ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 12(43), DOI: 10.17485/ijst/2019/v12i43/146432, November 2019*

Core Architecture and Design Issues of Collaborative Tagging Applications

Fouzia Jabeen* and Shah Khusro

University of Peshawar, Peshawar, Pakistan; fouzia.jabeenphd@uop.edu.pk, khusro@uop.edu.pk

Abstract

Objectives: Collaborative social tagging applications like Flickr, Del.icio.us, Pinterest, and Technorati have gained popularity among people in a very short period. The primary reason is the sharing and provision of metadata in the form of tags taking advantage of users' own vocabulary. There are two major benefits of user assigned tags, firstly, it enhances the understandability of the resources and secondly, it is used for search and retrieval. The study focuses on the analysis of aspects such as design decisions, candidate solutions, their implications on different important aspects of collaborative system, and available relevant technologies. **Method/statistical analysis:** This study gives review of the architectural and design aspects of collaborative tagging systems. **Findings:** The right choice of design decisions and technologies will greatly influence the performance of the systems and will be a step towards a well-built folksonomy system. **Application/improvement:** Our study will be beneficial for researchers, designers, and developers of collaborative tagging applications.

Keywords: Collaborative Tagging Application, Collaborative Application Design, Architecture, Folksonomy, Social Networks

1. Introduction

Collaborative tagging applications allow its users to annotate resources(s) with keywords, often called tags. The resource can be a photograph, URL, computer game, music or video, etc. Collaborative tagging is known by several names which are social tagging, social indexing, social classification, and folksonomy. Broad folksonomy and narrow folksonomy are alternative names used for collaborative and simple tagging, respectively.[1]

Comparative analysis of meaning of some of the related terms is necessary before going into depth of the design issues of collaborative tagging applications. *Folksonomy and Personomy*-folksonomy is the term used to refer to users' vocabulary. Personomy refers to the user's own vocabulary.[2] The essence of folksonomies is the combined wisdom of web users in the collection.[3] *Folksonomy and Taxonomy*-taxonomy is hierarchy of pre-defined categories which users can use to classify resources. Folksonomies, however, is opposite to taxonomy in many aspects. Firstly, folksonomies are flat (that is, there are no parent–child (hierarchical) relationships. Secondly, folksonomies are completely uncontrolled as opposed to taxonomy where an essential part of creating it is deciding what names are to be used for identifications of entities. In a folksonomy, there can be a variety of different words for the same concept. Any relationships (co-occurrence, subsumption, etc.) you see in a folksonomy can be derived mathematically (statistical clustering), utilizing external knowledge sources (Word Net, DBpedia, Wikipedia, Yago). In Taxonomy, tags are added by the content creator or by the author. On the other hand, folksonomy tags are added by anyone (content creator, annotator, consumer, reader, etc.).

For example, Flickr keywords (where the user add their own tags/keywords) for describing a photo.[4] *Folksonomy, Ontology and Folktology*-folksonomies and ontologies can be placed at the two opposite ends of a categorisation spectrum[5] in the sense that folksonomy is relaxed and ontology is very strict. Ontology, in the field of information science, is basically a study and definition of the categories of things that exist or may exist in some

**Author for correspondence*

domain. Combining folksonomy with ontology becomes folktology.[6,7]

The main contribution of this study is; it gives comparison of design issues, their possible solutions, impact of these decisions on important aspect of collaborative application and relevant available technologies. We focused on two core collaborative application architecture components which include the following: a) Tags, Resource, and User b) Browsing and Searching. The reason for choosing these two aspects are first, every collaborative application will have to store these three core components (Tag, Resource, and Tagger/user) and secondly, these tags are then used for searching and browsing resources. Precise search and retrieval result is one of the principal objectives of a successful quality collaborative application.[8,9]

The significance of the study is: it helps in selection of appropriate candidate solution(s) and available technologies that will have significant impact on the gathered users' vocabulary, quality of folksonomy, indexing depth and precision in retrieved results. In comparison with the work of[10,11] which only describes software components, this study covers the possible solutions for the design issues and their implications.

The study is organised according to the framework as shown in Figure 1. The study sections are planned as follows: Sections 2 and 3 outline the design issues, comparative analysis and impact of the available solutions related to tags, resources and user's management, as well as browsing and searching components of collaborative tagging application. Section 4 highlights the comparative study of the available technologies. Lastly, in Section 5 we have tabulated the summary and presented conclusions in Section 6.

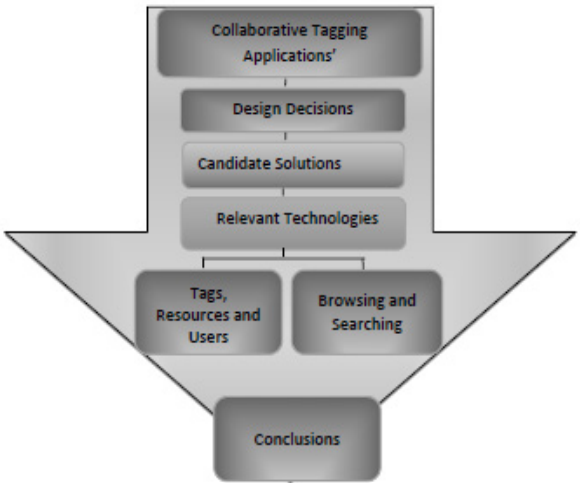


Figure 1. Framework

2. Collaborative Tagging Environment

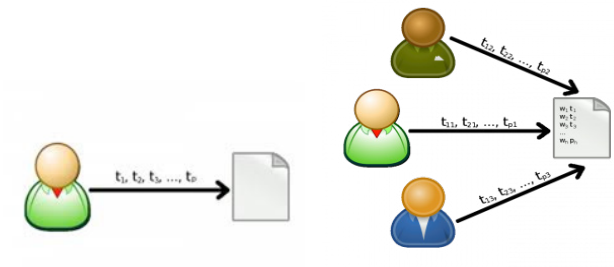
2.1. Tagging Types

Tagging approaches can be broadly classified as simple or collaborative in nature. Photo Gallery and Flickr when launched utilised simple tagging system. They allowed annotation to the original resource only. Combined view of the tags for a particular resource was not available. Figure 2(a) is pictorial representation of simple tagging in which user annotates a resource using different tags (tag_i), where *i* indicates the tag number.[12] However, in Del.icio.us and Amazon user tags pointer rather than a resource.

In collaborative tagging, each end-user has their own collection of tags for a resource and tags from all the users are grouped to give a combined view. That is the reason for calling this model a collaborating tagging model.[13] Figure 2(b) shows this model.[12] Every resource is annotated with tag tag_{ij}. Tagger (the annotating user) is denoted by *j and i* is the number of a tag. As more than one user annotates the same resource, a weighted collection of tags is there for each resource represented with $wtag_i$. The alternative names for collaborative tagging and simple tagging are broad folksonomy and narrow folksonomy, respectively.[1]

2.2. Collaborative Tagging Architecture and Design Styles

The architecture includes analysis of all the abstract design choices that require attention when developing a tagging application.



(a) (b)

Figure 2. (a) Simple tagging. (b) Collaborative tagging.[12]

2.2.1. Design Issues Related to Triple and Their Impact

A tagging system pyramid (also called triplet) is composed of three parts: users, tags, and resources. In this section, we have outlined the key decisions along with the possible solutions and their impact on different aspects of social applications.

2.2.1.1. User Part Design Decisions

The first part, *user*, the issues are membership and community.

- *Membership*-is concerned with the skill to add resources, to build tags, and to observe/view other folks tags and resources. In *Public sign-up* people become part of the system by simply signing up themselves. *Invitation only* option allows the existing members to invite other members. *External criteria/private* membership criteria are based on subjective factors. For example, if an organisation is using internal social bookmarking application, the staff members can be members.
- *Impact*-if membership is not determined explicitly then we end up in same category as Yahoo Podcasts where there is no obvious relationship between tags and users and one cannot even find their own submitted tags easily.[14] Through membership we can keep in touch with *turnover* (The rate at which users join and leave), and *activity* (how much a user posts resources and tags). User activity determines the dynamism and volume of the resources and tags.
- *Community*-means the way users interact. The connection can be in the form of followers, contacts, or groups. *Followers, is* a one-way linkage between a subscriber and content producer. A user can follow any other user(s). Del.icio.us is one of the most prominent applications supporting this concept. *Contacts,* a user requests to other user to add him/her to their contact list, upon acceptance a two-way linkage occurs. Facebook is an example that works on this model. *Groups,* means one or more users make a group to share same set of resources related to a topic with group members. Permission may be required from administrator in order to join the group.
- *Impact*-these connections act as building blocks for social navigation: discovering information by following the streams of other users. In addition, passive users watch and copy the active users' style of tagging rendering active users act as role model.

2.2.1.2. Resource Part Design Decisions

For the second pillar of collaborative application that is *resource*, we need to take decisions about contributions, original/pointer, privacy, and restrictions.

- *Contribution*-contribution to the system can either be *user-contributed or system-contributed resources.* Resources added by the users on regular basis are termed as user contributed resources while system *contributed resources* are comprised of pre-existing databases or artifacts. Images uploaded on Flickr are an opt example for user contributed resources. On the other hand, Amazon product catalog serves as system contributed resources and thus cannot be mutated by the end users.

  Impact-these models are not exclusive in usage, for example, Library Things combines both models (user + system contributed resources). If resources are being contributed by users, and there is a continual stream of additions. How fast that stream moves totally depends on activity of users.

If resources are already part of the system or if they are being added by another process, then the rate and size of change in the collection of tags and resources might not correlate with the users' activity.

For user-contributed resources, submission of resources and assigning tags to them is done at the same time. In contrast, for resources already present in the system, users may not feel strong motivation to tag them.

Depending on the dynamism of resource collection, navigation and visualisation tools are designed to be able to show relevant trends. In conclusion, the speed of resource collection stream influences navigation design decisions.

- *Original/pointer*-is the second design decision about what is being tagged. That is, either the original resource or the pointer. *Original resource* can be an actual video or a picture, etc. In case of multi users involved in a tagging activity with respect to a resource, all the tags are assigned to the resource itself. In case of pointer, the tags are not attached to actual resource but to the record in the database that exists for that particular resource. In other words, the record for that particular resource is tagged. One resource can have

multiple pointers (as many as the people annotating the resource).

Impact-the systems that allow tagging original resource lack collaboration, although it is possible to aggregate tags given by different users. Technically, there are very few cases where one can tag the actual resource.

- *Privacy*-refers to the ability to view or modify a resource. Privacy can be broadly categorised into four types. *Open to public* entails each link or resource is accessible to all web users, for example, Del.icio.us. *Configurable but public by default* allow to set the privacy settings selective for visibility of comments and tags, e.g., Flickr. By making the default setting public, people are motivated to share the resources and tags. *Configurable but private by default* make default setting to private and able to configure according to the needs. Lastly, in *everything is private* all resources and tags are private with no option to share.

  Impact-the sociability of tagging system depends greatly on default privacy settings.
- *Restriction* means what is not allowed. A system may put restriction on resources by their types. In case of *File Type* as Flickr only allows image files. In case of *Object* people a notate books of library with tags for example.

  Impact-these restrictions are usually obvious to the users, for example, it is not possible that a user joins a photo-sharing site and starts tagging/sharing web bookmarks.

2.2.1.3. Tag Part Design Decisions

Tags are keywords attached to the resource by the users. In[15] classified tags as *traditional tags* (e.g., #tag) and *non-traditional tags* (geo tags, system tags). Heckner[16] further classified the traditional tags into three categories linguistic, functional, and tag-to-text. Irrespective of tag type, decision choices fall in permission, truth and control categories.

- *Permission*-deals with the issue that who can edit, remove, or add a tag. Mostly, the contributor of the resource(s) is allowed to add, edit, or remove tags. However, the permission is extended to friends as well (Flickr is an example). In case of system resources, the first permissions-related issue is whether to allow tagging or not. Secondly, should user be allowed to tag just a subset of resource collection or entire collection is at its disposal. These choices are specific to tagging application.

  Impact-it is possible to gather a lot more metadata if users are allowed to tag resources without logging in, signing up, or performing some other action that confirms their identity. However, when users' ability to find resources is inhibited due to inefficacy to distinguish one user tags from another, users quickly lose motivation to tag.
- *Tag sharing*-means tag added by one user is available to other users. Precise choice can either be public, private or community/group.[17]

  Impact-conformity theory observed in folksonomies indicates that the tags that users observe from other users will influence their own choice of tags. The proof of this is that once a tag becomes popular it remains popular.[18] So, a worthy tag selection by the users will have significant impact on other users' tag selection.
- *Selection of tags*-to handle choice of the tags, we need to find out whether a user participates somehow in the selection of tags, or should the system handle the selection automatically. If the system gives opportunity of tag sharing at more than one level, then tag selection could be based on these levels. The two possible approaches for this can be either user is given a choice that from which level tags needs to be selected or system separates tags for the different levels, like public and private set of tags. In case of broad tagging, the tags' popularity (number of times a particular tag is chosen) can be used to select a subset representing the most popular tags.[17]
- *Tag scope*-scope of a tag refers to whether a tag belongs to an individual user or is shared by a community. There are two possible choices that are broad and narrow tagging as discussed in Section 2.1.[17]

  Impact-broad tagging is collaborative in contrast to narrow tagging. In narrow tagging there is restrictive collaboration.
- *Tag support*-this issue regards whether the user is provided with the help or suggestions when adding tag(s). Possible choices are blind tagging, viewable tagging and suggestive tagging. In blind tagging, a user is not able to see tags entered by other users while tagging. With viewable tagging, the user can view the tags added to an item by other users while tagging. The last category is suggestive tagging, where the system recommends appropriate tags to the user.

Impact-thinking about tags to be associated with a resource is time consuming.[19] Suggestive tagging reduces cognitive load on the users, as the users need to choose one or more recommended tags instead of thinking and summarizing a source to find out the best descriptive tags.

- *Tag format and control decisions*-format means the allowed character set for tags, and whether a tag will contain one or more words or even one or more sentences. It is not possible to list all possible tag formats sowing to large number of possibilities based on varying changes in character sets. However, a better approach can be to give some general categories, focusing on the degree of restriction. An open-minded approach is to let users make tags in any way they want, placing no constraints on the number of words or the type of characters permitted. This can lead to tags resembling full sentences. In contrast to this approach, tags can be limited to single words and only alphabetical characters.[17]

  Impact-degree of restriction has direct impact on collected tags' quality.
- *Control decision*-is all about restrictions on using certain words as tags, for example, Amazon do not permit common expletives to be used as tags.

  Impact-it is very critical not to exercise too much strictness and control. Users might use tags that express an opinion that may not be liked but is not otherwise objectionable. If users feel that they cannot express or are being unreasonably censored, they may leave the system.
- *Truth*-is the term used for storage place of tags. *Truth in the file* refers to tags deposition in a file (resource). The other substitute is *truth in the database* where tags are kept separate from the resource, mostly in the database.

  Impact-truth has some practical implications. If system is a social bookmarking and it is implemented on intranet, the tags will be included in database rather than the documents themselves. For the tags used to bring precision in search results, it is required that search engine indexes the tags in the database as well as with the documents.

2.3. Data Models/Candidate Solutions

To store tags, users, resources, and other related data, we can use relational databases like SQL Server, Microsoft Access, or open source MYSQL. Another alternate choice can be a triple store.

2.3.1. Del.icio.us Schemas

MySQLicious, *Scuttle*, and *Toxis chem as are* popularised and used by Del.icio.us.[20] *MySQLicious* is a denormalised solution. In this schema, a single table is used as shown in Figure 3.[20] The most prominent influence of this schema is its simple design. Usage and querying is based only on one table. The queries are very simple and straightforward. For a very small database, this schema might be noticeable because of its simplicity. Although simplicity is a positive side of this design, it suffers from duplication of data due to the denormalised nature of the schema. The schema is not flexible. Certain queries for obtaining the relationships that exist between users would be hard to code and slow to carry out. With a single table describing every entity, a database developer would require complex and time-consuming queries for retrieval of information on specific entities.

Scuttle schema shown in Figure 4 provides solution in two tables.[20] The *sc Categories* stores tags and contains a foreign key to the second table of sc *Bookmarks*. The prominent pros of this technique are; it is more normalised as compared to MySQLicious. User can have many tags for each bookmark. There is no restriction on repetition of tag textual data in the sc*bookmark* table. To show the relationship between bookmark and the bookmark's set of tags, foreign key can be implemented. If we look at the bookmarks table, only single tag is stored per row. This is the reason the Scuttle is not, however, precisely normalised and has limitations. A surrogate key is included in the sc Categories table (id) in place of primary key tuple on bid and category. Memory utilisation is increased because in tags table, Id of the tag text (Category) is not used and textual tag data is used. As a result, a small number of index records are being able to accommodate into a single

delicious
id
url
description
extended
tags
date
hash

Figure 3. MySQLicious.[20]

index block. Which will reduce the efficiency and speed of index accesses.

Existing schema does not seem to be successful in accurately retrieving statistical information like total number of bookmarks for each tag. The reason is one-to-many relationship is unable to represent or mis-represent the true many-to-many relationship that is present between the bookmarks and tags.

The three table solution called Toxi shown in Figure 5 came up to overcome the deficiencies of above two schemas.[20] The *tagmap*table manages n-to-m relation between the tags and bookmarks. Word Press utilises this schema. The good points of the schema are first, this is the most normalised solution (that is, you can go for 3NF). Secondly, doing so, removes the issue of duplicated information (particularly the textual tag data) that is present in the other two designs. However, the schema is not free from disadvantages. The cons are, when changing or deleting bookmarks one will get tag-orphans. The other main drawback of Toxi schema (as is the case with most properly normalised schema) is that there are now additional tables to manage and administer.

2.3.2. *Schemas for Simple and Collaborative Tagging*

The above presented solutions are altered and improved by[14] and give us general model for simple and collaborative tagging applications. These models can be adopted for a wide range of applications by making application specific changes and modifications.

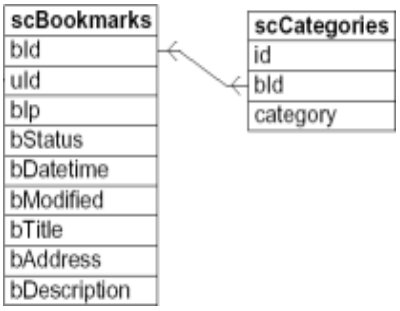


Figure 4. Scuttle solution.[20]

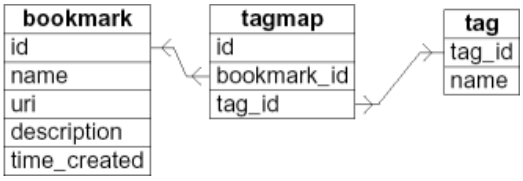


Figure 5. Toxi solution.[20]

Simple tagging model as represented in Figure 6,[14] make use of three tables (*user, tags, and resources*) to keep users, tags, and resources. One additional table called *resources tags* is included for normalisation purpose. In this model, each resource is unique to each user and has multiple tags.

Extraction of data and querying simple tagging model is easy. For example, the query shown in Figure 7 is used for finding tags assigned to a particular resource.[14] Figure 8 represents a query in which usage count of a tag is determined.[14] This metric can then be utilised as input to font size, representing the frequency of occurrence in tag cloud. To conclude, we can say that this simple model can be effectively used for personal information management applications.

Collaborative tagging model is represented in Figure 9.[14] There is a slight change. The table *resources_tags* is removed from and replaced with intermediary table called *users_resources_tags*. This table is required because in this model more than one user are allowed to assign tag(s) to a particular resource. This table maintains the entries for ids of user(s), tag(s) and resource every time tag is assigned by the user to a particular resource.

Two sample queries are shown in Figures 10 and 11.[14] The first query extracts tags given by any particular user and the second query retrieves hundred and fifty most popular tags, which can be used to be placed on the tag cloud. Furthermore, resources can also be filtered by

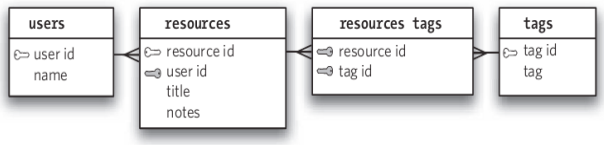


Figure 6. Simple tagging system data model.[14]

```
SELECT tag FROM resources_tags
INNER JOIN tags ON
resources_tags.tag_id = tags.tag_id
WHERE resources_tags.resource_id =1
```

Figure 7. Query for obtaining tags for a single resource.[14]

```
SELECT t.tag, COUNT(t.tag_id) AS tag_count
FROM resources r
    INNER JOIN resources_tags rt ON r.resource_id = rt.resource_id
    INNER JOIN tags t ON rt.tags_id =t.tag_id
    INNER JOIN users u ON r.user_id = u.user_id
GROUP BY t.tag, u.user_id
HAVING u.users_id = 1
```

Figure 8. Retrieving tags with number of times it occurs.[14]

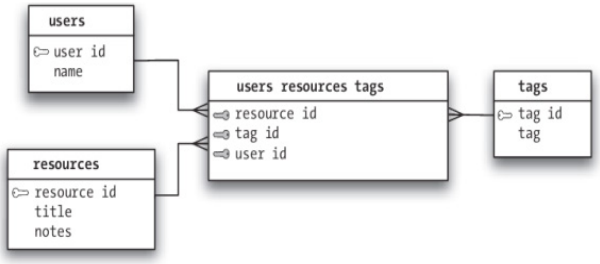


Figure 9. Collaborative tagging system data model.[14]

```
Select tag, COUNT (*) As tag_count FROM
users_resources_tags urt, tags t WHERE
urt.users_id =1 AND t.tag_id = urt.tag_id GROUP
BY tag
```

Figure 10. Retrieving individual user's.[14]

```
SELECT tag, COUNT(*) AS tag_count
FROM users_resources_tags urt, tags t
WHERE t.tag_id = urt.tag_id
GROUP BY tag
ORDER BY tag_count DESC
LIMIT 150
```

Figure 11. Retrieving 150 most accepted tags.[14]

using more than one tag in combination. We can say that this model give the provision of handling an arbitrary number of tag combinations.

3. Browsing and Searching

Searching, browsing, and querying are prominent and effective means used for information finding. These methods are utilised according to the objective of information seeker. For example, sometimes information seekers require a direct "right answer", while sometimes they intend to retrieve several relevant related sources and sometimes in-depth search is necessary and then there are times when required information have to be re-found. Tagging is effective in "exploratory seeking," that is, when the goal of searching is to gain a broad perspective of the information available about a particular topic.[21] In both ways of search either by browsing or by query, the difference is the cognitive load.

3.1. Search via Browsing Approaches, Issues and Their Impact

3.1.1. Facets

Browsing interface is designed in such a way that it contains a set of meaningful labels which reflect concepts of the domain. The name given to such design is called the facets. The major decisions involved in facets choice are; choice of facets, number of facets, facets chooser, and evaluation of quality of facets.[22]

- *Choice of facets*-given the vast scope of the subjects covered in atypical social tagging application, the questions that arise in mind are; how to choose facets that can be applied generally to all these subjects or domains? Whether the choice of facets will be tested by other parties or by the users, and how effectively these facets reflect subject domain and clients' interest?
- *Number of facets*-we need answers that how many facets will be sufficient that will cover site members' needs? There is no agreement on the optimal number of facets to use in social tagging applications.

  Impact-multiple facets empower a more flexible, improved and enhanced search. It supports classification of objects of interests along several dimensions of metadata.[23,24]
- *Facet chooser*-we need to decide who will be the chooser of the facets? The choices can be: site members, site editors, and researchers.

  Impact-the diverse nature of the site members in such applications will make reaching of consensus a difficult task.

Once a list of facets is chosen, the set of questions could be added to the user's interface, so that when the site members create or re-use a tag, the questions could help them determine in which facet to place the tag.[25]

- *Quality of facets*-we have to answer these questions: whether facets that are chosen reflect the needs of the site members? The extent to which the facets reflect the needs of the site members? Their ease of use, how much they are used, and whether they are used correctly (e.g., location-type tags are placed correctly in the location facet). Some site members may place tags in the correct facets in that case how the use of the facets will be monitored?

  Impact-if facets are not used correctly, then their potential benefits may not be realised fully.

3.1.2. Tag Cloud

Browsing is supported by visualisation techniques like Tag Cloud.[26] Tag clouds perform best in browsing scenarios where specific target is not clearly known. Tag cloud alone is not sufficient enough for information retrieval tasks, but could be used to boost a user's search activity.[27]

- *Shape (Layouts of tag clouds[28])*-which layout is a more appropriate circle, rectangle, a square, 2D, or 3D. Secondly, how tags should be arranged whether alphabetically or according to popularity (In circle form, least important to most important).

Impact-semantic, folksonomy-based layouts perform significantly better in searching as compared to random layouts. However, semantic layouts should only be used when the quality of tag organisation can be assured.[29]

- *Number of tags*-it is required to determine the number of tags to be displayed in a tag cloud. Impact-the selection should be based on the fact that it puts least cognitive load on the user.
- *Coloring, visual effects and interpretation of visual effects*-visual effects gain audience attention. Like large font size attracts user more towards a tag than a tag with small font size. The meaning hidden behind the font size is the frequency of use. Web Designer Wall[1] has a small tag cloud that uses size to indicate importance and frequency.

  Impact-font size strongly affects recall.[30]
- *Selection of tags*-how to select tags to be appeared on the tag cloud. Flickr Tag cloud displays the most popular tags of all times and hot tags of last 24 hours. Delicious displays list of top tags and also individuals have their own tag clouds.[31]

Impact-tags on tag cloud can be selected keeping one or more goals in mind. An example of tag selection based on single-objective would be to select only popular tags. A multi-objective algorithm, on the other hand, would keep balance among two or more goals, e.g., selecting tags that are popular and provide excellent coverage.[32]

3.1.3. Grouping Related Tags, Resources and Folks

Related tags, resources, and folks can be grouped to enrich browsing and to achieve required search objectives. The social bookmarking site Del.icio.us allows folksto put related tags into bundles. However, one tag may appear in many bundles[33].

Bibsonomy tag hierarchies[34]; sites do not limit on the ways of use of hierarchies. However, these are mostly used in finding out relationships (broader/narrower (most common)).

There are a number of researches conducted to find relatedness among tags, resources, and users.[35–38] Finding of related folks helps in finding group of people with common interests which further helps in finding related resources and tags.[35–38]

3.2. Query Based Searching

3.2.1. Query Based/Keyword Search

Searching via tag is analogous to search by keyword.[39] To facilitate searching via query mostly collaborative applications provide query expansion and recommendations.

- *Recommendations*-recommendations can be *Explicit or Implicit*. In Explicit recommendations, user personally sends his/her recommendation(s) to other user. Implicit recommendation depends on the user's behavior that is what he/she has searched for and what articles he/he has viewed.[39]

  Impact-quality and coverage of the recommended tags is a very important matter to look into: first, indexing terms automatically generated by search engines are good for textual data but are not significantly useful for labelling and organizing resources which include video, photos/pictures, and music.[40,41] Second, this facilitates the users/taggers beyond the annotation process. For instance, approaches that suggest tags considering relations "*is-a*", "*part-of*", "*tag inheritance*" are extremely useful in increasing indexing depth of a resource.
- *No. of tag recommended*-there is no hard and fast rule regarding number of suggested tags. However, two to twenty tags seem to be optimum.

  Impact-according to the author of[42] more tags of the recommendation are regarded, better will be the recall and worse will be the precision.
- *Query modification*-query can be modified either by substitution or by expansion. In substitution original term(s) are removed and replaced with new term(s). Query expansion aims to enrich and elaborate query with related classes or senses. The question arises what terms/tags should be provided for query expansion,

1

for example, choices can be ontology terms, local or global terms,[43] multilingual tags,[44] etc.

Global analysis is based on finding alike words for expansion of a query. These words can be obtained using different techniques like from co-occurrence information, external data source (Wikipedia/ DBpedia, WordNet, or web search results).[16] In local analysis, analogous terms are retrieved from most relevant documents, which are then used for expanding the original query.

Impact-an expansion is a much safer form of query modification as compared to substitution.[45,46] The objective of query expansion is to improve recall and precision by enriching result set with most relevant or at least equally relevant pages.[47] Expansion is much more effective to precise the search results when query contains abstract terms. Expanding queries will increase clarity, reduce ambiguity (disambiguation).[48] Major causes of query expansion failure are: topic drift and poor quality of query.[49]

- *Query types*-we need to decide whether search is possible with compound tags? Is there is support for boolean search?[22] Personalised search is allowed or not?[50]

  Impact-with facets, resource collection can be filtered and retrieved in a well organised way. However, authors Choi & Street[51] objected that the user point of view is not mirrored in facets. Personalised search considers user taste, interests and needs. This is very important as users' interest changes with time.

4. Relevant Technologies

At minimum, a social tagging system will have a *client interface*; users will be able to interact with the system using this interface. In addition, security measures are implemented to provide required access levels to the users and prevent the system from spammers or unauthorised users. Second, *application server layer*, for which Apache Tomcat web server can be one of the option or the tagging system, can be relegated to cloud-based services such as

Table 1. Categorical overview of API support for collaborative applications development

Category	API	Comments	Protocol	Data format	Usage limit
Search and recommendations	Jinni[1]	Movie discovery and search service	SOAP	XML	
	Sapo tags[2]	Tag search service	REST	RSS	
	TagTooga[3]	Tag based Internet directory	REST	XML	1000 queries/ day, 1/sec max
	Evri	Recommendations and semantic search service	REST, JavaScript, Cocoa	XML, JSON, JSONP	10,000 requests per day.
	tagthe.net	Tag recommendation service	REST	JSON XML	
	Tagyu	Tag recommendation service	REST	XML	
Content annotation	GroupMe	Social bookmarking	REST	RDF, RSS	
	OpenDover	Semantic tagging webservice	REST, SOAP	XML, JSON	100 requests within the hour
	Steve In Action	Museum social tagging project	REST-RPC	JSON	
	Tagatum API: Blog tags repository services	Blog tags repository services	REST	XML	
	Reflect	Biology and chemical tagging service	REST	XML	
	Wikimeta	Semantic tagging and content annotation service	REST	XML, JSON	100 requests & 1 MB data/day
Interface	Tagul	Tag cloud creation service	XML-RPC	SVG	
	Word Cloud Maker	Word clouds from blocks of text	HTTP POST	JSON	None

Table 2. Frameworks support for collaborative applications

Frameworks	Open source	Free	Developed in	Database support	Folksonomy support	Prominent features
Xoops with Yogurt Extension	No	Yes	PhP	MySQL	–	• Lightbox2 features + jquery tabs based simple front end
Mahara	Yes	Yes	Run on LAMP	MySQL	Yes	• Learning[54–56] management system Moodle[4]
Anahita Social Engine	Yes	**Gpl2** License	–	–	Yes[5]	• *Ajaxed* **interface** • Powerful profile management & messaging capabilities • Supports *OpenID* & integration with *Amazon Cloud Storage* to ease the management of growing data
Elgg	Yes	**Yes**	Apache 2 PhP	MySQL	cross-site tagging[6] comes with a **JSON API**	• Social networking • Strong access control lists • Multiple view support (e.g., iPhone, cell phones) • Templating engine • Widget framework and more
BuddyPress	Yes	**Yes**	PhP	–	–	• WordPress theme compatibility • bbPress discussion forums • BuddyPress widgets
YaCS	Yes		PhP	MySQL	Free form Tagging[7]	• Feature-rich chat • Multiple identities • Guest books • Multiple communities
Pligg	Yes	**Yes**	PhP	MySQL	No	• Five star voting mechanism similar to Digg, Mixx and Reddit • Vote ranking
SocialEngine	Yes	**No**	PhP	–	Plugin available for enabling tagging[8] Hashtag pluging9	• Allow multiple social frameworks(two way friendship, one way follower, subnetworks for grouping) • Customisable widgets • Membership levels for community
Ektron	No	**No**	Visual studio/.Net	–	Yes	• Full featured WYSIWYG editing • Powerful combination of a commercially licensed .NET CMS with a PHP-based open source search system[9]
Pinax	Yes	**No**	Django Web Framework	–	Yes	• Support multiple channels (e.g., technical vs business) • Previewing of blog posts before publishing • Ability to announce new posts on twitter • Review comments per post for multi-author workflows • Public but secret urls for unpublished blog posts for easier review

statusNet	Yes	**No**	PhP	MySQL	Yes	• Able to use NoSQL technologies (e.g., memcached) to increase performance • StatusNet rely on open protocols like oStatus and oAuth
KicKapps	Yes	**No**	–	–	Yes	• Its SaaS platform is powered by the first Social Graph Engine for web publishers • The platform includes social networking, user-generated content, programmable video players, drag-and-drop widget building, WidgeAds and other applications that are tightly integrated with Robust Media Moderation, member organisation and management, and reporting
Lovd By Less	Yes	Yes	Ruby on Rails	–	Yes	• Flickr and YouTube integration • Activity updates and user-to-user messaging • Insoshi a new social network platform similar to Lovd By Less

Table 3. Social bookmarking platforms

Bookmarking	Open source	Free	Development tools/ technology	Support tagging/ folksonomy support	Look like/similarity with
Akarru Social Bookmarking Engine	Yes	Yes	PHP MySQL	Yes	www.blogmemes.com
Bookmark4U	Yes	Yes	Apache + PHP + MySQL	No	
Scuttle[10]	Yes	Yes	PhP MySQL	Yes	Delicious
PressMark		Yes	PhP MySQL	Yes	Delicious
sabros.us	Yes	Yes	PhP My SQL	Yes	
GetBoo	Yes	Yes	PhPMySQL	Yes	Delicious
Pligg	Yes	Yes	PhPMYSQL	Yes	Digg
Laicos	Yes	Yes	asp.net and MySQL	Yes	Digg
de.lirio.us	Yes		Perl		Delicious clone
Hotaru CMS	Yes		PhP	Yes	Digg

Amazon Web Services (AWS). Third, suitable *database* like NoSQL-based database (e.g., MongoDB) or conventional SQL-based database (e.g., MySQL). Lastly, for *inter-layer server-client* communication, XML or JavaScript Object Notation (JSON) as data interchange scheme.

Wikipedia provides a comparison between some of the available social networking software.[52–58] We have organised relevant technologies in tables from three points of views. First, API support for collaborative tagging application's development. Table 1 is the feature wise analysis of API support. Second, available frameworks ready to download and use. The developer's choice mostly depends on key factors which include open source, freeware nature, and plug-in support. For example, developers prefer *elgg*over *social engine* because it is free, customisable and plugging are freely available.

Table 4. Summary of user-related design issues and possible solutions

User-related issues	Membership	Community
Possible solutions	• Public sign up • Invitation only • Private/external criteria	• Follower • Contacts • Groups

The technology chosen has significant impact on scalability of a system. Table 2 highlights the feature wise comparison of major available frameworks. Lastly, Table 3 gives an overview of the feature set support of available platforms for social bookmarking.

5. Summary

In this section, we have tabulated the summary in Tables 4–10. Tables 4–6 present key decisions, their possible candidate solutions related to user, resource, and tag. In

Table 5. Summary of resource-related design issues and possible solutions

Resource-related issues	Resource contribution	Original/pointer	Privacy	Restriction
Possible solutions	• User contributed resources • System contributed resources	• Original resource • Pointer to resource	• Public • Private • Configurable (public/private)	• File • Object

Table 6. Summary of tag-related design issues and possible solutions

Tag design issues	Permission	Tag sharing	Selection of tags	Tag scope	Tag support	Tag format and control decisions	Truth
Possible solutions	• Owner • Friend • Tag limited resources	• Public • Private • Group	• Level Wise • Public/private groups • Popular tags	• Narrow Tagging • Broad Tagging	• Blind tagging • Viewable tagging • Suggestive Tagging	• No restriction • Single word • Restriction on certain words	• File • Database

Table 7. Design issues and their impact on different aspect of social applications

Design issues	Portability	Quality of gathered vocabulary/ folksonomy	Collaboration	Turnover	Activity	Dynamism of resource and tags	Volume of collection	Motivation to tag
Membership	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Community	No	Yes	Yes	Yes	Yes	Yes	Yes	
User contributed resources	No	Yes	No	Yes	Yes	Yes	Yes	Yes
System contributed resources	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Original/ pointer	Yes	Yes	Yes	No	No	No	No	No
Privacy	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tag sharing	No	Yes	Yes	No	Yes	Yes	No	Yes
Tag scope	No	Yes	Yes	No	No	No	No	Yes
Tag support	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Tag format and control decisions	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Truth	Yes	No	No	No	No	No	No	No

Table 7, we have analyzed impact of these design issues on various aspects of collaborative applications. Tables 8 and 9 are all about the issues and candidate solutions related to browsing and query-based search. Lastly, in Table 10, we have analyzed implications of browsing and searching approaches. Both browsing and searching have same objective that is to bring precise results in front of users. However, facets or tag cloud work well when query objective is opaque and broad. In contrast, direct query-based approach is efficient if objective of search is clear. In addition to providing facilitation in browsing, facets are helpful in organizing folksonomy.

6. Conclusions and Developing Trends

Collaborative tagging applications gained popularity because of their wide spread adoptability among users. In

Table 8. Summary of browsing-related design issues and possible solutions

Facet navigation issues	**Choice of facets**	**Number of facets**	**Facet chooser**	**Testing quality of facets**
Major possible solutions	• Reflect subject domain • Reflect user interest	• No specific number • Must cove site member's needs	• Site member(s) • Editor (s) • List of question to help users	• Site editor • Researcher
Tag cloud issues	**Shape**	**Number of tags/selection of tags**	**Coloring, visual effects and interpretation**	**Arrangement of tags**
Major possible solutions	• Square • Circle • 2D, 3D • Semantic • Folksonomy-based	• Most popular tags • Representative tags of a site • Ontology tags	• Big Font represent most popular • Near tags represent related tags	• Alphabetical • Circular (least to most used tags) • Coverage
Browsing via tag bundles/related tags/ related resources/ related folks	**Techniques to discover tags relatedness**	**Techniques to discover resources relatedness**	**Techniques to discover folks relatedness**	**Ranking resources**
Major possible solutions	• Co-occurrence • Subsumption • Synonyms	Similar tags assigned to similar resources[57]	FolkRank[58]	FolkRank

Table 9. Summary of query-related design issues and possible solutions

Recommendations	**Recommended terms**	**Number of tags recommended**
Major possible solutions	• Implicit • Explicit • Personalised	• Depends on *p*-core[59] • Ratio of precision and recall
Query modification issues	**Types**	**Terms used for substitution/expansion**
Major possible solutions	• Substitution • Expansion	• Mapping with Wiki, DBPedia, WordNet • Global/Local • Ontology terms • Multilingual equivalent • With respect to user, resource and tag context • Concrete rather than abstract terms
Query types	**Types of queries support**	
Major possible solutions	• Support for compound tags • Support for boolean search • Personalised search • Temporal	

Table 10. Browsing and searching implications

Browsing and searching approaches	Folksonomy organisation	Precision in search and retrieval	Objective of a query is clear	Objective of a query is vague
Tag cloud	No	Yes	No	Yes
Facets	Yes	Yes	No	Yes
Query	No	Yes	Yes	No

this study, we have reviewed and highlighted the design issues, the possible solutions and their impact along with a comparative analysis of the feature set support provided by the available technologies.

This study covers key decisions that the designers' needs to take when developing a collaborative tagging application by keenly looking at the implications of these decisions on the performance of the system. These key decisions have great impact on the major aspects of a collaborative application which include, flexibility they provide to the users, the emerging folk vocabulary quality and indexing depth, precision in search and retrieval, success of an application measured by observing turnover and activity. In addition, choice of appropriate technology plays a significant role in scalability and security of a collaborative system. To end with the developing trends in this area of research includes (1) augmenting collaboration with semantics and context. Contextual collaboration has the potential to improve both the working experience and the experience of working together (2) visualisation based collaborative system (3) deep learning models are effectively used for dealing with unlabeled resources and/or recommendations.

Our critical review presented in the study will be beneficial in making right choices by the designers and developers of tagging applications.

References

1. Weller K. Folksonomies and ontologies: two new players in indexing and knowledge representation. Apple Web Creation. 2007;2:108–15. http://webdesignerwall.com/general/new-themify-flow.
2. Understanding the user: personomy translation for tag recommendation. https://www.researchgate.net/publication/268330617_Understanding_the_user_Personomy_translation_for_tag_recommendation.
3. Eltemasi M, Nadar N, Niusha ZF. Communication with folksonomy in technical libraries. In: 2011 international conference on social science and humanity; 2011. P. 479–82.
4. Difference between taxonomy and folksonomy wordpress news designbit. [cited 2015]. http://designbit.co.uk/2010/10/10/difference-between-taxonomy-and-folksonomy/.
5. Lachica R, Karabeg D. Metadata creation in socio-semantic tagging systems : towards holistic knowledge creation and interchange. In: International conference on topic map research and applications; 2007. P. 160–71.
6. Folktology. [cited 2012]. http://www.artificialmemory.net/(S(ghbqnlrfgoh13l3zavdexl45))/artificialmemory.aspx?ID=22294.
7. fuzzy.com:Folktology. [cited 2011]. http://fuzzzyblog.blogspot.com/2011_09_01_archive.html.
8. Measuring the quality of an e-government folksonomy. [cited 2008]. http://marksanderson.org/publications/my_papers/E-government_Folksonomy.pdf.
9. Enhancing navigability : an algorithm for constructing tag trees. [cited 2017 Mar 21]. https://content.sciendo.com/view/journals/jdis/2/2/article-p56.xml.
10. A collaborative tagging system design framework for supporting affective design. [cited 2015]. https://ieeexplore.ieee.org/document/7385677.
11. Discovering product feature and affective associations through collaborative tagging. [cited 2014]. https://ieeexplore.ieee.org/document/7058789.
12. What are social tags? [cited 2015]. http://blog.zubiaga.org/2009/02/what-are-social-tags/.
13. The structure of collaborative tagging systems. [cited 2005 Aug 18]. https://arxiv.org/abs/cs/0508082.
14. Tagging: people powered metadata for social web. [cited 2007]. https://www.researchgate.net/publication/252059367_Tagging_People-Powered_Metadata_for_the_Social_Web_Smith_G_2008_Book_Review.
15. Social tagging & folksonomies : indexing retrieving and beyond ? [cited 2011 Jan]. https://www.researchgate.net/publication/261541547_Social_tagging_folksonomies_Indexing_retrieving_and_beyond.
16. Tree, funny, to _ read, google: what are tags supposed to achieve ? A comparative analysis of user keywords for different digital resource types. [cited 2018 Jan]. https://www.researchgate.net/publication/29858360_Tree_funny_to_read_google_What_are_Tags_Supposed_to_Achieve_A_Comparative_Analysis_of_User_Keywords_for_Different_Digital_Resource_Types.
17. Social tagging of services to support end user development in ubiquitous collaborative environments, dissertation. [cited 2007 Jun]. https://core.ac.uk/download/pdf/52104578.pdf.

18. Tagging, communities, vocabulary, evolution. [cited 2007]. http://files.grouplens.org/papers/sen-cscw2006.pdf.
19. Tag recommendation for folksonomies oriented towards individual users. [cited 2008 Jan]. https://www.researchgate.net/publication/228632258_Tag_recommendation_for_folksonomies_oriented_towards_individual_users.
20. Faceted navigation of social tagging applications. [cited 2011]. https://dalspace.library.dal.ca/handle/10222/14297.
21. Partitioning instead of clustering: an alternative approach for mining structured content in folksonomies. [cited 2016]. http://docplayer.net/160699825-Partitioning-instead-of-clustering-an-alternative-approach-for-mining-structured-content-in-folksonomies.html.
22. Finding cultural heritage images through a dual-perspective navigation framework. [cited 2016]. http://d-scholarship.pitt.edu/34253/.
23. Faceted search. [cited 2009]. https://www.morganclaypool.com/doi/abs/10.2200/S00190ED1V01Y200904ICR005.
24. Li WD, Qiu ZM. State-of-the-art technologies and methodologies for collaborative product development systems. Int J Prod Res. 2006;44(13):2525–59.
25. Seeing things in the clouds: the effect of visual features on tag cloud selections. [cited 2008]. http://www.hci.usask.ca/publications/2008/fp039-bateman.pdf.
26. Lohmann S, Ziegler J, Tetzlaff L. Comparison of tag cloud layouts: task-related performance and visual exploration. In: IFIP conference on human-computer interaction; 2009. P. 392–404.
27. Semantically structured tag clouds : an empirical evaluation of clustered presentation approaches. [cited 2009 Jan]. https://www.researchgate.net/publication/221516486_Semantically_Structured_Tag_Clouds_An_Empirical_Evaluation_of_Clustered_Presentation_Approaches.
28. Getting our head in the clouds: toward evaluation studies of tagclouds. [cited 2007 Jan]. https://www.researchgate.net/publication/221517909_Getting_our_head_in_the_clouds_Toward_evaluation_studies_of_tagclouds.
29. What is a tag cloud and how do tags connect people. [cited 2009 Mar 06]. https://www.matrixgroup.net/thematrixfiles/blog/what-is-a-tag-cloud/.
30. On the selection of tags for tag clouds. [cited 2011 Jan]. https://www.researchgate.net/publication/221520128_On_the_selection_of_tags_for_tag_clouds.
31. From the crowd to communities: new interfaces for social tagging. [cited 2008 Jan]. https://www.researchgate.net/publication/228529099_From_the_crowd_to_communities_New_interfaces_for_social_tagging.
32. Constructing folksonomies from user-specified relations on flickr. [cited 2009]. https://www.isi.edu/~lerman/papers/fp500-plangprasopchok.pdf.
33. Tag relatedness using laplacian score feature selection and adapted jensen-shannon divergence. [cited 2014]. https://link.springer.com/chapter/10.1007/978-3-319-04114-8_14.
34. Content-based and collaborative techniques for tag recommendation: an empirical evaluation. [cited 2013 Feb]. https://www.researchgate.net/publication/257580447_Content-based_and_collaborative_techniques_for_tag_recommendation_An_empirical_evaluation.
35. Uddin MN, Duong TH, Nguyen NT. Semantic similarity measures for enhancing information retrieval in folksonomies. Exp Syst Appl. 2013;40(5):1645–53.
36. Xu G, Zong Y, Jin P. KIPTC: a kernel information propagation tag clustering algorithm. J Intell Inf Syst. 2015;45(1):95–112.
37. Folksonomies. Indexing and retrieval in web 2.0. [cited 2009]. http://books.google.com.pk/books?id=HBYg36gbnegC&pg=PA289&lpg=PA289&dq=Browsing+support+in++folksonomy+system&source=bl&ots=ZJOeG02ZhF&sig=QOQo8Oj-32KqVQ0R3UzGcPik2RM&hl=en-&sa=X&ei=qAhOU72REYXgygP-tYHwDA&ved=0CDYQ6AEwAQ#v=onepage&q&f=false.
38. Adapting K-nearest neighbor for tag recommendation in folksonomies. [cited 2009]. http://ceur-ws.org/Vol-528/paper8.pdf.
39. Folksonomic tag clouds as an aid to content indexing. [cited 2009 Nov 21]. https://arxiv.org/abs/0911.4178.
40. Tag recommendations in folksonomies. [cited 2007]. https://link.springer.com/chapter/10.1007/978-3-540-74976-9_52.
41. Query expansion in folksonomies. [cited 2011]. https://dl.acm.org/citation.cfm?id=2032131.
42. Jung JJ. Cross-lingual query expansion in multilingual folksonomies: a case study on Flickr. Knowl-Based Syst. 2013;42:60–67.
43. Query reformulation using anchor text. [cited 2010]. https://ciir-publications.cs.umass.edu/pub/web/getpdf.php?id=900.
44. Analyzing web text association to disambiguate abbreviation in queries. [cited 2008 Jan]. https://www.researchgate.net/publication/221299938_Analyzing_web_text_association_to_disambiguate_abbreviation_in_queries.
45. Wikipedia. Query expansion – Wikipedia, the free encyclopedia. [cited 2015]. http://en.wikipedia.org/wiki/Query_expansion.
46. Gold-standard evaluation of a folksonomy-based ontology learning model. [cited 2018]. https://iopscience.iop.org/article/10.1088/1742-6596/971/1/012045.
47. Studying query expansion effectiveness. [cited 2009]. http://terrierteam.dcs.gla.ac.uk/publications/he09ecir.pdf.
48. He B, Ounis I. Studying query expansion effectiveness. In: Advances in information retrieval; 2009. P. 611–9.
49. A personalized recommender system based on users' information in folksonomies. [cited 2013]. http://www2013.w3c.br/companion/p1215.pdf.

50. Bringing a more accurate user's perspective into web navigation: facet analysis of folksonomy tags. [cited 2012 May]. https://www.researchgate.net/publication/42759294_Bringing_a_More_Accurate_User's_Perspective_into_Web_Navigation_Facet_Analysis_of_Folksonomy_Tags.
51. Bringing a more accurate user's perspective into web navigation : facet analysis of folksonomy tags. [cited 2012 May].https://www.researchgate.net/publication/42759294_Bringing_a_More_Accurate_User's_Perspective_into_Web_Navigation_Facet_Analysis_of_Folksonomy_Tags.
52. Comparison of social networking software. [cited 2019 Jul 08]. https://en.wikipedia.org/wiki/Comparison_of_social_networking_software.
53. Collaborative deep learning for recommender systems. [cited 2015 Jun 18]. https://arxiv.org/abs/1409.2944.
54. Semi-supervised Learning with deep generative models. [cited 2019]. https://papers.nips.cc/paper/5352-semi-supervised-learning-with-deep-generative-models.pdf.
55. TRSDL : tag-aware recommender system based on deep learning – intelligent computing systems. [cited 2018 May]. https://www.researchgate.net/publication/325198795_TRSDL_Tag-Aware_Recommender_System_Based_on_Deep_Learning-Intelligent_Computing_Systems.
56. Effective retrieval of resources in folksonomies using a new tag similarity measure. [cited 2011]. https://dl.acm.org/citation.cfm?id=2063657.
57. Hotho A, Jäschke R, Schmitz C, Stumme G. Information retrieval in folksonomies: search and ranking. In: European semantic web conference; 2006. P. 411–26.
58. A combined method for mitigating sparsity problem in tag recommendation. [cited 2014]. https://ieeexplore.ieee.org/document/6758716.