned as to which of our selected aspects cover which criteria. Criteria 1–7 are outlined by Farhan and Sanderson (2008) for e-government folksonomy and 8–10 are pointed out in Helic *et al*. (2011, 2012), Al-Khalifa and Davis (2006), Strohmaier *et al*. (2012a), Solskinnsbakk *et al*. (2012). Here we feel the need that research is required to have updated evaluation criteria and folksonomy management metrics (Folksonomy Management, 2014) against which the folksonomies must be evaluated to measure and ensure worthiness of the folksonomy with passage of time (Qualitatively and Quantitatively).

Table 2 Folksonomy quality measurement evaluation criteria

Criteria	Description	Aspects(s) that mostly cover the criteria
Orthography	Tags of the folksonomy and it compatibility with standard English	1,2
Motivation	Tags selection expresses the tagger motivation	1
Relevancy	Representation of aboutness of resources in the folksonomy	1,2,3
Subject coverage	Extent to which folksonomy covers the subject domain	1,2
Consensus	How much taggers agree on individual tags present in the folksonomy?	2
Consistency	Frequency of use of particular tag by different users and/or user agreement on pattern of use of a tag	2,4
Findability/retrieval effectiveness	Folksonomy ability to find out relevant resource(s)	2
Navigation effectiveness	How much navigation is effective and helpful in searching and retrieval?	1,2
Evaluating the semantic quality of the folksonomy	How much folksonomy is semantically enriched?	2,3
Semantic quality of learned taxonomy from folksonomy	How much folksonomy is well-organized semantically?	2,4

(1) quality tag selection, (2) tag management features, (3) folksonomy cleaning, (4) Interoperability.

In the next sections, we are going to analyze the approaches in each aspect, with pros, cons, analysis of techniques and challenges.

3 Quality tag selection

3.1 Tag literacy

One possible approach to have value tags can be that there should be rules and regulation for formation of the tags, as proposed in the article by Mejias (2005) regarding tag literacy. Suggestions include the following:

- Use of plurals rather than singulars: the reason behind this is, the searchers make use of plural keywords more than twice as often as singular keywords (Baxter, 2009).
- Use lower case: the reason being is that mostly people write queries in lower case.
- Group words with an underscore.
- Follow tag conventions started by others.
- Add synonyms: for clarity and related option(s).

One might make a case for the establishment of a consortium to release a list of general guidelines (Guy & Tonkin, 2006). An attempt in this direction by Spiteri (2007a) evaluated tags of three famous folksonomies including Del.icio.us, Technorati and Furl against the National Information Standard Organization (NISO) guidelines. They concluded that the tags corresponded closely to NISO guidelines but were not entirely compliant due to the problem of ambiguous tags(homographs and unqualified abbreviations). Some researchers have suggested after experiments to utilize the clear guidelines for construction of terms established for taxonomies, but all the guidelines are not significant. However, some guidelines for choice and formation of concept terms seem relevant and can be utilized (Spiteri, 2007b).

Along with educating users, it is also required to outline metrics that can be used to justify whether a tag is a quality tag or not.

3.2 Tag quality measure metrics

The approaches discussed in this section are classified according to the metrics defined for the goodness of a tag.

3.2.1 Folk rating

Folk rating metric covers all the techniques in which users score is involved. The techniques can be seen from three perspectives, which are individual rating based, expert vs. non-expert opinion and inter-subjectivity based.

Analysis of the approach: the problem with individual rating-based techniques is that asking folks to rate tags as a quality measure is time consuming (Damme *et al*., 2008). In addition, an another observation is that if a tag receives four consecutive thumbs down (indication of dislikeness) ratings (regardless of user), then 90% of the remaining ratings will be thumbs down as well (Sen *et al*., 2007).

In *expert vs. non-expert opinion*-based techniques it is assumed that a tag is a quality tag if it is assigned by the more expert users as compared with the less experienced ones. The other name as suggested by Nauerz *et al*. (2009) is 'reputation-based tagging'. Kang and Lerman (2011) investigated and concluded that accuracy and detailed folksonomy aspects are well covered by annotations from experts, however, annotations from non-expert or novice users lead to more comprehensive folksonomies than annotations of experts only.

Inter-subjectivity means tags that are understood and used by many members of a group. This is one of the indicators of quality and basis of *inter-subjectivity-based techniques*. Three metrics which are high-frequency tag, tag agreement and tag frequency inverse resource frequency (TF-IRF) have been proposed by Damme *et al*. (2008) and among these three, TF-IRF metric gave promising results. The metrics were tested on data set of Del.icio.us.

Same quality measure, that is, the number of users applying a particular tag is explored by Sen *et al*. (2007) using the *num-user* method. Summary of result suggested that tags being used once have lowest average rating, tags applied 16–31 times have the top average rating (2.53) and the tags applied most often (256 or more times) have an average rating of 2.14. Another inter-subjectivity quality measure explored is *num-searches*, which give weight and quality to the tag depending upon how many users did search using that tag. Next measure is based on aggregate user selection *global-avg* that gives position/weight to the tags with respect to their overall average rating across all users. Lastly, the *consec-apps* selection method, which ranks tags based on the number of initial identical ratings and the similar *consec-users*, which requires that the ratings come from different users.

The challenge to be considered is the systems may want users to rate tags rather than tag applications. Second, rating given by a user is influenced by the rating already given by other users (Sen *et al*., 2007).

3.2.2 Validity of tags

The term validity of tags means that social terms provide added value relative to terms from the controlled vocabularies.

Analysis of the approach: one of the comparative studies on the potential role of tags vs. controlled vocabulary by Syn and Spring (2009) concluded that controlled vocabulary tends to represent the content more precisely and stably. Kiu and Tsui (2008) supported taxonomy-directed folksonomies to achieve the objective of consistency. However, a finding from Trant (2006) based on his work on museum vocabulary has discovered that >70% of the vocabulary used for tagging by the users was not listed in documentation. The reason being is the viewpoint and navigational behaviour of expert varies from that of (general) users.

3.2.3 Information value

Information value of tag means how well a tag is useful for thematic description, retrieval and distinction between resources.

Analysis of the approach: it is quite commonly observed that tags used to describe one resource may not be appropriate for another resource. The quality of a tag varies from resource to resource, therefore, there is a need to measure tag quality with respect to resource to which it is assigned (Krestel & Chen, 2008).

The major techniques in literature to find appropriate tags are based on query log file and graph. In case of query log file-based techniques, appropriate tags are found by exploiting similarities between the tags used by the users in a query and those either in official vocabulary or used by the ordinary library patron in a query. This set of tags identifies associated library catalog records (Pera & Lund, 2009; Daglas *et al*., 2012). However, in graph-based techniques the relationship among user, tag and resource is utilized (Krestel & Chen, 2008).

3.2.4 Preference

Preference means the tags to which users give priority for annotation of a resource. There are many factors that influence the preference of the users. Nandipati (2011) suggested that in order to find out the preference, an important issue is to understand the motivation of users. Körner *et al*. (2010) outlined two groups of people involved in tagging as categorizers and describers. Categorizers tag their resources using vocabulary based on their mental model, personal preference and mostly use less number of tags. On the other hand, describers keep easy retrieval in mind while annotating the content. Mostly, they use synonyms. Measures proposed for isolating people of each group are tag/resource ratio, conditional entropy, orphaned tags, overlap factor and tag/title intersection ratio. Authors concluded that all the measures are not equally helpful but tag/resource ratio appears to give good results in capturing human judgement. Strohmaier *et al*. (2009) worked in the same direction, experimentally evaluated eight data sets and found out that tag agreement between users who are motivated by categorizing resources is less as compared with users who are motivated by describing resources.

Recent tags represent user's present preference or interests. User significance for tags changes with the passage of time. In an environment in which the user gradually changes interests, the tag data close to the current temporal period is usually more significant than that which is temporally far from the current period (Zhang *et al*., 2012). Here we feel that a study is required to highlight factors (one of them is user's motivation (Strohmaier *et al*., 2012b)) that influence user's tagging behaviour and then reflection of this on folksonomy quality.

An important aspect that needs consideration is that it is challenging to design interfaces that collect ratings in preference along with other quality dimensions (validity, information value) without puzzling users. Systems such as YouTube and Netix only gather ratings in preference dimension. However, it is necessary to explore interfaces for differentiating preference and other quality metrics, and examine the role they play in different domains (Sen *et al*., 2009). In addition, there are factors that are important in quality of tags, including user interaction level, organization level and interest in content (Nandipati, 2011).

3.3 Tag generation models

Gupta *et al*. (2010) outlined the key features of each tag model, focussing on the key aspect of how valued tags are selected in each model.

- *Basic Polya Urn model*: proposed by Golder and Huberman (2006), captures already assigned tags but does not consider new tags.
- *Yule Simon model*: probability of new word selection is p, and of copied word is 1 – p.
- *Yule Simon model with long-term memory*: copy using a distribution over past × time steps where probability decreases as a power law.
- *Information value-based model*-Focus on tags assigned previously, in addition to their information value.
- *More parameters*: background knowledge of a user, number of prior used tags the user has access to and most well-accepted/popular tags.
- *Language model*: model generation of tags and words collectively, using the model like Latent Dirichlet allocation (LDA).

3.4 Characteristics of 'Good' tags

In the light of above discussion and the features pointed out by Peters (2009), we can identify Good Tags from seven characteristics. First, frequently used or high-frequency tags because they depict users

collective intelligence. Second, more distinctively appropriate metadata for a resource (distinctiveness of a tag). Third, the user should spend minimal effort to go through the tags of a particular resource so as to minimize cognitive load on the user. This is accomplished by having small number of tags per resource. Fourth, the tags which are specific to user's perspective often do not have meaningful value to other users and hence should be removed from public folksonomy. Fifth, terms from controlled vocabulary such that they do not come strictly from any specific domain jargon. Sixth, spelling and semantic variant tags should be placed together to avoid confusion and get high recall. Seventh, recommended tags should be evenly selected from different categories (multiple facets) to ensure diversity of terms.

After having a brief discussion on quality tag selection metrics and good tags characteristics, in the next section our focus will be on the aspect of tag management features provided by the folksonomy applications.

4 Tag management features provided by the folksonomy applications

The term tag management refers to capability of managing user-generated tags or a folksonomy within a collaborative software so that quality folksonomy could be maintained. Microsoft Sharepoint is an example of applying tag management to a content management system. In addition, Jumper 2.0 (Perry, 2009) and Knowledge Plaza (Whatever, 2009) are examples of tag management applied to enterprise bookmarking. For blogging wordpress is popular. We will review the approaches of tag management from two perspectives, first, the number of features considered at the time of resource submission, and second, the features which facilitate users in searching of resource(s) and to obtain precise search results.

4.1 Tag features for resource entry

Rather than developing efficient algorithms for making sense out of messy user-generated tags, clever algorithms should be developed and utilized to help users in selection of tags that make sense in the first place. Keeping this in view, one can improve systems which help in tag creation. These can be called preventive measures.

4.1.1 Choice of tags

Utilize predefined tags: rather than allowing users to create their own tags, the users can utilize predefined tags presented as suggested tags. The approach is common either when submitting or searching for a resource.

Analysis of the approach: manually defining tags for each resource is potentially time consuming (Lipczak, 2008). Second, if the users are freely allowed to apply any tag to any resource, tagging systems will contain large numbers of unneeded, ambiguous and personal tags, which causes difficulty in resource searching (Gemmell & Shepitsen, 2008). It reduces cognitive load on the users, as the users need to choose one or more suggested tags instead of thinking and summarizing a source to find out the best representative tags. Tag suggestions support a common vocabulary surrounding all the users (Dattolo *et al*., 2010).

Major techniques in the literature for tag suggestion are based on considering tag–tag relationship (similar tags, co-occurring tags, alternative tags), tag–resource relationship (tags of similar resources), mapping (with ontology terms, folksonomy terms), user personomy and usage pattern.

Looking into the *quality and coverage of the recommended tags* is an important challenge. Because this will facilitate the users beyond the annotation process. For instance, techniques that suggest additional tags based on relations like 'is-a', 'part-of', 'paradigmatic abstraction' and 'tag inheritance' are particularly very significant in increasing a resources indexing depth (Peters, 2009). *Asymmetric* tag co-occurrence provides a more suitable variety of candidate tags as compared with symmetric co-occurrence. In comparison, Jaccard symmetric co-occurrence gives good results in finding alike tags (Sigurbjrnsson & van Zwol, 2008).

In addition, *quality and coverage of the suggested tags* are particularly significant for the resources that are not evaluated by computers such as photos, videos and music. Indexing terms provided automatically

by search engines work well for textual data but are less useful for describing and organizing resources like video, music and photos (Gemmell *et al*., 2009; Harvey *et al*., 2009).

Tag editing/tag preprocessing: tag editing features allow users to change or modify tags, rename tags (e.g. typos), delete tags and merge tags (e.g. single and plural words).

Analysis of the approach: to handle lexical mistakes, such as misspelling, a tag editing feature is required. Type-ahead functionality helps the users to avoid misspelling. The other promising techniques used include Natural Language Processing (NLP) techniques, thesauri (aspell (GNU Aspell, 2011), Princeton University's word lexical reference system (WordNet, 2010), Perl Lingua (Rossi, 2006; Franz & Richardson, 2007). These can be used independently or in combination for correction of misspelled tags, finding word class and stem.

The drawback of type-ahead is it fails to answer many of *semantic-level* and *person-level* overlap requirements (Muller, 2007). One possible solution by Muller (2007) suggests that type-ahead functionality should provide the users with tags from his or her own personomy in order to perk up *self-consistency*.

4.1.2 Tags linkage

To keep contents organized, one prominent way is to allow users to organize and categorize tags into facets (Spiteri, 2011). Keeping contents organized will enhance navigation and search precision. Tag relatedness can be viewed from two prospective, that is, Explicit and Implicit. Explicit relationship can be achieved by mapping tags with external knowledge sources such as ontologies (Laniado *et al*., 2007) and implicit relationship is achieved by finding indirect connection among tags via usage of metrics, for example, subsumption.

4.1.2.1 Explicit connection. *Community-driven ontology development*: user interactive interface for representing the ontological concept with collaborative tags is one of the research directions to keep the folksonomy organized. Collaborative ontology evolution system allows community members to add, modify or delete existing and new ontology classes according to their own needs. The editing of the classes does not require any special skills and is as simple as adding tags to current tagging systems. Dimensions of workflow include size of the ontology, size of community (contributors, editors), discussion tools (mailing, chat, message boards), synchronization and editing techniques (Subversion (SVN), Concurrent versions system (CVS)) (Noy, 2008).

Analysis of the approach: community involvement enables members to actively contribute towards the ontology evolution process and it works, as it is a key factor to accomplish a community common ground (Gendarmi *et al*., 2006; Ga *et al*., 2011). In other words, it characterizes the understanding of a particular community or domain area. The major challenges are, first, the choice of a suitable Ontology Meta-Model, appropriate for a large audience. Second, provision of helper and guidance feature to help the community for development and maintenance of collaborative ontology. Third, handling conflicts, for example, as if two communities have contradictory views on the same subject, or if two communities use analogous terms for different concepts (Hepp *et al*., 2006).

Community-based ontology matching: many ontologies to be matched are analogous to existing ontologies, mostly in the case when they are representing the same application domain. Matching is a precise solution to handle semantic heterogeneity and disambiguation problem (Zhdanova & Shvaiko, 2006). Figure 1 represents the ontology alignment between two simple ontologies. 'Senior Researcher' and 'Research Associate' ontology on the left is mapped with 'Researcher ontology' on right.

Analysis of the approach: involving users in mapping process helps to handle challenges, such as appropriateness of mappings, especially when using them in the new application. The current challenge in the collaborative ontology matching is to find a way to handle the incomplete and contradictory alignments. Another prominent challenge is to handle appropriate annotation support and satisfactory description units that enable and facilitate user participation at large scale (Shvaiko & Euzenat, 2008).

Folk-based categorization: folksonomies with predefined organization, such as hierarchical, put limits on the flexibility of tagging because of its restricted and predefined vocabulary. In addition, it also requires an extra effort from users to select tags. These limitations divert direction of research to allow users to enter tags along with their categories.

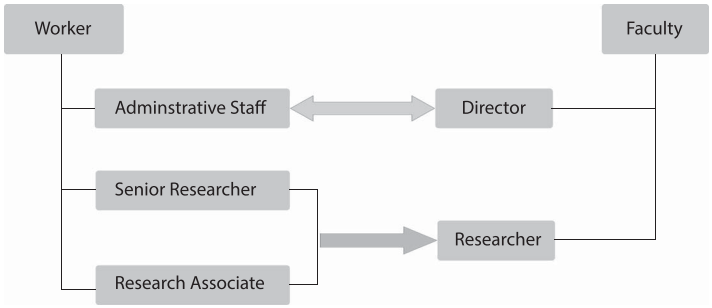

Figure 1 Ontology alignment (Zhdanova & Shvaiko, 2006)

Analysis of the approach: based on this approach users consensus-supported organized folksonomy is built. In addition to consensus it also gives flexibility to users in choosing tags and their corresponding categories. Yoo *et al*. (2013) claimed in their work that by adding this sort of organization retrieval is improved as compared with flat folksonomy. However, developed folksonomy using this approach by Yoo and Suh (2010) was limited to 'has' relationship.

4.1.2.2 Implicit connection. Explicit relationships are often inapplicable since they require that corresponding concepts must exist. For example, when a tag is mapped to an ontology it is required that corresponding concept must exist and also represents the meaning that was actually meant by the users. Apart from explicit relationships, there exists another linkage called implicit connection. Implicit relationships also play a key role in enhancing navigation, organization, portability and universal applicability (Matthes *et al*., 2012). Let us have a look at the approaches covering this aspect.

Grouping syntactical variation of tags: several reasons cause syntactical tag variations. That includes typographical mistakes, for example, semanticweb, semnticwev, zemantcweb, using singular or plural of the same word like semanticweb, semanticwebs, using separators, for example, semantic web, semanticweb or a combination of them like semntic-web, smanticweb, semntic-webs, etc.

Analysis of the approach: although all syntactically variant tags have identical sense, resources would be classified under diverse labels. This fact makes more confusion in locating the information, and navigating in the folksonomy. This problem can be rectified by identifying all of them as variations of the interchangeable label 'semantic web' and then group them under a single tag. The user could access this tag and would be able to obtain all the information relating to the resources associated with it and its syntactic variations (Echarte *et al*., 2009). Tag clouds, for example, could show a representative of each group of related variations and provide a doorway to all resources annotated with them. Major techniques in the literature are based on pattern matching (Echarte *et al*., 2009), similarity measures (Astrain *et al*., 2009) and linguistic sources like Wikipedia/DBPedia. However, accuracy is still a major challenge, that is, how correctly syntactic variant tags are detected, especially when dealing with short length tags.

Tag community: tag community detection is to find out groups of tags that are either semantically close to each other or used in the same context (Papadopoulos *et al*., 2011).

Analysis of the approach: Quattrone *et al.* (2011) proposed in their work that when a user is assigning tag(s) to a resource, similar and related tags already present in the folksonomy should be displayed. For example, in Del.icio.us and Flickr, when a user visits the page containing all the bookmarks/pictures tagged with a certain tag, a list of the tags related to that one are also shown inside a sidebar. The key benefit of displaying the tag community is to enhance help in searching via navigation.

There are many research efforts in the literature to compute similarity, like one based on linguistic resources (stemming/lemmatization/WordNet), or taking into account tag–tag relationships (clustering/co-occurrence), or tag–resource relations (tags associated to same resource are similar and vice versa).

As pointed out by Solskinnsbakk and Gulla (2001), the pitfall of linguistic sources is limited flexibility, since they are language dependent, therefore, difficulty in handling multilingual tags

(politics, politica) and sometimes proves to be expensive. However, it will help in organizing contents in hierarchy (Laniado *et al*., 2007).

The clustering/co-occurrence needs certain number of resources annotated with the same tag in order to recognize linguistic variations effectively. Second, co-occurrence may identify dual topics like Barcelona and Football and such tags cannot be moulded to be a single tag (Solskinnsbakk & Gulla, 2001). In comparison, tag–resource relation gives better results for tag disambiguation than co-occurrence measures (Wetzker & Bauckhage, 2010; Quattrone *et al*., 2011). Tag clustering techniques based on Hierarchical Agglomerative Clustering or *K*-mean results in good tag clusters if tags are distributed evenly but are not effective if distribution is not even (e.g. 'S' shape) (Xu *et al*., 2013).

Other approaches: *Equivalence* includes subsumption, identicalness and mutual exclusion (Matthes *et al*., 2012). *Abstract To Concrete Representation*, a hypothesis proposed in Benz and Christian (2011) is that specialists mostly make use of more specific vocabulary as compared with generalists, who rely mainly on abstract tags. *Relevancy Tag Ranking*, relevant tags need to be ranked keeping in view different ranking criteria such as, importance, semantic closeness (Liu & Zhang, 2011).

4.1.3 Other related aspects

Tagging styles and their evaluation: cost-wise analysis and evaluation of different tagging approaches like 'click2Tag' and 'SparTag.us' have been tested by Hong *et al*. (2008). Multimedia tagging by taking advantage of human intelligence and computers computation power is proposed by Wang *et al*. (2012). 'Automatic Tagging' or usage of alternatives such as a 'Tag Signature' in place of tags is proposed by Solskinnsbakk and Gulla (2011). Here we feel there is a need to evaluate different tagging styles keeping both quality and cost of tagging perspective in mind.

To end this subsection, Table 3 compares features supported by five popular folksonomy sites at the time of submission of a resource. The term tag gardening has been used for editing and organizing of tags by Peters and Weller (2008).

4.2 Precise resource searching via tags

Precision and accuracy in search results can also be achieved by incorporating supporting features for a user when he or she searches for a resource. In this section, we will look at the major approaches to achieve this objective. There are two directions, either to use visual interface for navigation-based search, or to write textual queries.

4.2.1 Visual interfaces

Tag cloud: to facilitate users in navigation and searching, most of the applications provide visual interfaces. Tag cloud is one of the popular types of interface being used in folksonomies.

Analysis of the approach: tag cloud design depends upon the type of the folksonomy application and its requirement. As an example, Diaz and Tory (2009) analyzed five diverse image and video search interfaces and gave a conclusion that search interfaces should maximize screen space used for visual representations of the media. In addition, it should present on-demand access to titles, tags and other metadata, and also make available contextual information about earlier viewed items, current keywords and alternate keyword possible options. The other important issues include selection of the tags and design (such as shape, colour and animation scheme and its interpretation) of the tag cloud.

Faceted browsing: with facets, resource collection can be retrieved and filtered in a structured way (Waltl, 2012).

Analysis of the approach: faceted browsing proved to be remarkable in organizing website resources. However, Choi and Street (2009) argued that the user point of view is not reflected in those facets and proposed to incorporate the users perspective in design of the interface.

Personalized view and search: personalized search is based on computing search results, considering user taste, interests and needs.

Table 3 Feature set support provided by five popular folksonomies when submitting a resource

Folksonomy sites	Editing features	Type of resource	Tag grouping	Adding resource	Tag recommendation	Facets classification	Tag community support	Restriction on choice of tags	Privacy level	Count number of taggers	Limit on number of tags given to resource
Del.icio. us	Add✓ Edit✓ Rename✓ Delete✓ Merge✓	Links, people	Allow tag bundles	Browser	✓	×	Recent tags	×	Public Private Friends Custom	✓	×
Flickr	Add✓ Edit✓ Rename✓ Delete✓ Merge✓	Photos, people	Relevant Recent Interesting	Browser e-mail Mobile Desktop	×	×	Recent Interesting Relevant	×	Public Private Friends Family	✓	Upto 75
Bibsonomy	Add✓ Edit✓ Rename✓ Delete✓ Merge✓	Links	Popular	Browser	✓	✓	Popular	×	Public Private Others	✓	×
Amazon	Tag features discontinued		Popular Frequently used tags	Browser	×	×	×	✓By guidelines	Public Private	×	×
Last.fm	Add✓ Edit✓ Rename✓ Delete✓ Merge✓	Songs, artist	Related tags	Browser	×	×	Related tags	×	Public Private	✓	×

Analysis of the approach: personalized view called 'Folkview' proposed by Dattolo and Pitassi (2011) works on creating personalized and adaptive paths for the users, so that association between the tags and resources can be modified. Major techniques to enhance precision and relevance in the search utilize user profiles and usage pattern.

By analyzing the profiles, users' interests can be outlined. Usage pattern-based techniques take advantage of following set of factors: the total number of tags a user has provided (as some user may have used few tags while other have many tags credited to them), the frequency of use of the tags in both ways, that is, the number of times a tag has been used by the users and how many users have used the tag (Golder & Huberman, 2006), the types of the tags like Personal, Factual and Subjective (Al-Khalifa & Davis, 2007), and the patterns of resources in which the users are more interested (prediction of reoccurring patterns) (Kawase & Herder, 2011). In the same way, Kwak *et al*. (2009) proposed that based on tag usage across different folksonomies the user with the same interest can be found. User interest pattern in temporal direction has been proposed by Zhang *et al*. (2012) meaning finding first and the last time usage of a tag. The more times the tag has not been used the higher is the probability of it not reoccurring and vice versa.

4.2.1.1 Interface design evaluation. Efficiency, effectiveness and user satisfaction criteria for evaluation are considered by Ravendran (2012) for evaluating the usability of the interfaces. McCarthy and Conners (2006) evaluated tag clouds on a descriptive basis. However, to design a visual interface that could be used for retrieval of contents continuously, we need to have a better understanding of the users goals, objectives and strategies (Diaz & Tory, 2009).

4.2.2 Query-based exploratory search

Query reformulation: query reformulation means disambiguation of abbreviation in queries, or to enrich it with meaning. This can be done by substitution and/or expansion. Substitution means dropping the original terms given in the query and adding a new term to replace them. Query expansion means to enrich queries with senses and/or with related classes.

Analysis of the approach: removing words as a component of substitution improves performance. However, an expansion is a much safer form of query amendment as compared with substitution (Wei *et al*., 2008; Dang & Croft, 2010). Expanding queries will reduce ambiguity and improve search precision. Major techniques in the literature for query expansion make use of relationships (e.g. co-occurrence), mapping (with external sources (ontologies/thesauri/anchor text) and senses (WordNet, Tagora Sense Repository) and retrieving top-ranked search results (Pseudo relevance feedback (PRF)).

The work done by Abbasi (2011) compared different query expansion methods used in folksonomies and evaluated these techniques on large scale. Query expansion based on co-occurring tags normalized using cosine, Dice and Jaccard coefficient are compared. The Jaccard coefficient does not favour tags that do not co-occur very often. The Dice coefficient gives more promising results in finding co-occurring tags as compared with Jaccard. Furthermore, Abbasi (2011) concluded that cosine-based expansion for queries seem suitable for sparse search results, PRF for queries having many appropriate resources (~11–50) and overlap-based global expansion for queries having a lot of relevant resources (approximately >50).

Finding senses by using WordNet is not suitable because it is either too coarse (does not cover senses of a word in all domains) or too fine (many of the tags matches to single sense and not to all the senses available for a particular word) (Chen *et al*., 2014).

Expanding abstract terms with more concrete terms is required to precise query and search results. Centrality and entropy measures are good at distinguishing abstract and concrete terms. Moreover, tag co-occurrence graph is more useful input to centrality measure than tag similarity graph to compute abstractness (Benz & Christian, 2011). In an image domain considering similarity of both visual and textual features gives 8% more improvement in detecting abstract tags as compared with if they are used individually (Xia *et al*., 2013).

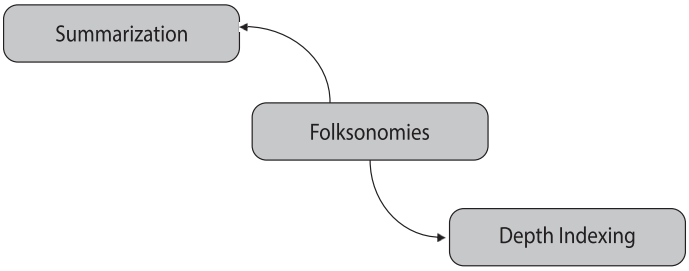

Figure 2 Folksonomies: a bridge between summarization and depth indexing

4.3 Other related aspects

Rank the search results/ranking the resources: given a keyword query (tag), the ranking strategy computes ranking of resources so that the resources that are most relevant to the keyword query become visible at the top of the ranking (Abel *et al*., 2010). As an example, Flickr offers search results considering relevancy, recentness and interestingness. Rásto *et al*. (2013) stated that faceted browsing navigational issues can be solved by classifying search results into hierarchical clusters. Hierarchical clusters make use of semantic properties of search results to get clusters and there is no need of domain experts, thus making them opt means for classifying search results.

Concluding remarks: folksonomies differ in results of precision and recall. The most likely reason behind these differences is their specific characteristics. Methods of collection building, tagging, socialization and ranking may differ from one folksonomy to the next (Morrison, 2008). Folksonomy-related difficulties outlined in Table 1 and discussed by Peters (2009) (such as, loss of indexing context, varying level of indexing and use of mixed languages) causes unsatisfactory search results, including inconsistencies in indexing depth.

Indexing depth affects both precision and recall. Indexing depth increases precision because more specific terms are used. Summarization increases recall because more broader terms are used (Taylor, 2004). According to Magnuson (2008) folksonomies may act as bridge between summarization and depth indexing by using folks analysis, paradigms and perspectives to index resources as shown in Figure 2. More suitably, the folksonomies make hierarchical power of information management more reasonable whose difficult classificatory taxonomies are beyond users vocabulary and understanding. However, in our opinion, quality of the folksonomy determines whether folksonomy is inclined more towards summarization or towards depth indexing. Three measures that are precision, recall and f measure are utilized by Vanderlei and Durão (2007) to test the search accuracy in folksonomies. They propose to combine keyword, facet-based and folksonomy search techniques to get the best retrieved results.

To test the navigation effectiveness of the folksonomy, Helic *et al*. (2011) worked on pragmatic (task oriented) evaluation. Broad and narrow folksonomies are evaluated for navigation criteria by Helic *et al*. (2012). Broad folksonomies result in better navigational utility since broad folksonomies give an opportunity of greater overlap among tags and also provide good alternatives for users to toggle between different portions of the network. Narrow folksonomies proved to be unable to provide such support. To end with, in Table 4 we have compared the searching features support provided by five popular folksonomies.

5 Folksonomy cleaning

Some tags expire with the passage of time (expired/invalid tags) while some are irrelevant to the resource to which they are assigned (false tag–resource relationship) or there is presence of some spam users. Folksonomy cleaning aspect deals with these cases to ensure folksonomy having resources tagged with correct and relevant tags, not with the noisy ones and also free from spam users. In this section, we will cover major research efforts to achieve the said objectives.

Table 4 Searching features supported by five popular folksonomies

Folksonomy sites	Query expansion	Visual interface or tag cloud	Personalized search	Search by tags	Allow friend or group search	Show same interest-based people
Del.icio.us	×	Display related tags/no tag cloud	✓	✓	✓	✓
Flickr	✓	Relevant, interesting, recent tags view in small, medium, detail and slide show	✓	✓	✓	✓
Bibsonomy	✓	Four views (alphabetic, list, frequency, tag cloud (limit on number of tags that appear in cloud))	✓	✓	✓	✓
Amazon	✓	No tag cloud and other views	✓	✓	✓	×
Last.fm	×	No tag cloud	✓	✓	✓	✓

5.1 Removing nonsense tags/expired tags

In portals that change dynamically, the tags do not remain appropriate all the time. This leads to the concept of tag expiry.

Analysis of the approach: tag expiry helps the community to focus on what is currently really relevant. Moreover, tag expiry allows us to neglect 'invalid' tags from being considered when doing content adaptation or recommendation. Major techniques include manual or filter-based removal of expired tags. Manual technique allows users to assign a 'lifetime' to the tags (Nauerz *et al*., 2009). In filter-based technique frequency threshold values have been used to filter out infrequent/invalid tags (Kiu & Tsui, 2008). However, identification of appropriate threshold values is important to judge the effectiveness of the filters.

5.2 Detection and refinement of noisy tags

The name *folksonomy-based item representation* is given to the practice of finding related tags for an item (resource) and utilizing these allied tags for illustration of each item. It is often the case that a tag is correct and is in frequent use but not relevant to the resource.

Analysis of the approach: noisy tags/spam posts or irrelevant tags assigned to a resource degrade performance because they give incorrect search results or low recall. Major techniques in the literature for finding relevancy (correct resource–tag relationship) are based on considering tag–tag relationship (similarity, co-occurrence). However, accuracy of tag similarity calculated from the corpus that is to be refined seems to be effected by noisy tags (Wang *et al*., 2009). Therefore, semantic and context-based approaches are more encouraged to use. The major challenges are, first, the decision of removal must be considered from view point of language dependency or society dependency. Second, refinement should not eliminate important tags.

5.3 Detection and removal of spam users

Spam users are major source of noisy tags. Spammers usually use different techniques for polluting the folksonomy. Examples include, by creating many accounts, publishing a specific post many times, using semantically diverse tags to annotate a resource and teaming up with other spammers to give good rating to each other (Yazdani *et al*., 2012).

Analysis of the approach: pollution in the folksonomy can be reduced if there is restricted access (Preventing Spammers) (Maurer *et al*., 2008). Detection-based techniques do not put restrictions on users access, however, for detection, a model of feature set based on information about the users, activities/posts and interaction is developed to distinguish between legitimate and spam user. Furthermore, features

Table 5 Feature set support, major approaches and corresponding techniques in tag management aspect

Feature set support	1 Correct spell/valid tags entry 2 Precise search results 3 Navigation effectiveness 4 Ambiguity resolution (disambiguation) 5 Multilingual support 6 Organized folksonomy contents	
Major approaches	Major techniques	Paper ID
Predefined tags/suggested tags	Rank tags	Jeong *et al*. (2011)
	Relationship [-Similar tags (cosine, Jaccard coefficient, Pearson correlation, Co-occurring tags, Tag clusters)	Sigurbjrnsson and van Zwol (2008) Zanardi and Capra (2008)
	-Alternative Tags (synonyms, expansion of acronyms)	Xu *et al*. (2013)
	-Similarity based on mutual reinforcement principal	Quattrone *et al*. (2011), Warden (2007)
	-Automatic Tags (tags of similar resources)	Sood *et al*. (2007)
	-Tags of content-based similar resources	Bartolini *et al*. (2013), Muller (2007)
	-User–resource and tag–resource relationship	Gemmell *et al*. (2009)
	-is-a, has-a, part-of	Peters (2009)
	-Tag inheritance]	
	User personomy	Lipczak (2008), Gemmell and Shepitsen (2008)
	Folksonomy terms	Wetzker and Bauckhage (2010)
	Usage pattern	Zhang *et al*. (2012)
	Mapping with ontology	Limpens *et al*. (2008)
	Type-ahead	
Tag editing	NLP techniques	Stock (2007)
	Thesauri	
Community-driven ontology	Collaborative Protégé	Tudorache *et al*. (2008)
Community-driven ontology matching	OntoWiki	Hepp *et al*. (2006)
	BIOPortal[1]	
	Pattern matching	Echarte *et al*. (2009)
Grouping syntactic variations	Similarity measures	Astrain *et al*. (2009)
	Linguistic sources (Wikipedia, online libraries)	
	Centrality and entropy measures	Benz and Christian (2011)
Abstract to concrete representation	Visual + Textual features (image domain)	Xia *et al*. (2013), Laniado *et al*. (2007)
	Linguistic source	Quattrone *et al*. (2011)
	Tag–tag relationships	Lipczak (2008)
	Tag signature	Solskinnsbakk and Gulla (2001)
Tag community	Resource–tag relationship	Quattrone *et al*. (2011)
	User model	Chow *et al*. (2009)
	Utag	Donghee and Yongmoo (2010)
Relevancy tag ranking	Tag Bundle[2]	Liu and Zhang (2011)
	Content based	
Equivalence	Subsumption, Equivalence, Mutual Exclusion	Matthes *et al*. (2012)
	Tag clouds	Sordo *et al*. (2013)
Browsing	Faceted browsing	Waltl (2012)
	Profile based	Vallet *et al*. (2010), Szomszor *et al*. (2008), Kwak *et al*. (2009), Kawase and Herder (2011)
Personalized view and search	Usage pattern	
	Mapping with	Pan *et al*. (2009)
	Thesauri/Ontology/Anchor Text (External sources)	Dang and Croft (2010)
Query reformulation	Relationships (co-occurring tags)	Jin *et al*. (2009)
	Mapping with senses	Dellschaft and Szomszor (2009), Hotho *et al*. (2006)
	Graph-Based FolkRank	Abel *et al*. (2010)
Ranking of the search results	Semantics and context-GRank	Abel *et al*. (2008)
	Clustering (hierarchical, c-clustering)	Rodenhausen *et al*. (2012), Rásto *et al*. (2013)

[1] http://bioportal.bioontology.org/
[2] http://www.Del.icio.us.com

(on basis of which model is developed) can also be defined considering different levels, including user level (spam user), post level (spam post/noisy tag) and resource level (spam resource). Most of the techniques in the literature for detection of the spam users are tested by defining features at post level and user level.

Cui *et al*. (2011) pointed out that most of the research done in this area usually utilizes partial information about users, resources and tags. However, research is needed in a specific direction to make use of all the features related to the users, resources and tags in the folksonomies. This would result in more precise expert findings and is more resilient to spammers. Keeping this in mind, Poorgholami *et al*. (2013) outlined set of features by considering the tags, resources, users and relationship among them. The authors claimed that these features are used by many machine learning algorithms to detect spammers with 99% of accuracy.

6 Interoperability of tags

In this section, we will analyze the issue of interoperability of the tags from two points of view that are to use the same tag set and an agreement on meaning/usage pattern of the tags across different platforms.

6.1 Same tag set utilization across platforms

If we look from the user's prospective then it is a more effective and friendly approach to use same set of tags on different platforms. TagCare (Golov *et al*., 2008) allows users to bring their personal tags from Del.icio.us, Flickr and Bibsonomy platform into one place. It helps the users to apply the same tags uniformly in different platforms. In the same direction, Bibsonomy, a popular social bookmarking service allows to import Del.icio.us., Firefox bookmarks and local BibTex files (Herwig, 2008).

Analysis of the approach: in Tagcare the statistical features provide a summary that gives an idea regarding the usage of a certain tag. This can be represented as tag cloud or as a ranking of the most often or least frequently used tags. This can help to find out tags which have been used frequently or very infrequently which may necessitate that they be grouped with others.

6.2 Semantic interoperability using metadata vocabularies

Metadata vocabularies like Dublin Core (DC) and folksonomy ontologies play a fundamental role for metadata interoperability within any particular domain and across domains.

Analysis of the approach: one of the studies conducted in an attempt to assign properties of DC to the folksonomy tags reveals that percentage of tags unassigned to DC elements is higher, approximately ranging between 35 and 45%. The reason behind this fact is that their meaning cannot be identified. Some new properties were identified which include action, rate, depth and utility. These fresh properties can be assigned to most of the tags (~26.5%). However, still 13% of the tags remained unassigned (Catarino & Baptista, 2010).

Saab (2011) user-generated tags in one application are exploited by another within a similar domain. Tags contained in RSS feed get accumulated during the interaction of the users with iCITY that provides services of cultural recommendations. These tags are mapped with the art-related concepts present in the CHIP, which is a personalized museum application. For mapping the relationships, skos:exactMatch, skos:broadMatch, skos:narrowMatch and skos:relatedMatch of Simple Knowledge Organization System (SKOS) have been utilized. However, this mapping is lexical mapping and still faces the problems of messy and ambiguous tags.

Knerr (2001) developed ontology of the folksonomy using semantic web techniques. For validation purpose, tagging data from Del.icio.us is obtained and incorporated in the developed ontology. Automatic generation of the ontology for the folksonomy by software agents makes this process easy and efficient.

In conclusion, researchers have made various efforts with partial and inadequate success for combining ontologies and folksonomies to achieve the objective of semantic interoperability. However, the

challenges faced in this include (1) structured nature of the ontologies and unstructured nature of the folksonomies, (2) lexicality, syntax and semantics, (3) social and cultural dimensions of the tags, and (4) appearance and reification (Saab, 2011).

7 Summary

Brief summary of the major approaches and corresponding techniques in *tag management facilitation provided by the folksonomy applications, quality tag selection, folksonomy cleaning and tags interoperability* aspects are presented in Tables 5–8, respectively. We have also itemized the prominent feature set coverage of approaches falling in each of the aspect. For example, if we look at Table 6 the approaches cover features of good and quality tags selection that improves precision in search and navigation and also provides coverage to non-conventional tags. In the same way, feature set support provided by the approaches in rest of the aspects are outlined.

8 Conclusion and future directions

To the best of our knowledge based on an extensive survey, we have outlined the major research efforts done for quality folksonomy maintenance and have systematized it in four aspects. The important questions covered in the analysis are as follows.

Aspects: the core aspects that will cover features required for keeping quality of the folksonomy protected.

Significance of the aspects: the way in which, defined aspects will cover the features and ensure a well-maintained quality folksonomy.

Approaches in each defined aspect: major approaches to cover the defined features in each aspect along with its pros, cons and challenges.

Efforts have been made in setting metrics for good and worthy tags, techniques for entering good tags in folksonomy-based systems, organizing folksonomy, disambiguation, removal of low-quality, irrelevant and spam tags along with spam users. These all efforts have significant effect on folksonomy quality,

Table 6 Feature set support, major approaches and corresponding techniques in tag selection aspect

Feature set support	1 Quality tag selection 2 Enhance navigation and search precision 3 Non-conventional tag coverage	
Major approaches	Major techniques	Paper ID
Tag literacy	Established guidelines	Guy and Tonkin (2006), Spiteri (2007b)
Folk rating	Individual	Lee and Han (2007), Sen *et al*. (2007)
	Expert vs. non-expert	Hsueh *et al*. (2009), Nauerz *et al*. (2009)
	Inter-subjectivity	Damme *et al*. (2008), Sen *et al*. (2007)
Validity of tags (taxonomy-directed folksonomy)	Taxofolk	Kiu and Tsui (2008)
Information value	Query log based	Pera and Lund (2009), Daglas *et al*. (2012)
	Graph based, TRP-Rank	Krestel and Chen (2008)
Preference	Temporal	Zhang *et al*. (2012), Strohmaier *et al*. (2009)
	Tagger motivation	Körner *et al*. (2010)
Tag generation models	Basic Polya Urn model	Gupta *et al*. (2010)
	Yule Simon model	
	Yule Simon model with long-term memory	
	Information value-based model	
	More parameters	
	Language model	

Table 7 Feature set support, major approaches and corresponding techniques in folksonomy cleaning aspect

Feature set support	1 Correct tag and resource relation 2 Removal of spam tags, invalid tags, expired tags and spam users	
Major approaches	Major techniques	Paper ID
Tag expiry	Manual	Nauerz *et al*. (2009)
	Filters	Kiu and Tsui (2008)
Refinement/removal of noisy tags	Semantic and context based	Liu *et al*. (2009), Lee *et al*. (2010)
	Quantitative/Qualitative	Tang *et al*. (2010) Yang and Lee (2011), Li *et al*. (2009) Lee *et al*. (2012), Wang *et al*. (2009) Noh *et al*. (2010), Damme *et al*. (2008)
Detection/removal of spam users	Prevention	Maurer *et al*. (2008), Markines *et al*. (2009)
	Detection (model of features)	Noll and Gibbins (2009)
	Demotion	Yazdani *et al*. (2012) Poorgholami *et al*. (2013)

Table 8 Feature set support, major approaches and corresponding techniques in tags interoperability aspect

Features supported	1 Folksonomy consistency 2 Semantic disambiguation 3 Interoperability	
Major approaches	Major techniqes	Paper ID
Using same tag set	Tag care	Golov *et al*. (2008)
	Common tag	Common Tag (2014
	JsongMiner	jSongMiner (2014)
Semantic interoperability	Dublin Core	Catarino and Baptista (2010)
	Folksonomy ontology	Knerr (2001), Cena *et al*. (2008)
	Upper tag ontology	Ding *et al*. (2010)

which is reflected by achieving precision in search. However, we still need to focus on following aspects which need researchers attention. These are as follows.

Updated evaluation criteria: it is necessary to have updated evaluation criteria (benchmarks) to test and ensure a well-maintained folksonomy in terms of quality. In addition, evaluation criteria will also be used for classification of algorithms/techniques/methods and estimation of their effectiveness on set-benchmarks.

Folksonomy management metrics: defining metrics for folksonomy management like quantitative metrics, qualitative metrics, time metrics, utilization metrics and financial metrics.

Factors influencing tagging behaviour: study is required to find out what factors (like social, environmental) influence tagging behaviour of the users and how much passage of time plays its role in changing their tagging behaviour. For this objective development of quantitative and qualitative measures are also required to test the influence of each factor. This study will also give answer to the question that how the change of tagging behaviour influences quality of the folksonomy.

Approach for tagging: evaluation of effectiveness is required both for cost and quality perspective with regard to every tagging style. This may lead to an option of having folksonomy type-dependent tagging style.

Acknowledgement

This research is undertaken by the first author for partial fulfilment of PhD degree requirements with support of Higher Education Commission of Pakistan (HEC).

References

Abbasi, R. 2011. Query expansion in folksonomies. In *Proceedings of the 5th International Conference on Semantic and Digital Media Technologies (SAMT10)*, 1–16.

Abel, F., Henze, N. & Krause, D. 2008. Ranking in folksonomy systems: can context help? In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, 1429–1430. ACM.

Abel, F., Kawase, R. & Krause, D. 2010. Leveraging multi-faceted tagging to improve search in folksonomy systems. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, 299–300.

Al-Khalifa, H. & Davis, H. 2006. Measuring the semantic value of folksonomies. *Innovations in Information Technology, 2006*, 1–5. IEEE. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4085397.

Al-Khalifa, H. & Davis, H. 2007. Towards better understanding of folksonomic patterns. In *Proceedings of the 18th Conference on Hypertext and Hypermedia*, 163–166. ACM Press. http://portal.acm.org/citation.cfm?doid=1286240.1286288.

Asch, S. & Solomon, E. 1956. Studies of independence and conformity: a minority of one against a unanimous majority. *Psychological Monographs: General and Applied* **70**(9), 1–70.

Astrain, J. A., Echarte, F., Crdoba, A. & Villadangos, J. 2009. A tag clustering method to deal with syntactic variations on collaborative social networks. In *ICWE 2009*, LNCS **5648**, 434–441. Springer-Verlag Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-02818-2_35.

Bartolini, I., Patella, M. & Romani, C. 2013. SHIATSU: tagging and retrieving videos without worries. *Multimedia Tools and Applications* **63**(2), 357–385.

Baxter, M. 2009. Reflections from the river—singular plural. http://www.reflectionsfromtheriver.com/tag/singular-plural/.

Benz, D. & Christian, K. 2011. One tag to bind them all: measuring term abstractness in social metadata. In *ESWC 2011, Part II*, LNCS **6644**, 360–374. Springer-Verlag Berlin Heidelberg.

Catarino, M. & Baptista, A. 2010. Relating folksonomies with Dublin Core. *International Journal of Metadata, Semantics and Ontologies* **5**(4), 285–295.

Cattuto, C., Baldassarri, A., Servedio, V. D. P. & Loreto, V. 2007. Vocabulary growth in collaborative tagging systems. CoRR abs/0704.3316. http://arxiv.org/abs/0704.3316.

Cena, F., Carmagnola, F. & Cortassa, O. 2008. Tag interoperability in cultural web-based applications. In *Proceedings of the 19th ACM Conference on Hypertext and Hypermedia (HT08)*, 221–222. http://dl.acm.org/citation.cfm?id=1379135.

Chen, J., Feng, S. & Liu, J. 2014. Topic sense induction from social tags based on non-negative matrix factorization. *Information Sciences* **280**, 16–25.

Choi, Y. & Street, E. D. 2009. Bringing a more accurate user's perspective into web navigation: facet analysis of folksonomy tags. *iConference 2009*, University of North Carolina at Chapel Hill.

Chow, K. O., Fan, K. Y. K., Chan, A. Y. K. & Wong, G. T. L. 2009. Content-based tag generation for the grouping of tags. In *International Conference on Mobile, Hybrid, and On-line Learning*, 7–12. http://www.computer.org/csdl/proceedings/elml/2009/3528/00/3528a007-abs.html.

Common tag 2014. Common tag. http://www.commontag.org/home.

Cui, B., Han, Q., Yao, J., Zhang, C. & Zhou, Y. 2011. Modeling user expertise in folksonomies by fusing multi-type features. In *Proceedings of the 16th International Conference on Database Systems for Advanced Applications*, 53–67.

Daglas, S., Kakali, C. & Kakavoulis, D. 2012. A methodology for folksonomy evaluation. In *Theory and Practice of Digital Libraries (TPDL 2012)*, LNCS **7489**, 247–259. Springer-Verlag Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-33290-6_27.

Damme, C. V., Christiaens, S. & Trog, D. 2008. Methodological approach to determine appropriately annotated resources in narrow folksonomies. In *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, Meersman, R., Tari, Z. & Herrero, P. (eds). Springer-Verlag Berlin Heidelberg, 230–237.

Damme, V. D., Hepp, M. & Coenen, T. 2008. Quality metrics for tags of broad folksonomies. In *Proceedings of International Conference on Semantic Systems (I-semantics 08), Journal of Universal Computer Science (J.UCS)*, 118–125.

Dang, V. & Croft, W. B. 2010. Query reformulation using anchor text. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, 41–50.

Dattolo, A., Ferrara, F. & Tasso, C. 2010. The role of tags for recommendation: a survey. In *3rd International Conference on Human System Interaction*, 548–555. IEEE. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5514515.

Dattolo, A. & Pitassi, E. 2011. Visualizing and managing folksonomies. In *Proceedings of the Workshop on Semantic Adaptive Social Web (SASWeb 2011)*, *CEUR Workshop Proceedings* **730**, 6–14. http://www.academia.edu/2971660/Visualizing_and_Managing_Folksonomies.

Dellschaft, K. & Szomszor, M. 2009. Sense aware searching and exploration with MyTag. In *The 8th International Semantic Web Conference (ISWC 2009)*, 6–7.

Diaz, J. & Tory, M. 2009. An exploratory study of tag-based visual interfaces for searching folksonomies. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, 410–417.

Ding, Y., Jacob, E. & Fried, M. 2010. Upper tag ontology for integrating social tagging data. *Journal of the American Society for Information Science and Technology* **61**(3), 505–521.

Donghee, Y. & Yongmoo, S. 2010. User-categorized tags to build a structured folksonomy. In *Second International Conference on Communication Software and Networks (ICCSN'10)*, 160–164.

Dotsika, F. 2009. Uniting formal and informal descriptive power: reconciling ontologies with folksonomies. *International Journal of Information Management* **29**(5), 407–415.

Echarte, F., Astrain, J., Crdoba, J. & Villadangos, J. 2009. Improving folksonomies quality by syntactic tag variations grouping. In *Proceedings of the 2009 ACM Symposium on Applied Computing*, 1226–1230. http://dl.acm.org/citation.cfm?id=1529559&dl=ACM&coll=DL&CFID=179648186&CFTOKEN=60450037.

Farhan, H. R. & Sanderson, M. 2008. Measuring the quality of an E-Government folksonomy. http://www.seg.rmit.edu.au/mark/publications/my_papers/E-government_Folksonomy.pdf.

Floeck, F., Putzke, J., Steinfels, S., Fischbach, K. & Schoder, D. 2011. Imitation and quality of tags in social bookmarking systems-collective intelligence leading to folksonomies. In *On Collective Intelligence: Advances in Intelligent and Soft Computing*, Bastiaens, T. J., Baumöl, U. & Krämer, B. J. (eds). Springer Berlin Heidelberg, **76**, 75–91. http://link.springer.com/chapter/10.1007/978-3-642-14481-3_7.

Folksonomy Management 2014. Folksonomy Management. http://www.if4it.com/SYNTHESIZED/DISCIPLINES/Folksonomy_Management_Home_Page.html#INTRODUCTION.

Franz, B. & Richardson, J. 2007. http://search.cpan.org/snowhare/Lingua-Stem-0.82/lib/Lingua/Stem.pod.

Ga, D., Zouaq, A., Torniai, C., Jovanovi, J. & Hatala, M. 2011. An approach to folksonomy-based ontology maintenance for learning environments. *IEEE Transactions on Learning Technologies* **4**(4), 301–314.

Geldart, J. & Cummins, S. 2010. The automatic integration of folksonomies with taxonomies using non-axiomatic logic. In *Proceedings of Information System Development*, 365–372. Springer Science + Business Media.

Gemmell, J., Schimoler, T., Ramezani, M. & Mobasher, B. 2009. Adapting K-nearest neighbor for tag recommendation in folksonomies. In *Proceedings of the 7th Worshop on Intelligent Techniques for Web Personalization and 21th International Joint Conference on Recommender Systems in Conjunction*, Artificial Intelligence, Pasadena, California, USA, 51–62.

Gemmell, J. & Shepitsen, A. 2008. Personalization in folksonomies based on tag clustering. In *Proceedings of the 6th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems*, 37–48. http://www.aaai.org/Papers/Workshops/2008/WS-08-06/WS08-06-005.pdf.

Gendarmi, D., Lanubile, F., Informatica, D. & Orabona, V. E. 2006. Community-driven ontology evolution based on folksonomies. In *OTM Workshops 2006*, LNCS **4277**, 181–188. Springer-Verlag Berlin Heidelberg.

GNU Aspell 2011. GNU Aspell. http://aspell.sourceforge.net/.

Golder, S. & Huberman, B. 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* **32**(2), 198–208.

Golov, E., Weller, K. & Peters, I. 2008. TagCare: a personal portable tag repository. In *International Semantic Web Conference*, 2–3. http://www.phil.hhu.de/fileadmin/Redaktion/Institute/Informationswissenschaft/1227004317iswc2008_t.pdf.

Gupta, M., Li, R., Yin, Z. & Han, J. 2010. Survey on social tagging techniques. *SIGKDD Explorations* **12**(1), 58–72.

Guy, M. & Tonkin, E. 2006. Folksonomies tidying up tags. *D-Lib Magazine*, **12**(1), Janauary 2006.

Harvey, M., Baillie, M., Ruthven, I. & Elsweiler, D. 2009. Folksonomic tag clouds as an aid to content indexing. CoRR abs/0911.4178. http://arxiv.org/ftp/arxiv/papers/0911/0911.4178.pdf.

Helic, D., Krner, C. & Granitzer, M. 2012. Navigational efficiency of broad vs. narrow folksonomies. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, 63–72. ACM. http://dl.acm.org/citation.cfm?id=2310008.

Helic, D., Lerman, K., Rey, M., Strohmaier, M. & Trattner, C. 2011. Pragmatic evaluation of folksonomies. In *Proceedings of the 20th International Conference on World Wide Web*, 417–426. ACM.

Hepp, M., Bachlechner, D. & Siorpaes, K. 2006. OntoWiki: community-driven ontology engineering and ontology usage based on Wikis. In *Proceedings of the 2006 International Symposium on Wikis*, 143–144. http://dl.acm.org/citation.cfm?id=1149487.

Herwig, J. 2008. Folksonomy the semantic puzzle. http://blog.semantic-web.at/tag/folksonomy/.

Hong, L., Chi, E. & Budiu, R. 2008. SparTag.us: a low cost tagging system for foraging of web content. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI08)*, 65–72. ACM. http://dl.acm.org/citation.cfm?id=1385582.

Hotho, A., Jschke, R., Schmitz, C. & Stumme, G. 2006. FolkRank: a ranking algorithm for folksonomies. In *FGIR*, 2–5. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.5271&rep=rep1&type=pdf .

Hsueh, P., Melville, P. & Sindhwani, V. 2009. Data quality from crowdsourcing: a study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, 27–35. http://dl.acm.org/citation.cfm?id=1564137.

Jeong, J. W., Hong, H. K. & Lee, D. H. 2011. i-TagRanker: an efficient tag ranking system for image sharing and retrieval using the semantic relationships between tags. *Multimedia Tools and Applications* **62**(2), 451–478.

Jin, S., Lin, H. & Su, S. 2009. Query expansion based on folksonomy tag co-occurrence analysis. In *Proceedings of IEEE International Conference on Granular Computing (GRC 09)*, 300–305.

jSongMiner 2014. jSongMiner. http://jmir.sourceforge.net/jSongMiner.html.

Kang, J. & Lerman, K. 2011. Leveraging user diversity to harvest knowledge on the social web. *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE 3rd International Conference on Social Computing (SocialCom)*, 215–222. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6113117.

Kawase, R. & Herder, E. 2011. Classification of user interest patterns using a virtual folksonomy. In *Proceeding of the 11th Annual International ACM/IEEE Joint Conference on Digital libraries (JCDL 11)*, 105–108. http://portal.acm.org/citation.cfm?doid=1998076.1998095.

Kiu, C. & Tsui, E. 2008. TaxoFolk: a hybrid taxonomy—folksonomy classification for enhanced knowledge navigation. *Knowledge Management Research & Practice* **8**(1), 24–32.

Knerr, T. 2001. Tagging ontology towards a common ontology for folksonomies, 3–8. http://tagont.googlecode.com/files/TagOntPaper.pdf.

Körner, C., Kern, R. & Strohmaier, M. 2010. Of categorizers and describers: an evaluation of quantitative measures for tagging motivation. In *21st ACM SIGWEB Conference on Hypertext and Hypermedia (HT2010)*, 157–166. ACM.

Krestel, R. & Chen, L. 2008. The art of tagging: measuring the quality of tags. In *The Semantic Web, 3rd Asian Semantic Web Conference (ASWC 2008)*, LNCS **5367**, 257–271. Springer-Verlag Berlin Heidelberg.

Kwak, H., Shin, H., Yoon, J. & Moon, S. 2009. Connecting users with similar interests across multiple web services. In *ICWSM*, 1–4. http://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewPDFInterstitial/189/506.

Laniado, D., Eynard, D. & Colombetti, M. 2007. Using WordNet to turn a folksonomy into a hierarchy of concepts. In *Semantic Web Application and Perspectives—Fourth Italian Semantic Web Workshop*, 192–201. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.4393&rep=rep1&type=pdf#page=200.

Layne, S. S. 1994. Some issues in the indexing of images. *Journal of the American Society for Information Science* **45**(8), 583–588.

Lee, S., Neve, W. D. & Ro, Y. M. 2010. Tag refinement in an image folksonomy using visual similarity and tag co-occurrence statistics. *Signal Processing: Image Communication* **25**(10), 761–773.

Lee, S., Neve, W. D. & Ro, Y. M. 2012. Towards data-driven estimation of image tag relevance using visually similar and dissimilar folksonomy images. In *Proceedings of the 2012 Workshop on Socially-Aware Multimedia (SAM 2012); held in Conjunction with ACM Multimedia 2012*, 3–8. http://www.citeulike.org/user/wmdeneve/article/11007815.

Lee, S. E. & Han, S. S. 2007. Qtag: introducing the qualitative tagging system. In *Proceedings of the 18th Conference on Hypertext and Hypermedia (HT 07)*, 35–36. http://dl.acm.org/citation.cfm?id=1286250&dl=ACM&coll=DL&CFID=179648186&CFTOKEN=60450037.

Li, J., Ma, Q., Asano, Y. & Yoshikawa, M. 2012. Improving folksonomy tag quality. In *WAIM 2012 Workshops*, LNCS **7419**, 264–275. Springer-Verlag Berlin Heidelberg.

Li, X., Snoek, C. G. M. & Worring, M. 2009. Learning social tag relevance by neighbor voting. *IEEE Transactions on Multimedia* **11**(7), 1310–1322.

Limpens, F., Gandon, F. & Buffa, M. 2008. Bridging ontologies and folksonomies to leverage knowledge sharing on the social web: a brief survey. In *23rd IEEE/ACM International Conference on Automated Software Engineering-Workshops (ASE)*, 13–18. IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4686305.

Lipczak, M. 2008. Tag recommendation for folksonomies oriented towards individual users. *ECML PKDD Discovery Challenge* **84**. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.183.2330&rep=rep1&type=pdf#page=92.

Liu, D., Zhang, M. W., Yang, Y. & Hong Jiang, X. S. H. 2009. Tag quality improvement for social images. In *International Conference on Multimedia and Expo (ICME)*, 350–353. IEEE.

Liu, D. & Zhang, X. H. H. 2011. Content-based tag processing for Internet social images. *Multimedia Tools and Applications* **51**(2), 723–738.

Magnuson, L. 2008. Folksonomies: meaning, discourse, and information retrieval, 1–15. http://www.cais-acsi.ca/proceedings/2009/Magnuson_2009.pdf.

Markines, B., Cattuto, C. & Menczer, F. 2009. Social spam detection. In *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb09)*, 41. ACM Press. http://portal.acm.org/citation.cfm?doid=1531914.1531924.

Matthes, F., Neubert, C. & Steinhoff, A. 2012. Structuring folksonomies with implicit tag relations. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, 315–316. http://dl.acm.org/citation.cfm?id=2310051&dl=ACM&coll=DL&CFID=179648186&CFTOKEN=60450037.

Maurer, B. Mcmillen, C., Abraham, D. & Blum, M. 2008. reCAPTCHA: human-based character recognition via web security measures. *Science* **321**(5895), 1465–1468.

McCarthy, L. & Conners, D. 2006. Folksonomies in action. http://folksonomiesinaction.blogspot.com/.

Mejias, U. 2005. Tag literacy. http://blog.ulisesmejias.com/2005/04/26/tag-literacy/.

Morrison, P. J. 2008. Tagging and searching: search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing & Management* **44**(4), 1562–1579.

Muller, M. J. 2007. Comparing tagging vocabularies among four enterprise tag-based services. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, 341–350. ACM Press. http://portal.acm.org/citation.cfm?doid=1316624.1316676.

Nandipati, A. 2011. Assessment of metadata associated with geotag pictures. Masters thesis, University of Muenster.

Nauerz, A., Bakalov, F., Welsch, M. & Birgitta, K. 2009. New tagging paradigms for content recommendation in Web 2. 0 portals. In *Workshop on Adaptation and Personalization for Web 2.0*, 143–147.

Noh, T., Lee, J., Park, S., Lee, S. & Kim, K. 2010. Tag quality feedback: a framework for quantitative and qualitative feedback on tags of social web. In *PRICAI 2010*, LNAI **6230**, 637–642. Springer-Verlag Berlin Heidelberg.

Noll, MG & Gibbins, N. 2009. Telling experts from spammers: expertise ranking in folksonomies. In *SIGIR09*, 19–23. ACM.

Noy, N. 2008. Community-based ontology development, alignment, and evaluation. http://kmi.open.ac.uk/events/sssw08/presentations/Noy-SSSW_2008_Noy.pdf.

Pan, J., Taylor, S. & Thomas, E. 2009. Reducing ambiguity in tagging systems with folksonomy search expansion. In *Proceedings of the 6th European Conference on the Semantic Web: Research and Applications*, 669–683. http://link.springer.com/chapter/10.1007/978-3-642-02121-3_49.

Papadopoulos, S., Vakali, A. & Kompatsiaris, Y. 2011. Community detection in collaborative tagging systems. *Community-Built Databases*, 107–131. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-19047-6_5.

Pera, M. S. & Lund, W. 2009. Folksonomies and similarity matching. *Journal of the American Society for Information Science and Technology* **60**(7), 1392–1406.

Perry, S. 2009. Jumper 2.0. http://theology.asin.web.id/IT/tulisan-770/Jumper-2.0-Enterprise_474_theology-asin.html.

Peters, I. 2009. *Folksonomies: Indexing and Retrieval in Web 2.0*, **1**. Walter de Gruyter.

Peters, I. & Weller, K. 2008. Tag gardening for folksonomy enrichment and maintenance. *Webology* **5**(3), 1–18.

Poorgholami, M., Jalali, M., Rehan, S. & Asgari, T. 2013. Spam detection in social bookmarking websites. *IEEE 4th International Conference on Software Engineering and Service Science*, 56–59. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6615254.

Quattrone, G., Capra, L., De Meo, P., Ferrara, E. & Ursino, D. 2011. Effective retrieval of resources in folksonomies using a new tag similarity measure. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 545–550. http://arxiv.org/abs/1207.6033.

Rásto, K., Tvaroek, M. & Bielikova, M. 2013. Web search results exploration via cluster-based views and zoom-based navigation. *Journal of Universal Computer Science* **19**(15), 2320–2346.

Ravendran, R. 2012. Usability evaluation of a tag-based interface. *Journal of Usability Studies* **7**(4), 143–160.

Rodenhausen, T., Anjorin, M., Garcia, R. D., Rensing, C., Steinmetz, R. & Dominguez, R. 2012. Ranking resources in folksonomies by exploiting semantic information categories and subject descriptors. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, 11. ACM.

Rossi, L. C. D. 2006. Folksonomies: tags strengths, weaknesses and how to make them work. http://www.masternewmedia.org/news/2006/02/01/folksonomies_tags_strengths_weaknesses_and.htm.

Saab, D. J. 2011. An emergent culture model for discerning tag semantics in folksonomies. In *Proceedings of the 2011 iConference (iConference11)*, 552–560. http://portal.acm.org/citation.cfm?doid=1940761.1940837.

Sen, S. Harper, F. M., Lapitz, A. & Riedlet, J. 2007. The quest for quality tags. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, 361–370. ACM. http://www.grouplens.org/system/files/group07-sen.pdf.

Sen, S., Lam, S. & Rashid, A. 2006. Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, 181–190. ACM Press. http://portal.acm.org/citation.cfm?doid=1180875.1180904.

Sen, S., Vig, J. & Riedl, J. 2009. Tagommenders: connecting users to items through tags. In *WWW 2009 MADRID! Track: Social Networks and Web 2.0/Session: Recommender Systems*, 1–10.

Shvaiko, P. & Euzenat, J. 2008. Ten challenges for ontology matching. In *On the Move to Meaningful Internet Systems: (OTM 2008)*, Meersman, R., Tari, Z. & Herrero, P. (eds). Springer Berlin Heidelberg, 1164–1182.

Sigurbjrnsson, B. & van Zwol, R. 2008. Flickr tag recommendation based on collective knowledge. In *Proceeding of the 17th International Conference on World Wide Web (WWW08)*, 327–336. ACM Press. http://portal.acm.org/citation.cfm?doid=1367497.1367542.

Solskinnsbakk, G., Atle, J., Haderlein, V., Myrseth, P. & Cerrato, O. 2012. Quality of hierarchies in ontologies and folksonomies. *Data and Knowledge Engineering* **74**, 13–25.

Solskinnsbakk, G. & Gulla, J. 2011. Semantic annotation from social data. In *4th International Workshop Social Data on the Web (SDoW 2011)*. https://files.ifi.uzh.ch/ddis/iswc_archive/iswc/pps/web/iswc2011.semanticweb.org/fileadmin/iswc/Papers/Workshops/SDoW/sdow2011_paper_3.pdf.

Solskinnsbakk, G. & Gulla, J. A. 2001. Mining tag similarity in folksonomies. In *Proceedings of 3rd International Workshop on Search and Mining User-Generated Contents*, 53–60.

Sood, S. C., Owsley, S. H., Hammond, K. J. & Birnbaum, L. 2007. TagAssist: automatic tag suggestion for blog posts. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM 2007)*.

Sordo, M. Gouyon, F., Sarmento, L. & Celmaet, O. 2013. Inferring semantic facets of a music folksonomy with Wikipedia. *Journal of New Music Research* **42**(4), 346–363.

Spiteri, L. 2011. Faceted navigation of social tagging applications. https://www.academia.edu/2374931/Faceted_navigation_of_social_tagging_applications.

Spiteri, L. F. 2007a. Structure and form of folksonomy tags: the road to the public library catalogue. *Webology* **4**(2), article no. 41.

Spiteri, L. F. 2007b. The structure and form of folksonomy tags: the road to the public library catalog. *Information Technology and Libraries* **26**(3), 13–25.

Stock, W. G. 2007. Folksonomies and science communication. *Information Services & Use* **27**, 97–103.

Strohmaier, M., Christian, K. & Kern, R. 2009. Why do users tag? Detecting users motivation for tagging in social tagging systems. In *Fourth International AAAI Conference on Weblogs and Social Media*, 339–342.

Strohmaier, M., Helic, D., Benz, D., Orner, C. K. & Kern, R. 2012a. Evaluation of folksonomy induction algorithms. *ACM Transactions on Intelligent Systems and Technology (TIST)* **3**(4), 74.

Strohmaier, M., Körner, C. & Kern, R. 2012b. Understanding why users tag: a survey of tagging motivation literature and results from an empirical study. *Web Semantics (Online)* **17**(C), 1–11.

Sturtz, D. N. 2004. Communal categorization: the folksonomy. *INFO622 Content Representation*. http://davidsturtz.com/drexel/622/communal-categorization-the-folksonomy.html.

Syn, S. Y. & Spring, M. B. 2009. Tags as keywords comparison of the relative quality of tags and keywords. *Proceedings of the American Society for Information Science and Technology* **46**(1), 1–19.

Szomszor, M., Alani, H., Cantador, I., Hara, K. O. & Shadbolt, N. 2008. Semantic modelling of user interests based on cross-folksonomy analysis. In *ISWC 2008*, LNCS **5318**, 632–648. Springer-Verlag Berlin Heidelberg.

Tang, R., Zuo, J., Xu, K., Zheng, J. & Wang, Y. 2010. An intelligent semantic-based tag cleaner for folksonomies. In *2010 International Conference on Intelligent Computing and Integrated Systems (ICISS)*, 773–776. IEEE.

Taylor, A. 2004. *The Organization of Information*, 2nd edition. Libraries Unlimited.

Trant, J. 2006. Social classification and folksonomy in art museums: early data from the steve.museum tagger prototype. *A Paper for the ASIST-CR Social Classification Workshop*, 1–27. http://www.archimuse.com/papers/asist-CR-steve-0611.pdf.

Tudorache, T., Noy, N., Tu, S. & Musen, M. 2008. Supporting collaborative ontology development in Protg. *The Semantic Web-ISWC 2008*, 17–32. http://link.springer.com/chapter/10.1007/978-3-540-88564-1_2.

Vallet, D., Cantador, I. & Jose, J. M. 2010. Personalizing web search with folksonomy-based user and document profiles. In *ECIR 2010*, LNCS **5993**, 420–431. Springer-Verlag Berlin Heidelberg.

Vanderlei, T. & Durão, F. 2007. A cooperative classification mechanism for search and retrieval software components. In *Proceedings of the 2007 ACM Symposium on Applied Computing (SAC07)*, 866–871. ACM Press. http://portal.acm.org/citation.cfm?doid=1244002.1244192.

Veres, C. 2006. Concept modeling by the masses: folksonomy structure and interoperability. In *Conceptual Modeling - ER 2006*, LNCS **4215**, 325–338. Springer-Verlag Berlin Heidelberg.

Waltl, B. 2012. Partitioning instead of clustering: an alternative approach for mining structured content in folksonomies. http://www.matthes.informatik.tu-muenchen.de/file/b14ej8z7mffa/sebis-Public-Website/-/Guided-Research-Bernhard-Waltl/Guided-Research-Bernhard-Waltl.pdf.

Wang, J., Hua, X. S. & Li, S. 2009. Tag refinement by regularized LDA. In *Proceedings of the 17th ACM International Conference on Multimedia*, 573–576. http://dl.acm.org/citation.cfm?id=1631359.

Wang, M., Ni, B., Hua, X. S. & Chua, T. S. 2012. Assistive tagging. *ACM Computing Surveys* **44**(4), 1–24.

Warden, P. 2007. Can you automatically generate good tags. http://petewarden.com/2007/12/03/can-you-automat/.

Wei, X. Peng, F. & Dumoulin, B. 2008. Analyzing web text association to disambiguate abbreviation in queries. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 751–752. ACM. http://dl.acm.org/citation.cfm?doid=1390334.1390485.

Wetzker, R. & Bauckhage, C. 2010. I tag, you tag: translating tags for advanced user models. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, 71–80. ACM.

Whatever, S. A. 2009. Knowledge plaza. http://theology.asin.web.id/IT/tulisan-770/Knowledge-Plaza_2639_theology-asin.html .

WordNet 2010. WordNet. http://wordnet.princeton.edu/.

Xia, Z., Peng, J., Feng, X. & Fan, J. 2013. Automatic abstract tag detection for social image tag refinement and enrichment. *Journal of Signal Processing Systems* **74**(1), 5–18.

Xu, G., Zong, Y., Jin, P., Pan, R. & Wu, Z. 2013. KIPTC: a kernel information propagation tag clustering algorithm. *Journal of Intelligent Information Systems* **45**(1), 1–18. http://link.springer.com/10.1007/s10844-013-0262-7.

Yang, H. & Lee, C. 2011. Identifying spam tags by mining tag semantics. In *3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA)*, 263–268. http://ir.nuk.edu.tw:8080/ir/bitstream/310360000Q/11836/2/icmia2011.pdf.

Yazdani, S. Ivanow, I., Analoui, M., Berangi, R. & Ebrahimi, T. 2012. Spam fighting in social tagging systems. In *Sociinfo 2012*, LNCS **7710**, 448–461. Springer-Verlag Berlin Heidelberg.

Yoo, D., Choi, K., Suh, Y. & Kim, G. 2013. Building and evaluating a collaboratively built structured folksonomy. *Journal of Information Science* **39**(5), 593–607.

Yoo, D. & Suh, Y. 2010. User-categorized tags to build a structured folksonomy. In *Second International Conference on Communication Software and Networks (ICCSN10)*, 160–164. IEEE.

Zanardi, V. & Capra, L. 2008. Social ranking: uncovering relevant content using tag-based recommender systems. In *Proceedings of the 2008 ACM Conference on Recommender Systems*, 51–58. ACM. http://dl.acm.org/citation.cfm?id = 1454018.

Zhang, L., Tang, J. & Zhang, M. 2012. Integrating temporal usage pattern into personalized tag prediction. In *Web Technologies and Applications*, Sheng, Q. Z., Wang, G., Jensen, C. S. & Xu, G. (eds). LNCS **7235**, 354–365. Springer-Verlag Berlin Heidelberg.

Zhdanova, A. & Shvaiko, P. 2006. Community-driven ontology matching. In *The Semantic Web: Research and Applications*, Sure, Y. & Domingue, J. (eds). LNCS **4011**, 34–49. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/11762256_6.