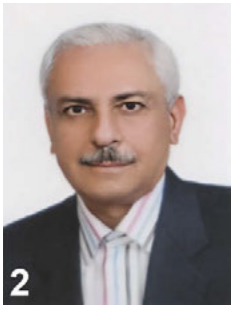
Identifying the Patterns of Author-Generated Tags to Library and Information Science Papers in The Academic Social Networks: Focusing on Academia.edu

Rasul Saadat*, Ahmad Shabani**, Asefeh Asemi***,
Mehrdad Cheshmeh Sohrabi**** and Mohammad Tavakolizadeh Ravari*****

* Islamic Sciences and Culture Academy, Pardisan, University Blv. Qom, Iran
** ****Department of Knowledge and Information Science, University of Isfahan, Azadi Ave., Isfahan, Iran
*** School of Business Informatics, Corvinus University of Budapest, Fővám tér 8, 1093 Hungary, Budapest
***** Department of Knowledge & Information Science, Yazd University; Safaeieh town, Yazd, Iran
*saadat.rasul@gmail.com, **shabania@edu.ui.ac.ir, ***asemi.asefeh@unicorvinus.hu,
****mo.sohrabi@edu.ui.ac.ir, *****tavakoli@yazd.ac.ir

Rasul Saadat holds a PhD in Library and Information Science and a researcher and librarian. He is currently in charge of the collection building and organization of the Islamic Sciences and Culture Academy library in Qom, Iran. Knowledge organization and scientometrics are among his research interests.



Ahmad Shabani is a Professor and researcher with more than thirty years of experience in Knowledge and Information Science. He is a Professor at the the University of Isfahan's Department of Knowledge and Information Science. His main research topics include Information Management, Knowledge Management, Research Methods, and Theoretical Issues of Information Science.

Asefeh Asemi is PhD in Business Informatics. She is a faculty member and researcher at Corvinus University, Budapest. Her fundamental research interests are Information Systems, Intelligent Business, Investment Systems, Crypto Currency, Recommender Systems of investment, Business Informatics, Big Data, and Data Mining.

Mehrdad Cheshmeh Sohrabi holds a Ph.D. in Information and Communications. He is an Associate Professor in the Department of Knowledge and Information Science at the University of Isfahan. His research topics are Semantic Web, Semantic Retrieval, Ontology, Data Mining, Quantitative Science Studies, and Research Ethics.

Mohammad Tavakolizadeh Ravari is PhD in Scientometrics. He is currently an Associate Professor in the Department of Knowledge and Information Science at Yazd University. His research interests are scientometrics, technologies, automatic indexing, computer programming in libraries, and information science.

Saadat, Rasul, Ahmad Shabani, Asefeh Asemi, Mehrdad Cheshmeh Sohrabi, and Mohammad Tavakolizadeh Ravari. 2024. "Identifying the Patterns of Author-Generated Tags to Library and Information Science Papers in the Academic Social Networks: Focusing on Academia.edu". *Knowledge Organization* 51 (1): 26-37. 29 references. DOI:10.5771/0943-7444-2024-1-26.

Abstract: This research aims to identify some patterns of author (as user) generated tags to the papers of library and information science field in Academia.edu. The research method is typically based on text analysis and word frequency distribution. The population contains over 6000 papers tagged in Academia.edu, and their abstracts were extracted from 159 English journals of the library and information science (LIS) field in the Scopus database. The growth of different types of tags in terms of the number of their words (one-word, two-word, three-word, and four-word and more), as well as the total number of tags over time, appeared as a logistic curve. It was also found that two-word tags had the most matching (54.92%) and four-word tags or more the least matching (1.76%) with different sections of papers (title, abstract, and authors' keywords). The total tags matched 7.5% with the title, 76.61% with the abstract, and 15.89% with the authors' keywords. Regarding the reuse of tags, it was revealed that on the one hand, 38.8% of the tags had been reused; on the other hand, 16% of the tags were reused in the first year, and more than 50% of the tags were reused in the first three years. Finally, it can be said that the users' consensus on specific terms can identify the new patterns of users' tagging at least partially compatible with professional indexing concepts, and by focusing on the most widely used tags and their sustainable distribution, the weighting of indexing terms and even classification schemes may be achieved.

Received: 27 May 2022; Revised 07 December 2023; Accepted 15 December 2023

Keywords: author-generated tags; folksonomy; social tagging, Academia.edu, subject indexing

1.0 Introduction

Advances in information technology, especially the Internet and Web 2.0, have led to new approaches to organizing and retrieving information. Among these approaches, ontologies and folksonomies have emerged to help improve and complement the previous knowledge organization systems (Alexiev et al. 2005).

Folksonomies, as a new indexing approach in social networking platforms, have different characteristics than traditional indexing methods, and the interaction and collective participation of the users and the direct involvement of their mind and language in organizing information and documents is particular to this type of approach.

In fact, the problems of using professional organization of information methods, as well as the subject and structural diversity of information distributed on the Web, have led to the folksonomies as a new approach to organizing and controlling Internet content. The philosophical basis of folksonomies is the use of web users' collective wisdom in collecting, organizing, and disseminating information. During this process, the user categorizes the information he obtains on the web in various forms, using keywords called tags, and then enters their information into specialized directories using social software in order to manage and share them with other users (Golder and Huberman 2006).

Thomas Vander Wal (2005) has divided folksonomies into two types: broad and narrow. In the broad folksonomies, multiple users are able to assign their tags to a resource (e.g., Flickr). In fact, in this type of folksonomies, a user can share his chosen resource with others so that they can assign their desired tags to that resource as well. Narrow ones are used when the users (e.g., authors of papers) who posted a resource are able to assign tags to that resource only once (e.g., Academia.edu, Bibsonomy). In fact, in this type of social network, a user's chosen source is not available to others for tagging purposes.

According to Rafferty (2018), the production of metadata in the context of folksonomies is done by users, and this type of knowledge organization system is an unstructured method for indexing documents. While professional indexers are most responsible for document indexing, folksonomies allow content producers or users to take responsibility for indexing themselves.

Today, due to the great importance of folksonomies (social indexes) in organizing and retrieving information, numerous studies have been conducted on various aspects, including knowledge organization and, particularly, professional indexing. According to Chen and Ke (2013), researchers in the field of information science try to understand the priorities and patterns of the users. In this way, they can create a bridge between the tags that the users of social networks and professional indexing freely produce.

Therefore, it can be said that indexing documents based on users' natural language is one of the prominent features of social networks; that is, the same language used by users when searching for information. In fact, the formation of the folksonomies based on the mind and language of the user is their strong feature. At the same time, professional indexing, discussed today in the field of library and information science, is done with less understanding of users' information needs and what is going on in their minds. As the findings of researches demonstrate, index terms generated by professional indexers have little overlap with the social tags generated by the folksonomies' users, and there is a very different understanding of documents between these two groups (Lee and Schleyer 2010, 2012; Kipp 2006, 2007b, 2011; Rorissa 2010; Lyer and Bungo 2011; Qanavati et al. 2018; and Movahedian et al. 2020).

Accordingly, the question that researchers have focused on is whether it is possible to improve the organization of information, particularly professional indexing, by relying on the strengths of folksonomies and social tags and, in fact, by patterning these systems, in such a way that more atten-

tion is paid to the users' mind and language when indexing documents? It should be noted that assigning tags to academic papers by their authors is one of the new methods that has become possible due to the widespread of information technology. This process, called social tagging, results from a collaborative effort of subject experts and authors of papers as users of academic social networks.

Issues in the field of professional subject indexing literature can be generalized to the field of social tagging because these are related to the issue of determining the subject content of documents. Topics in the field of indexing include defining single-word or simple terms and multi-word or compound (ISO 25964-1 2011), the importance of different parts of a document to identify important words and content (Lancaster 2003), and time series analysis of the presence and departure of subject terms from a field. Addressing these issues can reveal hidden facts in identifying or attributing subjects to documents (manually or automatically). The primary motivation for this research is the facts and patterns that exist in the background of assigning social tags to academic papers.

Academia.edu is one of the platforms that meet the requirements for studying social tags. In this site, any person in any scientific discipline, including library and information science, can create an account, upload his/her work(s) and increase the visibility to a great extent by assigning desired tags (research interests) to those works (Academia.edu 2023).

For the present study, we selected the field of library and information science, which is also an interdisciplinary field of interest to the researchers. Hence, selecting this field can complement previous research and open new research horizons from the perspective of social tagging toward knowledge organization, particularly professional indexing, to those interested in this field.

Accordingly, the present study seeks to answer the following questions: 1) what is the growth pattern of different groups of tags assigned to papers related to library and information science, in terms of the number of constituent words, over time in Academia.edu?; 2) in terms of compliance, what is the relationship between the number of words that make up the tags assigned by users and the different sections of papers (titles, abstracts, authors' keywords) related to library and information science in Academia.edu?; 3) what is the time pattern between the first time a tag is assigned to papers related to the library and information science until it is reused in different time units in Academia.edu?

2.0 Related works

In previous studies, we could identify two types of research that have been conducted so far: a) a type of research that has compared social tags with words in information resources (including web pages and papers) in terms of their compliance (Heckner et al. 2008; Haustein and Peters 2012; Qanavati et al. 2018; Vaidya and Harinarayana 2019) and; b) a type of research based on the characteristics of social tags to discover the growth patterns, and reuse of the tags (Farooq et al. 2007; Yin et al 2011; Ma 2012; Santos-Neto et al. 2014; Choi and Syn 2016; Xu et al.2018)

Heckner et al. (2008) by conducting a research on computer technology papers on the Connotea, found that 49% of the tags were selected from the title keywords, and 9% from the abstracts of the papers, They also showed that only 30% of the tags overlapped with the keywords assigned to the documents by the authors, and users were more inclined to assign more general and simpler terms to the documents than the authors. Haustein and Peters (2012), by analyzing title keywords, abstracts, subject headings, and index terms of 45 journals of physics and related social tags at CiteUlike, Connotea, and Bibsonomy, revealed that the most consistency was related to abstract and title. In fact, 77.6% of social tags matched the abstract and 66% matched the title words. Also, 29.3% of the tags matched the authors' keywords. Qanavati et al. (2018), with the aim of identifying the degree of conformity of the language of indexers, authors, and taggers in Eric and Mendeley, found that the degree of compatibility of keywords assigned by authors with tags assigned to the same documents was 15%. The compatibility of the descriptors assigned by the indexers to the documents in the Eric with the tags assigned by the taggers to the same documents on the Mendeley was 3%, and finally, the overlap of all three languages studied was 1.1%. Vaidya and Harinarayana (2019), by comparing the tags assigned to marine science papers in CiteUlike with the keywords of authors and titles of the same papers, showed that the overlap between social tags with the authors' keywords was 44.47% and with the title was 36.29 %.

Evaluation of user tagging behavior on CiteUlike by Farooq et al. (2007) and the calculation of the reuse rate of tags showed that each tag's average occurrence of reuse was 3.9%. By examining the temporal aspects of users' interests as well as the phenomenon of changing subject trends in three tagging systems (del.icio.us, Bibsonomy, and Flickr), Yin et al. (2011) clarified that users were 13.9% more likely to use new tags. Such a point indicates that in 86.1% of cases, users chose based on the previously used tags. Ma (2012) studied the growth patterns of tags in CiteULike with the aim of investigating the stability of tag distribution. The findings of the study in the form of time series and analysis of specific trends indicated the continuous growth of the tagging system. Santos-Neto et al. (2014) studied three social networks, including CiteULike, Connotea, and Del.icio.us, with the aim of investigating the patterns of information production, dynamic time factors affecting the

words of tagging, and the social aspects of tagging systems. The findings showed that the target audience and the type of content that received the tag played an important role in the tagging behavior of users. Also, the tagging rate was constantly growing, and at the same time, the rate of this growth depends on the user's history in the tagging system. In fact, the average growth rate of tags by older users was higher, reaching 10%. In order to find out which features of documents users pay the most attention to, Choi and Syn (2016) examined NINES' social tags in the field of history relative to textual, non-textual, and bibliographic sources. The research findings revealed that out of 1540 unique tags, nearly 76% were single-word and 24% were multi-word. Also, among the unique tags, 53.44% were used only once, 13.12% of the tags were used twice and 6.49% were used three times. Of these, only 52% were used more than 100 times. Xu et al. (2018), by examining the tagging behavior of active and inactive users in CiteULike based on three types of growth models, namely the damped exponential model, normal model, and fluctuating model, showed that the tagging behavior of the most active and inactive users followed the fluctuating growth model.

A comparison of the research results of the first and second types shows that the first focuses more on the amount of tags taken from different parts of information sources, regardless of the number of words making up the tags (simple and compound). In contrast, the second emphasizes discovering rules about users' activity, stability in the patterns of tagging activities, clustering users' interests, exploiting the similarities of users' interests in tagging activities to improve exploration, and discovering and analyzing the hidden structure and patterns in the tagging systems regardless of the number of tags' words.

Accordingly, conducting a research based on the number of constituent words of the tags (simple and compound tags) is essential once one can identify the users' tagging patterns from various aspects, including the organization of information and hidden patterns in the tagging of such a significant academic social network (i.e., Academia.edu).

3.0 Methodology

This quantitative research is conducted by text analysis at the level of words, including social tags assigned to papers. After collecting data from April to September 2020 from Scopus and Academia.edu, the analysis was performed by Excel 2016 and SPSS 26. Also, to check which part of the papers the tags were exactly repeated, a computer program was prepared by C#, and with its help, the correspondence between the tags with different sections of the papers (title, abstract, and author keywords) was done. The choice of Academia.edu for extracting the papers and their tags was due to its adoption by a wide range of users worldwide, especially in the library and information science for uploading and tagging the papers. Hence, it can be a prominent example of academic social networks for studying. The choice of Scopus for extracting bibliographical information from journal papers was due to its international reputation and the standards observed in this database regarding to the ranking of the journals.

Regarding the questions raised in the introduction, in the first one, the growth rate of tags assigned to papers over time was calculated based on the year of publication of journal papers from 1972 to 2020. The reason for using article publishing dates instead of user tagging dates was to study the exponential growth and saturation of user-assigned tags and terms, which could not be studied in less than 10 years, for example. In the second question, the amount of compliance of the tags with the different sections of the papers (title, abstract, and authors' keywords) was obtained. In the third question, the amount of reuse of tags by users was calculated year by year (i.e., from 1972 to 2020), and tags that were used more than once, along with the amount of their use and the duration of their presence in the collection under study was achieved.

In order to collect data, first, we extracted from Scopus the bibliographical records in the Txt file format of papers in English language journals in the library and information science field, which in April 2020 included 159 English language journals and 194.337 papers. At the time of data collection, the total number of LIS journals on Scopus was 189 titles in different languages. In such cases, according to the census method, 8.340 bibliographic records of papers in the field and their abstracts in English language journals, based on a search in Academia.edu with two criteria, (i.e., "exact title" and "having tag(s)") were extracted. In summary, the steps of data collection and preparation for analysis were as follows:

1. Extracting 194.337 bibliographic records of papers from 159 English language journals in the LIS field, available in Scopus in txt format and based on search in LISTA database;
2. In order to convert the txt files to the appropriate format, it was entered into MiMFa RAVAR (already released in C#) software and the desired output was obtained;
3. The files obtained from the previous step were called Excel 2016 and entered into the MasterCoderRobotArticle (released in C# by the researchers);
4. Searching for papers with the help of an article finder robot (released in C# by the researchers) in the Academia.edu and extracting those papers that had tag(s) (8.340 bibliographical records of papers);
5. Importing bibliographical records of papers retrieved using ScrapeFromResult (released in C# by the researchers)

in Excel 2016 and pre-processing and unifying them and obtaining 6.086 ones;
6. Building of document-tag matrix;
7. Dividing tags assigned to papers' records into one-word, two-word, three-word, and four-word and more, as demonstrated in Table 1.

4.0 Findings

4.1. Findings of Question 1

Regarding the first question of the research, the growth of one-word, two-word, three-word, four-word, and more tags, as well as the total tags assigned to journal papers published between 1972 and 2020, can be observed in Figure 1, in the diagrams a to e, and in Table 2.

N. of papers	One-word tags		Two-word tags		Three-word tags		Four-word tags & more		Total tags	
	By repeat	No repeat	By repeat	No repeat	By repeat	No repeat	By repeat	No repeat	By repeat	No repeat
6086	7119	1354	18396	4046	3594	1375	5087	958	34196	7733

Table 1. Statistical population of the study

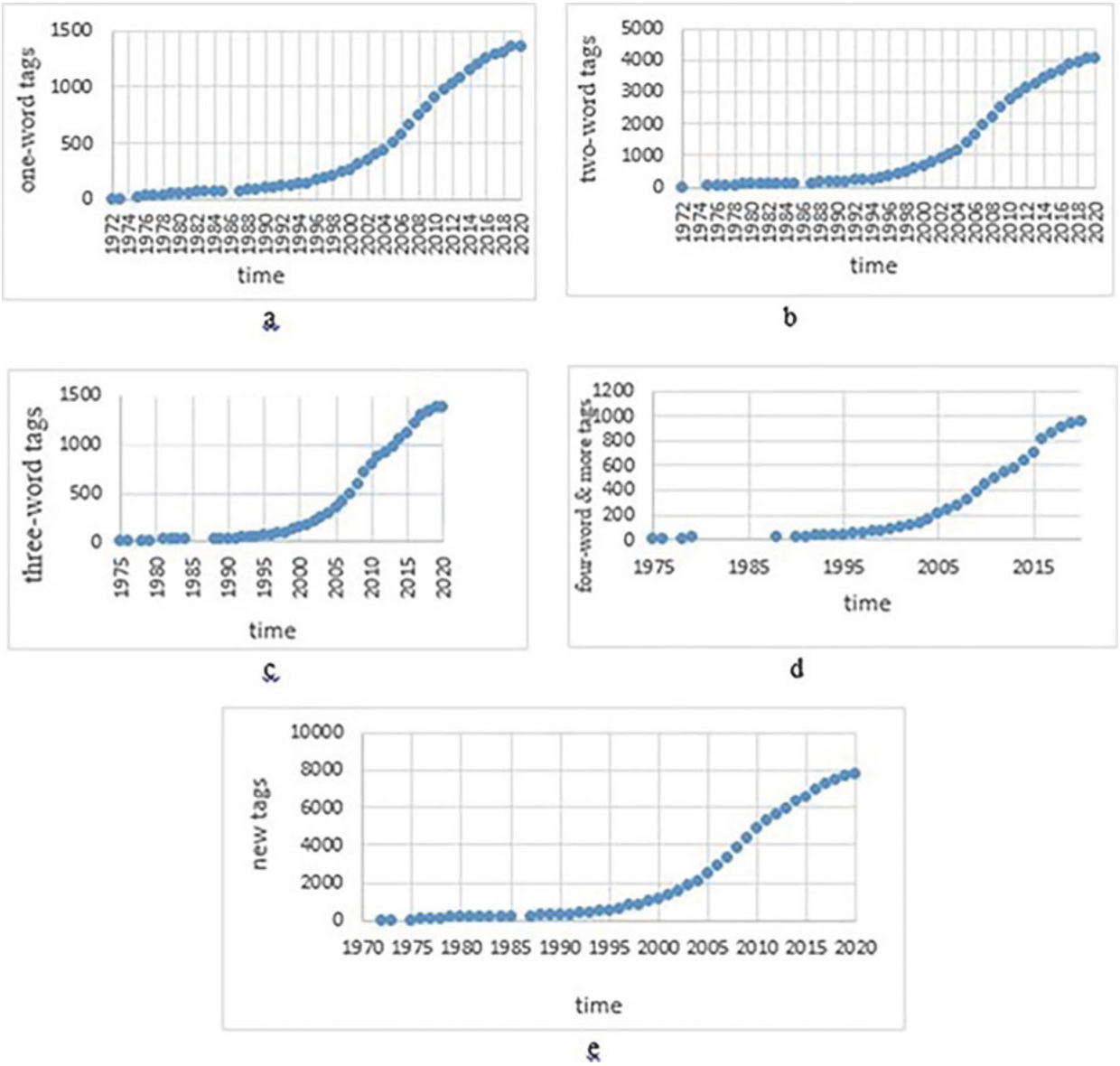


Figure 1. Tag growth during the time

Charts a to e show the cumulative growth of one-word, two-word, three-word, four-word, and more tags, as well as the total number of new tags over time for papers published between 1972 and 2020. Table 2 identifies the number of years of exponential growth and frequency of tags.

According to Table 2, the first year of exponential growth of tags varies between 1994 and 1998, the last year of exponential growth varies between 2014 and 2019, and the number of exponential years of growth varies between 19 and 21 years.

4.2. Findings of Question 2

Regarding the second question of the present study, the existence of one-word, two-word, three-word, four-word and more tags, as well as all new tags in different sections of papers (title, abstract, authors' keywords) can be observed in Table 3 and in diagrams 1 to 2:

Table 3 shows that one-word tags (2.73%), two-word tags (4.13%), three-word tags (48%), and four-word tags and more (16%) correspond to the title. Also, one-word tags (29.44%), two-word tags (41.53%), three-word tags (4.34%), and four-word tags or more (1.3%) are consistent with the abstract. On the other hand, one-word tags (5%), two-word tags (9.26%), three-word tags (1.33%), and four and more tags (3%) correspond to the authors' keywords.

Figure 2 shows the compatibility of different groups of tags with different sections of papers by comparison and percentage. Two-word tags had the most compliance (54.92%), and four-word tags and more had the least compliance (1.76%) with different sections of papers (titles, abstracts, and authors' keywords).

Also, according to Figure 3, the total tags were compliant 7.5% with the title, 76.61% with the abstract, and 15.89% with the authors' keywords.

4.3. Findings of Question 3

About the third question posed in the introduction, the findings about the reuse of tags in total and in terms of different groups of tags, (i.e. one word, two words, three words, and four words and more) based on the date of publication of journal papers (2020-1972) are presented in Tables 4 to 7 and Figure 4.

According to Table 4, out of 7.733 new tags (non-repetitive), 2.997 (38.8%) have been reused and 4.736 (61.2%) have been used only once. It also shows the number of censored tags that have only been used once. Therefore, these tags were removed from the test and the time pattern between the first assigning time and reuse was calculated using the tags that were reused at least once using survival analysis and life table in the first 10 years (Table 5).

Table 5 shows that 16% of the tags were reused in the first year and 24% of the tags not reused in the first year were reused by the end of the second year.

Tag frequency	Number of exponential growth years	The last year of exponential growth	The first year of exponential growth	Tag Group
1.354	19	2014	1995	One-word
4.046	20	2014	1994	Two-word
1.375	20	2016	1996	Three-word
958	21	2019	1998	Four-word &more
7.733	20	2014	1994	Total tags

Table 2. The exponential growth of tags over time and their frequency

Tags' group	Tags' population		Title		Abstract		Authors' keywords		All sections	
	Freq.	Perce.	Freq.	Perce.	Freq.	Perce.	Freq.	Perce.	Freq.	Perce.
One-word	1354	17.51	786	2.73	8.487	29.44	1.440	5	10.713	37.17
Two-word	4.046	52.32	1191	4.13	11.975	41.53	2.670	9.26	15.836	54.92
Three-word	1.375	17.78	140	48	1.253	4.34	382	1.33	1.775	6.15
Four-word &more	958	12.39	45	16	374	1.30	89	3	508	1.76
Total tags	7.733	100	2162	7.5	22.089	76.61	4.581	15.89	28.832	100

Table 3. Adaptation of types of tags with different sections of papers (titles, abstracts, and authors' keywords)

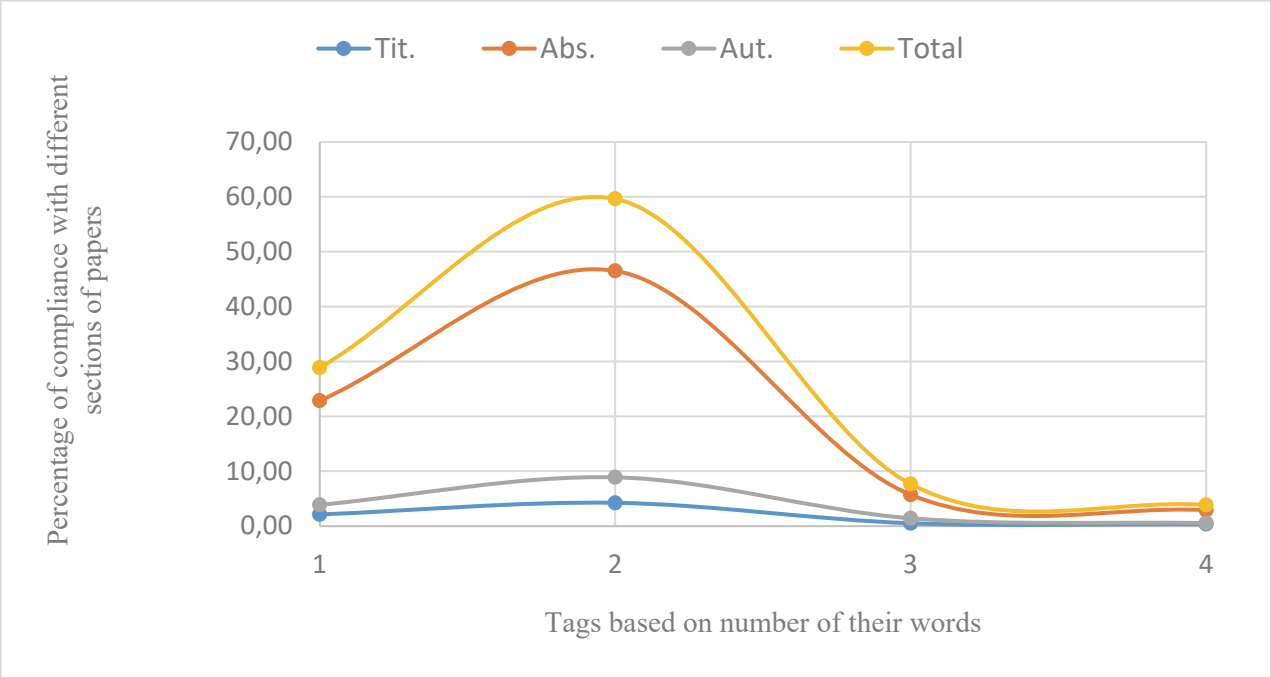


Figure 2. Comparison of the compatibility of different groups of tags with different sections of papers

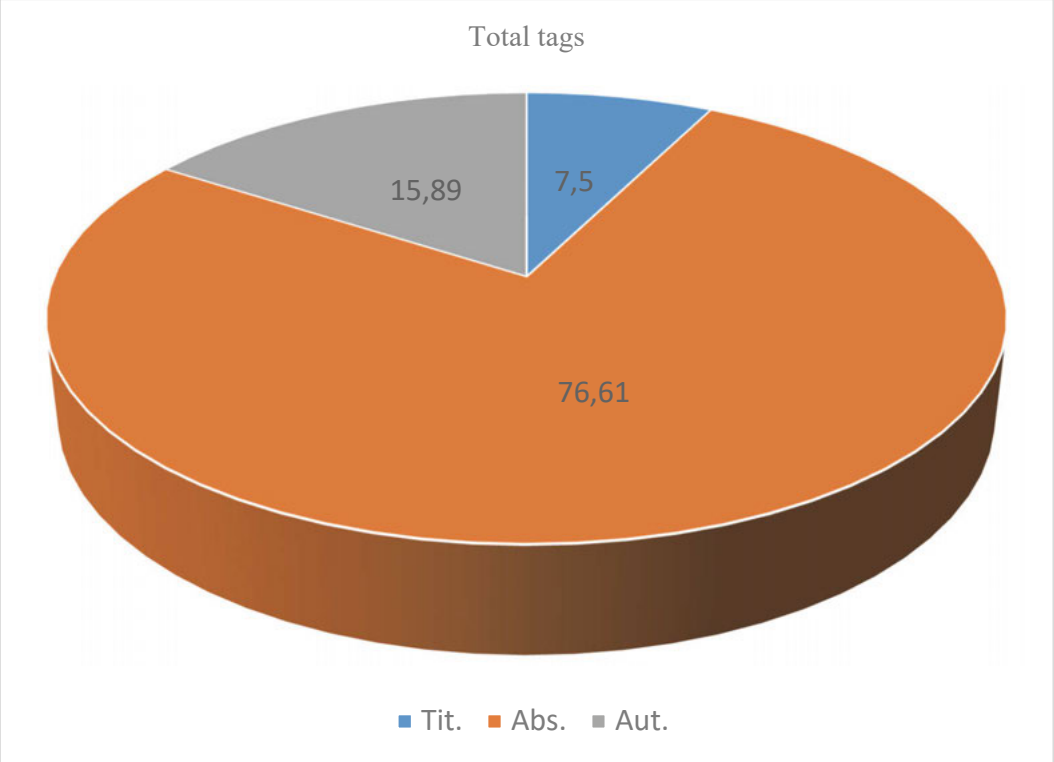


Figure 3. Compliance of all tags with different sections of papers

In Figure 4, the index line indicates that half of the tags have been reused up to 49.47 months (i.e. about 4 years after the tag was created).

In the next step, the reuse of tags in terms of different groups (one-word, two-word, three-word, and four-word and more ones) was verified using Kaplan Meyer's estimation, presented in Table 6.

According to Table 6, 57.59% of two-word tags were the most reused, and four-word tags and more, with 7.54%, were the least reused.

Tag reuse	Freq.	Perce.
Reused tags	2.997	38.8
Censored tags	4.736	61.2
Total tags	7.733	100

Table 4. Reuse of tags in total

Time series	Frequency of tags remaining until the end of each time interval	Frequency of reused tags	Possibility of reusing the remaining tags at any time interval	Standard deviation
1	2997	474	.16	.001
2	2523	595	.24	.001
3	1928	390	.20	.001
4	1538	322	.21	.000
5	1216	235	.19	.000
6	981	185	.19	.000
7	796	147	.18	.000
8	649	103	.16	.000
9	546	118	.22	.000
10	428	65	.15	.000

Table 5. Life table of the time pattern between the first allocation time and the probability of reuse of tags

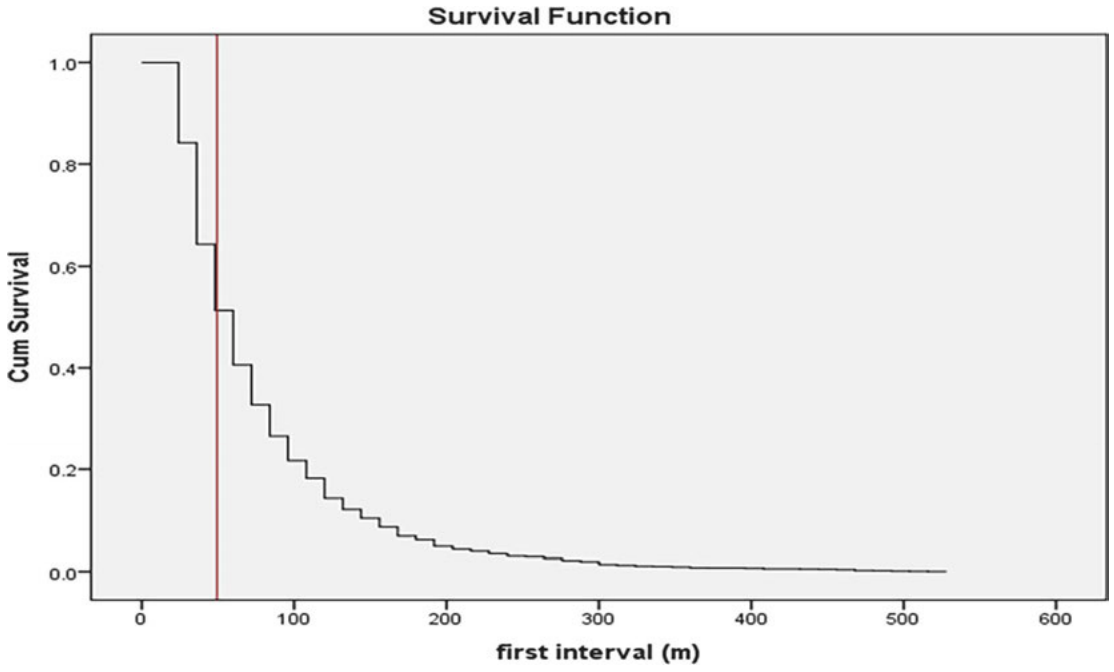


Figure 4. Probability of reusing tags over a period of time

Table 7 lists the 20 most used tags among the other tags, which is another finding of the present research.

5.0 Discussion

After presenting the findings in the previous section, we now return to the three questions posed in the introduction, seeking to answer them accordingly.

The first question was: What is the growth pattern of different groups of tags assigned to papers related to the library and information science field, in terms of the number of constituent words, over time in Academia.edu? Here, it was found that the growth of new one-word, two words, three words, and four words and more tags is first exponentially (19, 20, 20, and 21 years old) and then logarithmically in the form of logistic curves (Figure 1). In this type of curve, the growth is first slowly and then exponentially and rapidly, after which exponential growth is saturated and enters logarithmic growth. Logarithmic growth is rapid at first and then continues in a horizontal and straight line.

Also, based on the findings of the present study, in total (regardless of the number of words), the growth of new tags was first as a 20-year exponential, and then logarithmic growth and this growth showed itself in the form of a logistic curve (diagram e).

These points reveal that in academic social networks such as Academia.edu, the growth of tags, depending on the type of tagging system, reaches a saturation stage after a

Tags' group	Initial frequency	Percentage of Initial frequency	Frequency of tag reuse	Percentage of tag reuse
One-word	1354	17.51	604	20.15
Two-word	4046	52.32	1726	57.59
Three-word	1375	17.78	441	14.72
Four-word &more	958	12.39	226	7.54
Total	7733	100	2997	100

Table 6. Reusing of tags in terms of the number of their words

No	Tag	Frequency of use	First year of use	Last year of use
1	LIBRARY AND INFORMATION STUDY	2.705	1975	2019
2	INFORMATION SYSTEM	977	1975	2020
3	SCIENTOMETRIC	467	1980	2020
4	INFORMATION RETRIEVAL	332	1980	2020
5	INFORMATION MANAGEMENT	279	1991	2019
6	DESIGN METHODOLOGY	242	1981	2015
7	LIBRARY AND INFORMATION SCIENCE	240	1978	2019
8	HUMAN	224	1981	2017
9	INFORMATION TECHNOLOGY	216	1979	2017
10	CASE STUDY	207	1988	2015
11	DISTRIBUTED COMPUTING	190	1991	2015
12	LIBRARY SCIENCE	190	1972	2020
13	CLINICAL SCIENCE	190	1982	2016
14	DIGITAL LIBRARY	178	1976	2020
15	INFORMATION LITERACY	178	1997	2019
16	NURSING	172	1995	2016
17	INFORMATION SCIENCE	170	1981	2020
18	ACADEMIC LIBRARY	164	1991	2019
19	KNOWLEDGE MANAGEMENT	159	1995	2019
20	INDEXATION	143	1987	2017

Table 7. 20 most used tags among Academia.edu users in the field of Library & Information Science

while, and then the tags grow at an almost constant rate. The exponential growth of tags showed that in the LIS field, the growth of subjects, which is manifested in the form of tags assigned to papers, is saturated after about 20 years. In fact, in that period, new issues experienced exponential growth, and then it reached a stage of saturation in which new issues were rarely raised and, according to Price (1965), reached a critical point or senile limit.

In this regard, although Price (1965) has estimated the exponential growth of the subjects to be about 30 to 45 years, it is a combination of exponential growth and a logarithm obtained in the form of a logistic curve, which is in line with the findings of the present study. Ma's research findings (2012) about CiteULike showed that in a period of 5 years (2005-2009), new tags grew 12.4 times and had exponential growth. He emphasized that the growth of non-duplicate tags (i.e., the creation of new tags over time) is an indicator of the stability of collaborative tagging systems, and his research findings in the form of time series and analysis of specific trends indicate the continuous growth of the tagging system. Ma's study (2012) time period was not more than 5 years. Therefore, the growth of tags was not saturated, and the author talked about the continuous growth of the tagging system. Also, Santos-Neto et al. (2014) confirmed the continuous growth of tags in the social networks CiteULike, Connotea, Del.icio.us over a period of 6 years and considered this growth rate to be conditional on the user's history in the tagging system. Yunhong et al. (2018), who surveyed user behavior and tag growth at CiteULike over a 4-year period, also acknowledged that most users' tagging behavior followed a fluctuating growth model. Farooq et al. (2007) evaluated the tagging behavior of users on the CiteUlike over a period of 3 years and found that the cumulative frequency of new tags followed a linear relationship over time, and such a finding indicates that new tags are always increasing over time, which is not consistent with the findings of the present study; the probable reason is the relatively short period of time that has been studied.

The second question was: In terms of compliance, what is the relationship between the number of words that make up the tags assigned by users and the different sections of papers (titles, abstracts, authors' keywords) related to the field of library and information science in Academia.edu? The findings of the second question of the research once again showed the tendency of the authors of the papers to use two-word terms; as two-word tags had the most compliance (54.92%) and four-word tags and more had the least compliance (1.76%) with different sections of papers (title, abstract, and authors' keywords). Also, the total tags were 76.61% consistent with the abstract (Table 3), which indicates the prominent situation of papers' abstracts in obtaining index terms during indexing. Of course, this situation can be due to the large volume of the abstract and the greater number of important keywords of the papers in their abstract.

At the same time, the findings of the present study showed that the rate of tag compliance with titles was the lowest (7.5%) among previous studies, as Heymann et al. (2008) compared social tags of del.icio.us (without data preprocessing) which showed that 16% of the tags appeared in web page titles, which can be attributed to differences in the type of social network and the purposes of user tagging, as well as in the lack of data preprocessing. Heckner et al. (2008) conducted a study in the field of computer technology and examined the social network tags of Connotea (without data pre-processing) and found that 49% of the tags were selected from the title keywords. Also, Haustein, and Peters (2012) by analyzing related tags in the field of physics in CiteUlike, Connotea and Bibsonomy (after data pre-processing), showed that 66% of the tags corresponded to the title words and the reason for the discrepancy with the findings of the present study can be considered due to the area under study and the social network type. Vaidya and Harinarayana (2019) compared the tags assigned to marine science papers in CiteUlike (after preprocessing) and revealed that the rate of overlap between user tags and article title words is 36.29%, which may be the reason for the difference in the findings of the present research in the field of study. In fact, science and engineering are different from the humanities and social sciences in this respect. Hjorland (2015) pointed out that one of the issues about searching in different fields of databases is that the importance of titles of works for search varies according to how titles are used in different scientific fields. In social sciences, for example, metaphors may diminish the importance of searches in the title field.

The findings of the present study are somewhat consistent with the research of Haustein and Peters (2012), who stated that the conformity of tags with the abstracts of physics papers was 77.66%, but inconsistent with the findings of Heckner et (2008) who stated that this conformity in the field of computer technology was 9%.

Also, the findings of the present study indicated that 15.89% of the tags were in accordance with the authors' keywords, which is partly in line with the findings of Qanavati et al. (2018), which (after data preprocessing) stated that 15% of tags are consistent with authors' keywords.

Regarding question 3: what is the time pattern between the first one a tag is assigned to papers related to the library and information science field until it is reused in different time units in Academia.edu?; we concluded that, in general, 38.8% of the tags have been reused. Farooq et al. (2007) calculated the reuse rate of tags on the CiteUlike (during two years and four months) 3.9%, which is not consistent with the present study, and the study duration can affect the

amount of reuse of tags. In addition, Yin et al. (2011) verified three tagging systems (del.icio.us, Bibsonomy and Flickr) and found that users reuse previous tags in 86.1% of cases. In this case, tagging systems with different purposes and the sample under study can be the probable reason for the difference in the present study's findings. Also, by examining the tags of NINES's users in the field of history (8.310 tags, 1.540 non-duplicate tags), Choi and Syn (2016) found that 46.56% of the tags have been used more than once, which is not consistent with the findings of the present study.

Although in previous studies the reuse of tags was not examined separately in terms of the number of words, in the present study the prominent role of two-word tags in social indexing was revealed by the authors in library and information science.

Also, with the help of survival analysis and life table (Table 5), it was revealed that, first, it can be determined survival of subjects by their emergence and decline in the field under study; second, in library and information science 16% of the topics that have been reused in the first year (hot topics), can be expected to be extended to the following year. Thus, the trajectory of using topics in the future can be determined, as well as the current direction of research in the field.

In the final part of the third question, Table 7 shows different tags used along with their frequency and duration of presence. According to the frequency of use and the duration of each tag in the study set, a formula for determining the weight of tags for data retrieval in the field can be obtained and used in database retrieval algorithms. Also, according to the table, it can be expected that the topics that are the basis of a field are constantly present in the literature of that field.

6.0 Conclusion

Tags often contain the views and mental context of the taggers. Therefore, social tags can be considered more prosperous than the index terms generated by indexers. Also, significant user consensus on certain words or phrases indicates that the new patterns of user tags are at least partially compatible with the concepts of professional indexing of document content. The other point is that by focusing on the most commonly used tags and their consistent distribution, we can formulate a weight and even design the classification schemes. In addition, authors' (as users) activities on academic social networks can be used to increase the quality of suggestions in collective tagging systems. Another point is that there is a connection between professional indexing and user tagging, and the two are not alien to each other.

Finally, we propose some research suggestions that can enrich the studies about folksonomy and academic social networks: a) examining the growth patterns of user tags to papers in the fields of humanities and social sciences, sciences and engineering over time in academic social networks; b) adaptation of user tags to different sections of papers (title, abstract, author's keywords, and text) in the fields of humanities and social sciences, sciences, and engineering in academic, social networks; c) examining the time series of reuse of tags by users of other scientific fields in academic, social networks and; d) studying the factors affecting the reuse of tags and the effectiveness of these factors in other scientific fields in academic, social networks.

References

Academia.edu. 2023. "About Academia.edu." www.academia.edu/about

Alexiev, Vladimir, Michael Breu, Jos de Bruijn, Dieter Fensel, Ruben Lara, and Holger Lausen (Eds.). 2005. *Information Integration with Ontologies: Experiences from an Industrial Showcase.* Chichester, England: Wiley & Sons.

Chen, Ya-Ning and Hao-Ren Ke. 2013. "An Analysis of Users' Behavior Patterns in the Organization of Information: A Study of CiteULike." *Online Information Review* 37, no. 4: 638-56. DOI: 10.1108/OIR-05-2012-0086/full/html

Choi, Youngok and Sue Y. Syn. 2016. "Characteristics of Tagging Behavior in Digitized Humanities Online Collections." *Journal of the Association for Information Science and Technology* 67, no. 5: 1089–1104.

International Organization for Standardization (ISO). 2011. *Information and Documentation. Thesauri and Interoperability with Other Vocabularies. Part 1: Thesauri for Information Retrieval (*ISO Standard n. 25964-1: 2011).

Farooq, Umer, Thomas G. Kannampallil, Yang Song, Craig H. Ganoe, John M. Carroll, and C. Lee Giles. 2007. "Evaluating Tagging Behavior in Social Bookmarking Systems: Metrics and Design Heuristics." In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*. New York: ACM Press: 351-60.

Golder, Scott A., and Bernardo A. Huberman. 2006. "Usage Patterns of Collaborative Tagging Systems." *Journal of Information Science* 32, no. 2: 198-208. DOI: 10.1177/0165551506062337

Haustein, Stefanie, and Isabella Peters. 2012. "Using Social Bookmarks and Tags as Alternative Indicators of Journal Content Description." *First Monday 17*, no. 11. http://firstmonday.org/ojs/index.php/fm/article/view/4110/3357

Heckner, Markus, Susanne Mühlbacher, and Christian Wolff. 2008. "Tagging Tagging. Analyzing User Keywords in Scientific Bibliography Management Systems".

Journal of Digital Information 9, no. 2: 1–19. https://jodi-ojs-tdl.tdl.org/jodi/article/view/246

Heymann, Paul, Georgia Koutrika, and Hector Garcia-Molina. 2008. "Can Social Bookmarking Improve Web Search?" In *Proceedings of the International Conference on Web Search and Web Data Mining – WSDM '08*. DOI: 10.1145/1341531.1341558

Hjørland, Birger. 2015. "Classical Databases and Knowledge Organization: A Case for Boolean Retrieval and Human Decision-Making During Searches." *Journal of the Association for Information Science and Technology* 66, no. 8: 1559–75. DOI: 10.1002/asi.23250

Kipp, E. Margaret. 2006. "Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator and Intermediary Keywords." In *Proceedings of the Annual Conference of CAIS*. DOI: 10.29173/cais174

Kipp, E. Margaret. 2007. "Tagging Practices on Research Oriented Social Bookmarking Sites." In *Proceedings of the Annual Conference of CAIS.* DOI: 10.29173/cais223

Kipp, E. Margaret. 2011. "Controlled Vocabularies and Tags: An Analysis of Research Methods." In *Proceedings from North American Symposium on Knowledge Organization* – NASKO 3, no. 1: 23-32. DOI: 10.7152/nasko.v3i1.12787

Lancaster, Frederick W. 2003. *Indexing and abstracting in theory and practice*. 3rd edition. Champaign: University of Illinois Graduate School.

Lee, Danielle. H. and Titus Schleyer. 2010. "A Comparison of Mesh Terms And Citeulike Social Tags as Metadata for the Same Items." In *Proceedings of the First ACM International Conference on Health Informatics, 11-12 November 2010, Arlington Virginia USA,* edited by Tiffany Veinot. New York: Association for Computing Machinery, 445–48. DOI: 10.1145/1882992.1883060

Lee, Danielle. H. and Titus Schleyer. 2012. "Social Tagging is no Substitute for Controlled Indexing: A Comparison of Medical Subject Headings and CiteULike Tags Assigned to 231,388 Papers." *Journal of the American Society for Information Science and Technology* 63, no. 9: 1747- 57. DOI: 10.1002/asi.22653

Lyer, Hemalata and Lucy Bungo. 2011. "An Examination of Semantic Relationships Between Professionally Assigned Metadata and User-Generated Tags for Popular Literature in Complementary and Alternative Medicine." *Information Research* 16, no. 3, paper 482. http://InformationR.net/ir/16-3/paper482.html

Ma, Jin. 2012. "The Sustainability and Stabilization of Tag Vocabulary in Citeulike: an Empirical Study of Collaborative Tagging." *Online Information Review* 36, no. 5: 655–74. DOI: 10.1108/14684521211275966

Movahedian, Qasem, Ahmad Shabani, A, Mehrdad Cheshmesohrabi and Asefeh Asemi. 2020. "Otherness in Library Organization Systems vs. Social Tagging." *Librarianship and Information Organization Studies* 31, no 3: 58-71. http://46.209.25.211/article_2419_c0998356dd9fafae1b7a84f83a02f9ba.pdf [Persian].

Price, Derek de Solla. 1965. *Little Science Big Science*. London: Columbia University Press.

Qanavati, Maryam, Alireza Noruzi, Maryam Nakhoda, and Ashkan Khatir. 2018. "A Study of Language Compatibility of Indexers, Writers, and Taggers in Eric and Mendelian Databases." *Journal of Information Processing and Management* 33, no. 4: 1715-36. https://jipm.irandoc.ac.ir/article_699494.html [Persian]

Rafferty, Pauline. 2018. "Tagging". *Knowledge Organization* 45, no. 6: 500-16. Doi: 10.5771/0943-7444-2018-6-500

Rorissa, Abebe. 2010. "A Comparative Study of Flickr Tags and Index Terms in a General Image Collection." *Journal of the American Society for Information Science and Technology* 61, no. 11: 2230–42. DOI: 10.1002/asi.21401

Santos-Neto, Elizeu, David Condon, Nazareno Andrade, Adriana Iamnitchi, and Matei. Ripeanu. 2014. "Reuse, Temporal Dynamics, Interest Sharing, And Collaboration in Social Tagging Systems". *First Monday* 19, no. 7. https://firstmonday.org/ojs/index.php/fm/article/view/4994/410

Vaidya, Praveenkumar and N.S. Harinarayana. 2019. "Comparison of User-generated Tags with Subject Descriptors, Author Keywords, and Title Terms of Scholarly Journal Articles: A Case Study of Marine Science." *Journal of Information Science Theory and Practice* 7, no. 1: 29-38. DOI: 10.1633/JISTaP.2019.7.1.3.

Vander Wal, Thomas. 2005. "Explaining and Showing Broad and Narrow Folksonomies." *Vanderwal.net* (blog). February 21, 2005. www.vanderwal.net/random/entrysel.php?blog=1635

Xu, Yunhong, Dehu Yin, and Duanning Zhou. 2018. "Investigating Users' Tagging Behavior in Online Academic Community Based on Growth Model: Difference between Active and Inactive Users." *Information Systems Frontiers* 21, no. 4. DOI: 10.1007/s10796-018-9891-2

Yin, Dawei, Liangjie Hong, Zhenzhen Xue, and Brian D. Davison. 2011. "Temporal Dynamics of User Interests in Tagging Systems." In *Proceedings of The Twenty-Fifth AAAI Conference on Artificial Intelligence* 7 –11 August 2011, San Francisco, California USA. http://www.cse.lehigh.edu/~brian/pubs/2011/AAAI/temporal-dynamics.pdf