JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 10(3): 40-56, 2022

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2022.10.3.4

An Exploratory Analysis of Online Discussion of Library and Information Science Professionals in India using Text Mining

Mohit Garg*
Central Library, Indian Institute of Technology Delhi, New Delhi, India
School of Social Science, Indira Gandhi National Open University, New Delhi, India
E-mail: gargmohit@library.iitd.ac.in

Uma Kanjilal
School of Social Science, Indira Gandhi National Open University, New Delhi, India
E-mail: ukanjlal@ignou.ac.in

ABSTRACT

This paper aims to implement a topic modeling technique for extracting the topics of online discussions among library professionals in India. Topic modeling is the established text mining technique popularly used for modeling text data from Twitter, Facebook, Yelp, and other social media platforms. The present study modeled the online discussions of Library and Information Science (LIS) professionals posted on Lis Links. The text data of these posts was extracted using a program written in R using the package "rvest." The data was pre-processed to remove blank posts, posts having text in non-English fonts, punctuation, URLs, emails, etc. Topic modeling with the Latent Dirichlet Allocation algorithm was applied to the pre-processed corpus to identify each topic associated with the posts. The frequency analysis of the occurrence of words in the text corpus was calculated. The results found that the most frequent words included: library, information, university, librarian, book, professional, science, research, paper, question, answer, and management. This shows that the LIS professionals actively discussed exams, research, and library operations on the forum of Lis Links. The study categorized the online discussions on Lis Links into ten topics, i.e. "LIS Recruitment," "LIS Issues," "Other Discussion," "LIS Education," "LIS Research," "LIS Exams," "General Information related to Library," "LIS Admission," "Library and Professional Activities," and "Information Communication Technology (ICT)." It was found that the majority of the posts belonged to "LIS Exam," followed by "Other Discussions" and "General Information related to the Library."

Keywords: discussion forum, Lis Links, text mining, topic modelling, Latent Dirichlet Allocation

Received: April 1, 2022 **Revised:** May 18, 2022
Accepted: July 10, 2022 **Published:** September 30, 2022

***Corresponding Author:** Mohit Garg
https://orcid.org/0000-0001-5787-7143
E-mail: gargmohit@library.iitd.ac.in

1. INTRODUCTION

The Internet and Web 2.0 technologies have facilitated the academic community, including library professionals, with many platforms like mailing lists and discussion forums to disseminate or seek information (Muñoz-Cañavate et al., 2017a, 2017b; Shukla & Dawngliana, 2018). Common mailing list, discussion forums in Library and Information Science (LIS) include Lis-Forum[1], IFLA-L[2], and Lis Links[3]. Despite the growth in many social media platforms, these platforms are popular among students and working/retired professionals to seek professional knowledge on different topics, disseminate professional information, or announce conference/workshops/seminar/job advertisements (Muñoz-Cañavate et al., 2017a; Sawant & Sawant, 2016). The professionals articulate their information needs on these platforms and are more active in discussion forums than on blogs for the information needed (Muñoz-Cañavate et al., 2017b). In different fields, online discussion forums have been used commonly for seeking or sharing information (Arden et al., 2014). The asynchronous nature of these forums makes them more popular as they can be accessed anytime from anywhere at the convenience of the community's user (Coulson, 2005).

The plethora of intellectual interactions available on online platforms makes them a knowledge repository of solutions to the various problems faced by professionals. This vast amount of valuable information in the form of online conversations is the ideal data source for deriving valuable insights into the information needs of professionals (Abdillah & Adriani, 2015). In recent times, the analysis of these online discussions has become an emerging area of research interest (Singh et al., 2021) and has attracted researchers to evaluate the needs of professionals. The study of these online conversations can unveil hidden insights about the information needed by the user. There are multiple advantages of considering online discussions as data sources for behavioral information studies compared to the traditional questionnaire, survey, and interview. Firstly, users explicitly expresses their issues whenever they encounter them. The discussion forum stores the problem posted and the respective solution or information posted by the different users of the community. This makes the online forums warehouse all types of information needed by community members at other times. Secondly, many people's interactions can be captured and analyzed using different computer-assisted analyses, whereas formal surveys and interviews capture limited interactions only (Pandapotan et al., 2015). Hence, it saves time in data collection and analysis. Lastly, this online data is unbiased in the research context. Among the many advantages, online discussion data has some drawbacks too. The demerit of this data is that forum participants are the people who have Internet access through mobile/computer appliances, thus they cannot be representative of the whole population. The data is most suitable for exploratory analysis and less for hypotheses testing (McKenna & Thomson, 2014).

In this study, we have explored the online discussion of Indian LIS professionals on Lis Links, an online social networking platform for LIS professionals of India. It has more than 29,000 members posting over 10,000 messages in its discussion forum (Barman, n.d.). This study aimed to answer the following two research questions:

RQ1: What topics are discussed by LIS professionals on the online forum of Lis Links?
RQ2: What are the trending topics of discussion among the LIS Professionals of India?

2. RELATED WORK

In recent times, online discussions on different platforms have become a goldmine of data for researchers to understand people's behavior on various parameters deeply. The analysis of this online data is broadly classified into two categories: manual content analysis, and computer-assisted analysis like text mining or social network analysis. Both approaches have been widely used to analyze data from popular social media platforms, like Facebook, Twitter, Youtube, or StackOverflow. However, in the case of online forums, a great deal of research has been conducted using manual content analysis, and very few studies using computer-assisted analysis like text mining or social network analysis. In this review, the studies based on these techniques have been discussed.

Omidvar et al. (2014) proposed a novel method for finding experts in the AskMe forum. The data were extracted using a crawler and processed utilizing a wordnet dictionary and social network analysis. Abdillah and Adriani (2015) developed a recommendation system that used reviews posted on the forum as the backend data source. The author used Indonesia's most famous restau-

[1]http://ncsi.iisc.ernet.in/mailman/listinfo/lis-forum
[2]https://mail.iflalists.org/wws/info/ifla-l
[3]http://www.lislinks.com/

rant review forum data, i.e. Open Rice Indonesia. The system used Latent Dirichlet Allocation (LDA) for extracting the topics from the reviews. The results show that the method significantly recommends the restaurant based on the user's interest. Wang and Tang (2016) investigated online debates on traditional Chinese medicine in one popular Chinese BBS Tianya Forum. The authors used text mining methods to perform stance analysis. The data was extracted using the crawling mechanism. Later, this data was pre-processed before LDA analysis to discover the topics. Kahani et al. (2016) computed LDA analysis on the Eclipse forum to identify prominent discussion topics. MALLET software was used for LDA analysis. The study highlights the key insight about issues in Eclipse's plugin architecture and documentation, and the support for setting up the initial tool. Kim et al. (2017) examined 17,381 forum articles and 627,122 user comments on the online bitcoin forum to predict the price fluctuation of bitcoin. The crawled data was analyzed using topic modeling to extract the topics from the user comments. The study indicated the fluctuation in bitcoin prices with an accuracy rate of over 80%. The study showed the most prominent words of ten topics as a word cloud.

Munezero et al. (2017) employed text analysis techniques on open source software discussion forums to identify sales leads. The results highlighted new insights from the dataset, which were impossible by manual analysis of this vast data. Zarra et al. (2018) visually analyzed the two sets of a corpus of 15,000 and 25,000 discussions on StackOverflow. LDA analysis in R was used for extracting the topic, and D3.js was used for the visualization. Özcan-Tok et al. (2019) used text mining and regression analysis techniques to investigate four years (2013-2017) of comments of farmers and traders related to potatos, onions, lemons, and apples on a discussion forum. They found that these forums contain valuable information. Qian and Gui (2021) investigated 14,933 health-related posts posted by users of senior online communities. They crawled the corpus of data from Yinling and Keai and used text mining methods to conceptualize the health information needed. The study found that the senior users seek four types of remedies, i.e., coping with aging, dietary nutrition, physical exercise, and mental health. However, the main focus of the senior user is physical health issues. Ahn et al. (2021) examined tweets and retweets in the English language related to disasters to explore the information providers on Twitter. The data was collected using Twitter API and the weepy library of Python. The data was cleaned by removing duplicate tweets and standardized by lemmatization. The authors implemented LDA on standardized data using Python Scikit-learn's Library.

Saranya and Geetha (2020) implemented a topic model on clothing reviews of 2,000 users on the YELP website. The pre-processing of the data to clean the data included converting text to lowercase, removing punctuation, stopwords, and white spaces, and stemming. They used LDA

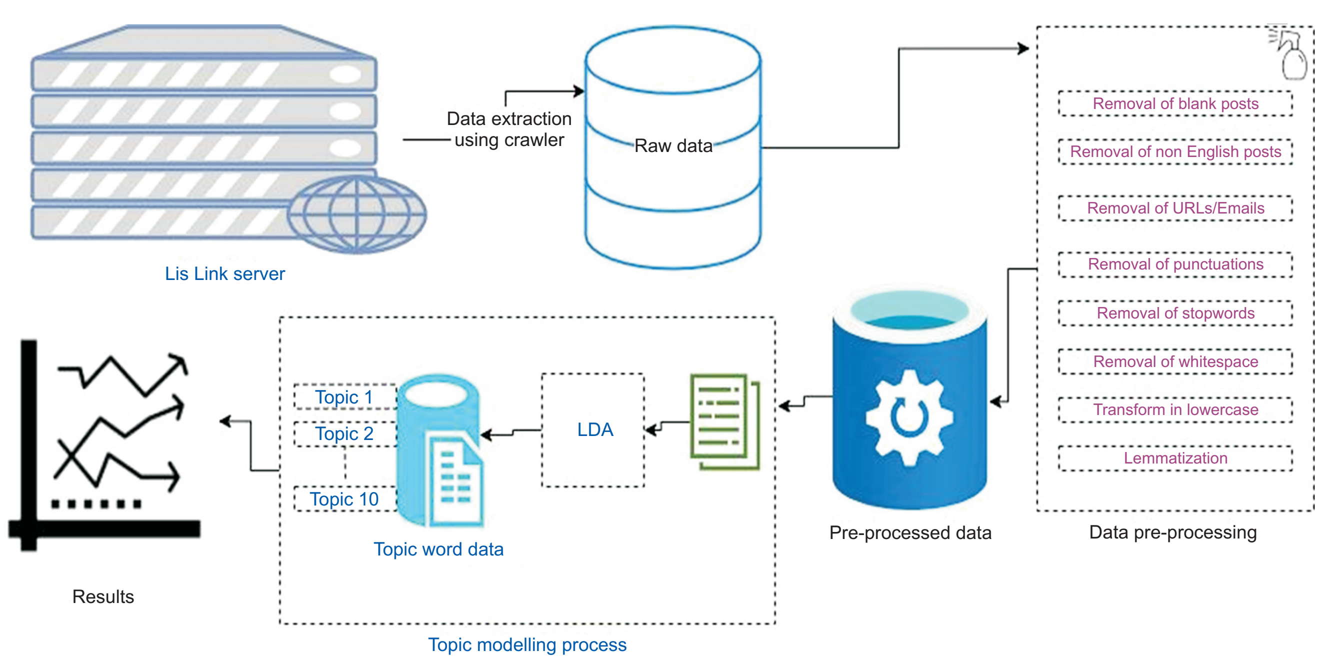

Fig. 1. Methodology of data analysis.

 https://doi.org/10.1633/JISTaP.2022.10.3.4

as topic modeling algorithms to extract key topics from the dataset. The most frequent words with a minimum of 50 were plotted as a wordcloud. They found that the word "dress" was the most frequent.

The above review highlighted the studies of different domains conducted using LDA. This shows that LDA can be regarded as promising for analyzing online discussion forum data. Therefore, the present research is concentrated on the LDA-based topic modeling of online discussion on Lis Links.

3. METHODOLOGY

In this section, we discuss a detailed approach to performing topic modeling on the discussion forum of Lis Links. Fig. 1 shows the methodology of the analysis of text data of Lis Links forum based on the framework presented by Garg and Kanjilal (2019). This includes data collection, pre-processing, and topic modeling.

The data was extracted from Lis Links using R programming. This raw data was pre-processed with operations like removing blank posts, non-English posts, URLs, and email stopwords. Finally, the data was analyzed, and the ten topics were extracted using LDA with ten as the value of k. The post with high gamma scores was selected under that topic. Lastly, based on the gamma score, we visualized the topic-wise word cloud of ten words from 10 topics.

3.1. Data

A total of 10,286 posts were extracted using the crawler from Lis Links. The present study's analysis is focused on the text dataset, so it is essential to harvest only text content. Therefore, the crawling program was written in R using the package "rvest" to extract the text data.

The extracted data was stored in CSV format. The extracted dataset consists of 10 attributes, namely PostID, PostUser, PostDate, PostTime, PostCategory, PostTitle, PostBody, PostView, PostReplyCount, and PostUrl. However, the goal of this study will suffice with the analysis of text data of PostBody. Therefore, we have considered the content of PostBody as it contains detailed information about the post. The sample of the dataset is shown in Table 1.

3.2. Data Pre-Processing

Data pre-processing is one of the key steps in all data-based research to remove unwanted noise from the data. However, this is essential in enhancing text data quality

Table 1. Sample dataset

PostID	PostUser	PostDate	PostTime	Post View	PostReply Count	PostTitle	PostBody	PostCategory	PostUrl
4102	Mohit Garg	February 21, 2014	17:49	734	2	DRTC Admision Notification	DRTC exam dates are out for two year MS in Library and Information Science. They are the best institutes of library science where you will get stipends also. There are no fees for the course. Highly motivated course. Kindly consult the attached file for admission notification.	Other Discussions	http://www.lislinks.com/forum/topics/drtc-admision-notification

(Hariharakrishnan et al., 2017). Many studies have shown that the pre-processing of data is a critical and most time-consuming process (Press, 2016). The standard methods for pre-processing text data include lower casing of text data, stemming, lemmatization, and removal of punctuation, numbers, HTML tags, special characters, stopwords, etc. However, the type of methods used for pre-processing the data is application-dependent (Ignatow & Mihalcea, 2017). For example, we called two additional ways in this study before going for general pre-processing. The process was divided into two stages. Firstly, the blank posts and non-English posts were removed. The Lis Links also provide the functionality of posting images, documents, etc., in the discussion forum. The developed program for data extraction from Lis Links is limited to working with text data only. It does not extract anything other than text from the post, so the program will not harvest anything from the post containing only an image, document, etc., resulting in blank posts. Fig. 2 shows the example of a post where the image was embedded in the post body. Also, there is a possibility that somebody has posted blank posts without any messages. Fig. 3 shows the example of one such post without any content.

The character count of the PostBody of each post was calculated using R to identify the blank posts. The character size of each post was stored in the primary data frame. Analysis found that the posts with a character count from 0 to 7 were blank posts, which have no relevant content, except for the one case where the post only has one word. Therefore, we have removed all these blank posts that were either blank or having less informative text content.

India is a diverse country with multi-language-speaking people. The 22 languages (including Hindi and English) are officially listed as per the Eighth Schedule of the Constitution of India (The Editors of Encyclopaedia Britannica, 2019). Lis Links supports the features of posting content in regional languages with the help of Unicode. It makes Lis Links a national platform to cater to the needs of all LIS professionals in India. Therefore, there is a possibility that posts may contain regional languages also. Each post of a regional language on the Lis Links is stored as a sequence of Unicode. The original post posted in the Hindi language is shown in Fig. 4. The table shows the same post extracted from the Lis Links server data.

The one word of Hindi language is stored as a sequence of Unicode character tags. The R program has fetched the text content and the Unicode of non-English posts.

Fig. 3. Blank post.

Fig. 2. Example of post containing image and extracted blank.

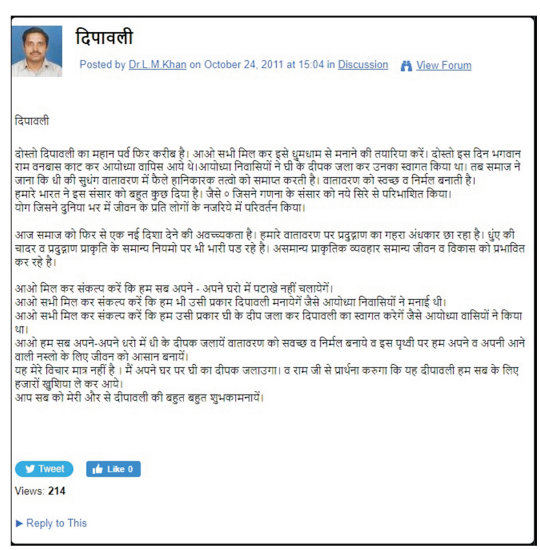

Fig. 4. Regional language post.

The word “दिपावली” is made up of seven Unicode character tags: द (U+0926), “◌ि” (U+093F), “प” (U+092A), “◌ा” (U+093E), “व” (U+0935), “ल” (U+0932), and “◌ी” (U+0940). This is how each post in regional language can easily be identified consisting of Unicode characters. The presence of pattern <U+ was searched and stored as a separate data frame using the pattern matching function in R. These posts were checked manually to avoid any English posts and have some Unicode characters due to some unwanted space or character. The sample of one such post is shown in Table 2. Finally, those posts were considered for the final dataset, which was in English but had one or two Unicode characters. The other posts of regional language were identified and excluded from the final dataset.

In the second stage, URLs were removed from the data. Firstly, a regex was defined for all possible pattern matching of URLs. The URLs were extracted and removed based on this regex. The regex used for URL matching is defined as “(http[^]*)|(ftp[^]*)|(www\\.[^]*).” Similarly, the email mention was removed and stored as a separate CSV file. The regex used for email matching is defined as “\\S*@\\S*.”

The removal of URL and email strings may again leave some of the posts empty. These posts were identified and removed from the primary dataset. After removing URLs and emails from the posts, punctuation and white spaces were removed from the text corpus. The stopwords were removed from the dataset in the next step. The stopwords are words that do not contain much information about the analysis. The list of stop words of SMART (System for the Mechanical Analysis and Retrieval of Text) information retrieval system is the most popular in text analysis. This system was developed in the 1960s at Cornell University (Lewis et al., 2004). The SMART list consists of 571 stop words removed from the dataset (Hvitfeldt & Silge, 2021). The other popular stopword list is snowball stopwords given by Porter (n.d.). Apart from these, there can be some domain and platform-specific stopwords. The removal of these stopwords depends on the purpose of the study. Words like “dear,” “greetings,” “hello,” “hi,” “please,” “kindly,” etc., are commonly used words in writing any correspondence to a person or organization. These are also known as salutations. The use of salutations is done to establish strong communication. Previous studies have found that salutations were common practice in writing an email (Hiranburana, 2017; Lee, 2004). The web-based discussion forum interface and workflow are different from email-based forums. Therefore, we have not excluded these words from exploring the use of salutations on the discussion forums of Lis Links.

The text corpus of Lis Links was converted to lowercase to avoid the issue of similar words considered as different tokens. For example, the system finds Book, BOOK, book, BooK as four tokens, while the literal meaning of all these words is the same. After the lower-casing of the words, the system will identify the word “book” with a frequency of four.

The transformation of different words into root words is an important phenomenon in text analysis. It helps identify the similar token used differently as per the language's grammar. For example, “Libraries” and “Library” convey the same meaning of context related to the library. It can be achieved by applying either stemming or lemmatization. Lemmatization is mapping the word with the root word with the dictionary meaning, while stemming is just matching to words after removal of smaller suffixes like *ies*, *y*, etc. This is why we have used lemmatization in this study to consider dictionary words for the analysis. It will also help to interpret the results quickly with the right words. The removal of numbers is a common practice in pre-processing of online data. However, we did not remove the numbers, as numbers also contain helpful information in the context of this study. The year of examination, admission, or question paper is the key information element to classify the discussion of LIS professionals. This is why we have not removed the numbers from the dataset.

Table 2. Extracted data of regional language

PostID	PostUser	PostDate	PostTime	Post View	Post ReplyCount	PostTitle	PostCategory	PostUrl
8732	Dr. L. M. Khan	October 24, 2011	15:04	194	0	<U+0926><U+093F><U+092A><U+093E><U+0935><U+0932><U+0940>	Other Discussions	http://www.lislinks.com/forum/topics/2013205:Topic:224441

3.3. Post-Pre-Processing

The pre-processed data was considered to visualize the most frequent terms and modeling of the topics. In this research, we have used LDA for extracting the topics from the text corpus. LDA is the most popular topic modeling technique in big data analysis (Garg & Rangra, 2022; Li & Lei, 2021).

The "topic models" package was used for LDA analysis (Grün & Hornik, 2011). The document term matrix (DTM) was created from the pre-processed corpus of the Lis Links discussion forum posts. As the term suggests, DTM is a matrix that consists of the number of occurrences of a term in the document.

> Post 1: "pgdlan ignou clear write exam focus thesis work humble request send format synopsis proceed thank"
> Post 2: "dear friend kindly suggest net library information science coach chennai effective online coach upcoming net exam kindly suggest good study material book syllabus advance"

The above two posts are pre-processed posts. The sample DTM of five terms from these two posts is shown in Table 3.

The above matrix shows that the term "exam" appeared once in both posts. The words "clear," "focus," "format," and "humble" appeared only in the first post, while the second post has a zero count for these terms.

Now, this DTM was used as an argument for the LDA function of the package "topic models." The value of k was set to 10. The k represents the number of topics to be extracted from the dataset. The value of k is the essential parameter in LDA analysis, describing the number of topics to be extracted from the text corpus. The smaller value of k represents the lesser number of general topics, while the larger value of k represents more specific topics. There is no best value of k for all studies, and its value varies for different datasets (Kahani et al., 2016).

Previous studies on mailing lists, an online forum of LIS professionals, have categorized discussions into 7 to 12 categories (Pujar et al., 2014; Siddique et al., 2020). Also, at present, Lis Links has nine (9) pre-defined categories for structuring discussions on the forums. Therefore, we have assumed the value of k to 10 based on the evidence of such previous studies on LIS domain and existing categories on Lis Links.

The extracted topics with their corresponding words were visualized as a Wordcloud. Finally, based on these ten topics, the posts were labeled. The frequency distribution of these topics was calculated.

3.4. Ethical Considerations

The analysis of content posted by people on online and social media platforms has become an emerging research topic among scholars of different academic discipline. These research efforts are supported by the easy availability of vast amounts of data and advanced tools for the whole analysis process. Popular platforms like Facebook, Twitter, and YouTube even provide APIs for accessibility of bulk content posted on these platforms. One essential question that arises here is, who has the right to the content posted on these platforms. Is it the user who has posted the content or the organization that owns the platform?

Sometimes online researchers misunderstand this. If they have access to a tool (data scraping tool) or know how to program, then no other permission is required to access the data. But this is not always the case. The users who posted their experience, reviews, etc., on the platform are generally unaware that the content will be analyzed in different research contexts. The user's consent is a key aspect of any research involving human participants. Before starting research, mutual agreement between researchers and participants should be done regarding how their data will be used. It is common to declare this agreement when collecting data through questionnaires/surveys/interviews, etc. to maintain the confidentiality of the respondent. However, this is quite impossible in online spaces, as the number of users is large compared to traditional survey-based research.

Therefore, it is important for researchers to first understand privacy and data concerns before starting any

Table 3. Document term matrix example

Posts	Terms				
	Clear	Exam	Focus	Format	Humble
Post 1	1	1	1	1	1
Post 2	0	1	0	0	0

 https://doi.org/10.1633/JISTaP.2022.10.3.4

research based in an online environment. A good number of studies have considered the analysis of public data as per ethical standards. The data is publicly available if the content on the online platforms is accessible without any restrictions or requirements of login/passwords (Betts et al., 2014; Buck & Ralston, 2021; Eastham, 2011).

The discussion forum of Lis Links is publicly available and does not require any password to see content posted by any of the users. However, a login is required to start a new post or reply to existing posts. Therefore, the discussion forum of Lis Link can be considered a publicly accessible platform. In this study, we have extracted only the text of the discussions posted on Lis Links. No personal information or images were collected for this research.

Because of the above discussion it was not deemed necessary to obtain consent from each user who posted content on the Lis Link discussion forum. However, written permission has been taken from the creator of Lis Links to pursue this research.

Fig. 5. Wordcloud of frequent words in discussions on Lis Links.

4. RESULTS

The analysis of wordcloud and topic modeling was conducted on pre-processed data.

4.1. Wordcloud

Wordcloud is an easy-to-understand and powerful visual representation of the most frequent words in a text dataset (Bashri & Kusumaningrum, 2017; Dewi et al., 2020). The word cloud originated from the tag cloud used for Flickr (a social networking site) for multiple grouping of photographs (Miley & Read, 2011). In text analytics, word clouds are commonly used for presenting the overall picture of the text content, also known as text summarization (Heimerl et al., 2014). The font size of a word in the wordcloud is the indicator of the frequency of a word (Miley & Read, 2011). The most frequent word is displayed in larger font sizes than other words. This helps get an intuitive idea about the topics discussed in the text.

Fig. 5 illustrates the wordcloud of the 800 most frequent words in the Lis Links posts. The minimum frequency of the words was set to 20. Fig. 6 shows the frequency chart of the 25 most common words. The term "library" appeared in a larger size in pink color, followed by the word "information" in green. This shows the highest occurrence of the words "library" and "information" in the postings of Lis Links. This is not surprising, as Lis Links is a library domain-specific platform. The existence of the words "book," "university," and "librarian" shows that

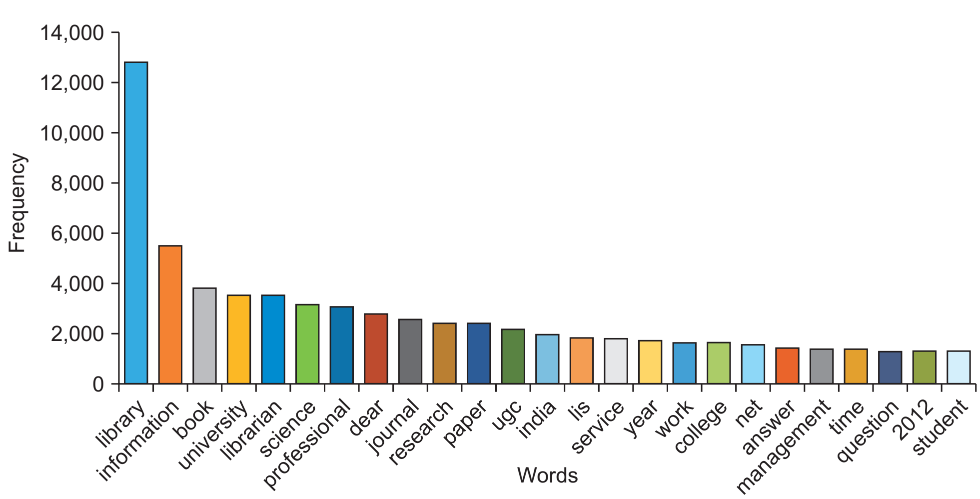

Fig. 6. Twenty-five most frequent words.

Table 4. Word wise beta score

Topic	Word	Beta	Word	Beta
1	koha	5.12E-57	software	3.86E-32
2	koha	9.70E-62	software	0.00022
3	koha	2.44E-20	software	0.000809
4	koha	0.000149	software	0.001037
5	koha	1.26E-77	software	0.000978
6	koha	6.54E-61	software	1.21E-31
7	koha	8.06E-26	software	0.001942
8	koha	6.85E-36	software	5.02E-05
9	koha	2.82E-70	software	0.000133
10	koha	0.01039	software	0.014984

users have discussed institutional or profession-related things in the posts. The terms "UGC," "question," and "answer" indicate posts related to UGC examinations. The word "2012" in a significant font size refers to one of the issues related to the 2012 UGC NET (University Grant Commission National Eligibility Test) Exam. This was the year when Lis Links had the maximum number of posts. Upon further analysis of the dataset, the words "research," "journal," and "paper" were mainly used as a means of posting the discussion related to research. Some posts also appeared to be related to call for papers, data collections using surveys, etc. Thus the frequent words gave an intuitive idea that discussions related to examinations were common in the forums. The most frequent words related to examinations were "application," "question," and "answer." The salutation words "dear" and "kindly" show that LIS professionals post discussions with due gratitude.

Fig. 7. Comparative Wordcloud of words in each ten topics.

4.2. Topic Modeling

The basic assumption of the topic modeling algorithm is that each text document is distributed over a certain fixed number of topics and these topics consist of a specific group of words. The analysis retrieves a group of terms of 10 topics from the text corpus. The most prominent ten words of each topic are listed in Table 4. These groups of words for each topic were selected based on the beta score. The beta score is per topic per word probability. It estimates the occurrence of a word in each of the topics. The table shows the probability of the words "koha" and "software" in all ten topics. Analysis shows that both words have the highest probability for Topic 10 compared to other topics. Thus, the words "koha" and "software" were selected for Topic 10. Similarly, the words of other topics were considered based on high beta score.

Fig. 7 shows the comparative word cloud of the most probable words of all ten topics. The font size of the word is dependent on the beta score. The higher the beta score, the bigger the font size of the word will be. The analysis found that the words "librarian," "post," "college," and "pay" belong to Topic 1; "2012," "India," and "current" are related to Topic 2; "dear," "professional," and "kindly" belong to Topic 3; "code," "information," and "classification" relate to Topic 4; "journal," "research," and "paper" relate to Topic 5;

"ugc," "net," and "candidate" form Topic 6; Topic 7 consists of the words "library," "book," and "public"; "university," "science," and "department" are generated from Topic 8; "librarian," "post," and "pay" belong to Topic 9; and "software," "koha," and "online" form Topic 10.

LDA analysis identifies the per-document-per-topic probability of representation of all ten topics in the post. It is a probability of estimating the proportion of words from the post that formed the given topic. In the next step, the important task is to select the best topic discussed in the post. This can be achieved in multiple ways based on the type of application, i.e., either by the highest probability or by setting the threshold value (Barravecchia et al., 2022). Each one of the approaches has its limitations. But in the literature, the standard approach for selecting a topic is based on the highest probability (Jelodar et al., 2019). All the posts were classified based on the high per-document-per-topic probability, also known as a gamma score.

Table 5 shows the assignments of ten topics from one sample post. The score of 0.2482 of Topic 1 of the post indicated that Topic 1 has a 24.82% representation in the post. The score of 0.5721 shows that 57.21% of the proportions of the post are related to Topic 10; similarly, 14.93% for Topic 3. A minor percentage of representation was observed for Topic 2 and from Topic 4 to Topic 9. This shows that the post consists mainly of Topic 10, Topic 1, and Topic 3, with a major representation of Topic 10. Therefore, Topic 10 has been assigned to the post. Similarly, the topic with a high probability (gamma score) was assigned as a topic with each post. Table 5 shows an example of all ten topics and their related probability of one post.

In the final step, it is necessary to transform these topics into human understandable topics. The title of the topics as Topic 1, Topic 2,...., and Topic 10 make no sense unless it is known what these topics are broadly discussing. Therefore, we have assigned the human-readable topics based on the constituents of words of each topic. The categorization of the posts by earlier studies on LIS discussions was also considered (Pujar et al., 2014; Siddique et al., 2020). The posts' ten categories or topic names are shown in column 3 of Table 6.

Topic 1 is described by the words "librarian," "post," "college," "assistant," "pay," "library," "university," "ugc," "interview," "school," etc. It shows that Topic 1 is related to the "LIS Recruitment" post. Recruitment is the process of finding suitable employees for the organization. Generally, the discussion of recruitment is surrounded by the advertisement of the post, pay or remuneration of the post,

Table 5. Gamma score of post per topic

PostUser	PostDate	PostTime	PostView	PostReply	PostTitle	PostCategory	PostURL	PostClassified	Gamma score
Ramkisan more	April 6, 2020	16:44	158	5	Enabling Web OPAC in e-Granthalaya 3.0 Software	Other Discussions	http://www.lislinks.com/forum/topics/enabling-web-opac-in-e-granthalaya-3-0-software	Topic 1	0.2482
								Topic 2	0.0043
								Topic 3	0.1493
								Topic 4	0.0043
								Topic 5	0.0043
								Topic 6	0.0043
								Topic 7	0.0043
								Topic 8	0.0043
								Topic 9	0.0043
								Topic 10	0.5721

Table 6. List of topic word and new category

Topic	Words	Category name
Topic 1	librarian, post, college, assistant, pay, library, university, ugc, interview, school	LIS Recruitment
Topic 2	2012, india, download, current, affair, indian, week, jun, world, bank	LIS Issues
Topic 3	dear, professional, lis, link, friend, kindly, library, information, member, share	Other Discussion
Topic 4	library, information, iii, answer, correct, ans, code, classification, TRUE, system	LIS Education
Topic 5	journal, library, information, research, paper, science, publish, international, author, management	LIS Research
Topic 6	ugc, net, candidate, paper, exam, answer, mark, question, result, test	LIS Exams
Topic 7	library, book, journal, college, public, year, professional, school, student, dear	General Information related to Library
Topic 8	university, library, science, information, india, research, librarian, department, phd, prof	LIS Admission
Topic 9	library, work, information, make, read, librarian, time, knowledge, book, professional	Library and Professional Activities
Topic 10	library, software, user, koha, service, open, datum, access, system, online	Information Communication Technology (ICT)

specification/profile of the job, etc. Hence Topic 1 was named LIS Recruitment, consisting of messages where professionals discussed job-related matters. The sample post labeled as "Topic 1" or "LIS Recruitment" is shown in Table 7. The post discusses an advertisement for school librarians by DSSSB.

Topic 2 talks about the "LIS Issues" through the words "2012," "India," "download," "current," "affair," "Indian," "week," "jun," "world," "bank," etc. The LIS issues are all about the problems related to the LIS domain faced by people. The sample of the posts labeled as Topic 2 in Table 7 shows the concern of professionals.

Topic 3 represents the topic of "Other Discussion," which consists of the words "dear," "professional," "lis," "link," "friend," "kindly," "library," "information," "member," "share," etc. It consists of posts related to multiple subtopics of LIS.

The fourth topic highlights the post related to "LIS Education" which consists of the words "library," "information," "answer," "correct," "ans," "code," "classification," "TRUE," "system," etc. This means the discussions are related to LIS courses, theoretical concepts of LIS, etc.

The words "journal," "library," "information," "research," "paper," "science," "publish," "international," "author," and "management" suggest that Topic 5 discusses "LIS Research." The components of research include data collection, analysis, and publication. Therefore, we concluded that Topic 5 is related to LIS research.

Topic 6 consists of the words "ugc," "net," "candidate," "paper," "exam," "answer," "mark," "question," "result," "test," etc. These words revealed that Topic 6 talks about the post related to LIS exams. The examination process involves the candidate, question paper, answer key, marks of the question paper, exam name like UGC NET, exam results, etc.

The words "library," "book," "journal," "college," "public," "year," "professional," "school," "student," or "dear" highlight that the posts were broadly discussing "General Information related to the Library." It includes information about some special libraries and their services, or achievements of any particular library.

Topic 8 discusses admission notification and its related subtopics. The words include "university," "library," "science," "information," "India," "research," "librarian," "department," "phd," or "prof."

The analysis found that in Topic 9, library and professional activities were widely discussed. The top 10 words in this category are "library," "work," "information," "make," "read," "librarian," "time," "knowledge," "book," and "professional." The professional has a keen interest in organizing activities in the library to enhance usage.

In Topic 10, the generated words contain "library," "software," "user," "koha," "service," "open," "datum," "access," "system," and "online." A professional posted a post related to information and communication technology on this broader topic. The terms "software" and "koha" signify discussions related to open source software for library man-

Table 7. Classification of posts as per revised category

Sl. No.	PostTitle	PostCategory	PostClassified	Revised PostCategory
1	DSSSB Librarian 197 Post: Advertisement No. 04/20, Post Code 92/20 Kept in Abeyance	Circular of Importance to LIS	Topic 1	LIS Recruitment
2	How to Get Experience When No One is Interested to Recruit a Fresher	Other Discussions	Topic 2	LIS Issues
3	Where to Purchase Rare Books?	Other Discussions	Topic 3	Other Discussion
4	How many Auxiliary tables are there in DDC 23rd Edition	Other Discussions	Topic 4	LIS Education
5	Call for Chapter for the Book on "Virtual and Online Classrooms: Digital Resources, Internet Learning and Open Courseware"	Call for Articles for Publication	Topic 5	LIS Research
6	NTA UGC NET June 2020 Notification	UGC NET / SET Examination	Topic 6	LIS Exams
7	IFLA Recognised the Initiative taken by Kota Public Library in Covid-19 and Global Public Libraries	Other Discussions	Topic 7	General Information related to Library
8	CUTN, M.Lib.I.Sc. and Ph.D. Online application is extended till 23rd May 2020	Admission Notification	Topic 8	LIS Admission
9	How to Conduct "Meet the Author" Program in CBSE School	Other Discussions	Topic 9	Library and Professional Activities
10	Enabling Web OPAC in e-Granthalaya 3.0 Software	Other Discussions	Topic 10	Information Communication Technology (ICT)

Table 8. Distribution of post as per revised category

Topic	Human readable topic	Number of posts	%
Topic 1	LIS Recruitment	1,067	10.73
Topic 2	LIS Issues	185	1.86
Topic 3	Other Discussion	1,557	15.66
Topic 4	LIS Education	636	6.40
Topic 5	LIS Research	845	8.50
Topic 6	LIS Exams	1,621	16.30
Topic 7	General Information related to the Library	1,442	14.50
Topic 8	LIS Admission	937	9.42
Topic 9	Library and Professional Activities	707	7.11
Topic 10	Information Communication Technology (ICT)	946	9.51
Grand total		9,943	100.00

agement systems like koha.

Table 8 shows the distribution of posts as per the revised category. The analysis found that most of the discussions were in the category of "Lis Exams." It shows that LIS professionals are keen to seek or share information related to LIS exams. LIS exams is one of the key topics where study material, question papers, and proper guidance are required for passing the exam. In India, the popular LIS exams include UGC NET, UGC JRF (University Grant Commission Junior Research Fellowship), SET (State Eligibility Test), KVS (Kendriya Vidyalaya Sangathan), NVS (Navodaya Vidyalaya Samiti), MOOC, and university entrance exams. The UGC NET exam is a national level exam for qualifying the eligibility test for an assistant professor, assistant librarian, and jobs of the same cadre. These jobs are the highest job level at the entry level in India. UGC JRF is the national level exam for getting a fellowship for pursuing the MPhil/Ph.D. in India. Under this scheme, Rs. 25,000/- per month is given to each qualified candidate for the initial two years and Rs. 28,000 per month for the next two years with the annual contingency of Rs. 10,000/- per annum. It was also found that professionals actively interact on topics related to "General Information related to the Library" (14.50%, 1,442), LIS Recruitment (10.73%, 1,067), ICT (9.51%, 946), and LIS Admission (9.42%, 937). Few discussions were observed on the theme "LIS Issues" (1.86%, 185).

5. DISCUSSION

This study employed text mining-based methods to conceptualize the online discussions of LIS professionals. The raw corpus of the discussions in the text format was extracted from Lis Links and pre-processed using different packages of R. Frequency analysis found that the most frequent words were library, information, university, librarian, book, professional, science, research, paper, question, answer, management, etc. The word cloud was plotted to show the frequency of all such common words compared to other words.

Further, the popular topic modeling algorithm LDA was implemented on the text corpus to identify the topic of discussion among LIS professionals. The document per topic probability (gamma score) and topic per word probability (beta score) are the important parameters for identifying the right topic for each document and constituent words for each topic. The beta score of each topic per word was calculated, and words were assigned to each topic based on the highest beta score. Similarly, the gamma score for each post per topic was calculated, and topics were assigned based on the highest gamma score. The ten topics identified were "LIS Recruitment," "LIS Issues," "Other Discussion," "LIS Education," "LIS Research," "LIS Exams," "General Information related to Library," "LIS Admission," "Library and Professional Activities," and "Information Communication Technology (ICT)." The topics were closely related to various previous studies of content analysis of online discussions of LIS professionals (Choi et al., 2019; Pujar et al., 2014; Siddique et al., 2020).

Each post on the forum was labelled with one of these topics and the frequency distributions of all ten topics were calculated. The results revealed that most discussions were tagged under the topic of "LIS Exams." This showed the most significant interest of the LIS professionals in virtually sharing Information related to various jobs exams like KVS and NVS, and professional examinations like NET/SET, with the LIS community.

These findings affirm that LIS Links is a good platform for discussing LIS exams. For example:

> "Hi.. Everybody can anyone tell me any trusted institution or place or person name and who provide the coaching for UGC-NET exam for library science in Delhi..."

The example shows that professionals are looking for a coaching institution or tutor for UGC-NET (University Grant Commission-National Eligibility Test). LIS professionals also post on the forum to grab the attention of fellow community members on different issues like disparity in salary grade or anomalies in the exams. For example:

> "Please anyone move Supreme Court in Unequally and Unfair Result in UGC-NET June 2012. Like West Bengal School Service Commission. Please any body move this situation. I hope This result not to fair in our society. please move. hope success."

> "Sign the Petition Kindly join hands for the cause of disparity created by UGC in the pay level of Library Information Assistant and sign the petition to support the profession and the professionals. Dhttp://chng.it/r6GZ9pFp"

The above two are examples of encouraging professionals to participate in raising their voices against issues. This study's overall findings show that the discussion forum of Lis Links is an effective channel of communication for a wide range of topics. It has become a support system for the community of Indian LIS professionals in the online sphere.

6. THEORETICAL AND PRACTICAL IMPLICATIONS

This study contributes to research in the literature on the analysis of online data/user-generated content and management of online forums for LIS professionals. For a better understanding of online engagements of LIS professionals, this study implemented text mining-based methods and identified the key topic of discussions. LIS researchers commonly use survey methods to know the various challenges professionals face.

Rather than depending on survey data, LIS practitioners can analyze online conversations on different platforms to explore the trends of discussions among LIS professionals. The findings of this study can help organizations/policy makers/senior LIS professionals to understand the issues faced by professionals in their day-to-day life. The findings offer user-based categories that can be used for managing and organizing discussions on Lis Links. This study provides a detailed implementation workflow, and the same can be applied on similar forums, email lists, or different social media platforms.

7. LIMITATIONS

In addition to the number of contributions, this study has a few limitations. First, the present study only focused on LIS professionals registered with Lis Links online discussion forums; this could be a case of a statistically biased sample (Barbierato et al., 2022).

Second, this study could not extract the replies of community members on each post, and this limitation restricted the investigation of response analysis of the professionals and how they reacted to the particular post. Social media platforms like Facebook have the feature of reacting in five different ways and with comments, but Lis Links has the feature of commenting only. Social network analysis of these responses could provide insightful results. Third, the present analysis framework considered online discussion in the text format only and excluded other formats like images, pdf, or URLs of any external source. However, posts with both attachments and enough text content to model the discussion were considered for the analysis. One such example is shown below:

> "Here is the total list of LIS candidates qualified from each center in July 2018 UGC NET Attached here: List1.pdf"

Posts which did not have any information about the attachment or URL mentioned in the posts could not be included in the analysis. One such sample of a post is given below:

> "Kindly consult the attached file or consult this link: http://kvsangathan.nic.in/EmploymentDocuments/EMP-NTC-27-10-14.PDF"

Fourth, the LDA topic model is based on the bag of words model, where semantic context is ignored. Lastly, we have assumed the value of k (number of topics) to be 10, based on the themes created by the previous studies, which resulted in a general category of topics. The study with high value of k will generate more specific topics.

8. CONCLUSION

In recent times, topic modeling algorithms, especially LDA, have attracted researchers from all domains for extracting topics from text data originating from different sources. The present study results are explorative and help in understanding the topic of discussion among LIS professionals. It also provides implications for the development and management of online discussion platforms for LIS professionals. This paper provides a structured procedure for implementing the LDA technique and models the online discussions of LIS professionals expressed on Lis Links within ten topics. The LIS professionals have actively posted their discussions on these ten topics. However, most of the conversations were concentrated on "LIS Exams," and very few discussions were found on "LIS Issues."

To the best of our knowledge, the present study is the first attempt in the domain of LIS to use a text-mining-based methodology for analysis of an online discussion forum of LIS professionals. Earlier studies on the forum or the email list used the manual content analysis methodology for classifying topics. Future research could compare these study results with the analysis of discussion on other community-based and social media platforms. As discussed in the limitations, future research can also be taken up to analyze these images and PDFs through image processing to present the granularity of the results. Further exploration of LDA on other Indian LIS forums is needed for a deeper understanding of the categories of the posts. The combination of high and low value of k may be used to identify both general categories and their specific subcategories.

ACKNOWLEDGMENTS

The authors are grateful to all the anonymous reviewers for their deeply helpful feedback to enhance the qual-

ity of this research work.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Abdillah, O., & Adriani, M. (2015, March 24-27). Mining user interests through internet review forum for building recommendation system. In L. Barolli, M. Takizawa, F. Xhafa, T. Enokido, & J. H. Park (Eds.), *Proceedings of the IEEE 29th International Conference on Advanced Information Networking and Applications Workshops* (pp. 564-569). IEEE.

Ahn, J., Son, H., & Chung, A. D. (2021). Understanding public engagement on Twitter using topic modeling: The 2019 Ridgecrest earthquake case. *International Journal of Information Management Data Insights,* 1(2), 100033. https://doi.org/10.1016/j.jjimei.2021.100033.

Arden, M. A., Duxbury, A. M., & Soltani, H. (2014). Responses to gestational weight management guidance: A thematic analysis of comments made by women in online parenting forums. *BMC Pregnancy and Childbirth,* 14, 1-12. https://doi.org/10.1186/1471-2393-14-216.

Barbierato, E., Bernetti, I., & Capecchi, I. (2022). Analyzing TripAdvisor reviews of wine tours: An approach based on text mining and sentiment analysis. *International Journal of Wine Business Research,* 34(2), 212-236. https://doi.org/10.1108/IJWBR-04-2021-0025.

Barman, B. (n.d.). *Lis Links*. http://www.lislinks.com.

Barravecchia, F., Mastrogiacomo, L., & Franceschini, F. (2022). Digital voice-of-customer processing by topic modelling algorithms: Insights to validate empirical results. *Journal of Quality & Reliability Management,* 39(6), 1453-1470. https://doi.org/10.1108/IJQRM-07-2021-0217.

Bashri, M. F. A., & Kusumaningrum, R. (2017, May 17-19). Sentiment analysis using Latent Dirichlet allocation and topic polarity wordcloud visualization. In H. S. Lim, Y. H. Pang, Y. Rusmawati, & J. Tirtawangsa (Eds.), *Proceedings of the 5th International Conference on Information and Communication Technology (ICoIC7)* (pp. 1-5). IEEE.

Betts, D., Dahlen, H. G., & Smith, C. A. (2014). A search for hope and understanding: An analysis of threatened miscarriage Internet forums. *Midwifery,* 30(6), 650-656. https://doi.org/10.1016/j.midw.2013.12.011.

Buck, A. M., & Ralston, D. F. (2021). I didn't sign up for your research study: The ethics of using "public" data. *Computers and Composition,* 61, 102655. https://doi.org/10.1016/j.compcom.2021.102655.

Choi, S., Dukic, Z., & Hill, A. (2019). Professional networking with Yahoo! Groups: A case of school librarians from international schools in Hong Kong. *Journal of Librarianship and Information Science*, 51(4), 1077-1090. https://doi.org/10.1177/0961000618763488.

Coulson, N. S. (2005). Receiving social support online: An analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. *Cyberpsychology & behavior,* 8(6), 580-584. https://doi.org/10.1089/cpb.2005.8.580.

Dewi, I. N., Nurcahyo, R., & Farizal. (2020, April 16-21). Word cloud result of mobile payment user review in Indonesia. *Proceedings of the IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 989-992). IEEE.

Eastham, L. A. (2011). Research using blogs for data: public documents or private musings? *Research in Nursing & Health,* 34(4), 353-361. https://doi.org/10.1002/nur.20443.

Garg, M., & Kanjilal, U. (2019). A framework to process text data of web discussion forums a study of LisLinks. *DESIDOC Journal of Library & Information Technology,* 39(06), 315-321. https://doi.org/10.14429/djlit.39.06.15145.

Garg, M., & Rangra, P. (2022). Bibliometric analysis of Latent Dirichlet allocation. *DESIDOC Journal of Library & Information Technology,* 42(2), 105-113. https://doi.org/10.14429/djlit.42.2.17307.

Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software,* 40(13), 1-30. https://doi.org/10.18637/jss.v040.i13.

Hariharakrishnan, J., Mohanavalli, S., Srividya, & Sundhara Kumar, K. B. (2017, January 10-11). Survey of pre-processing techniques for mining big data. *Proceedings of the 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)* (pp. 1-5). IEEE.

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014, January 6-9). Word cloud explorer: Text analytics based on word clouds. In R. H. Sprague, Jr. (Ed.), *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 1833-1842). IEEE.

Hiranburana, K. (2017). Use of English in the Thai workplace. *Kasetsart Journal of Social Sciences,* 38(1), 31-38. https://doi.org/10.1016/j.kjss.2015.10.002.

Hvitfeldt, E., & Silge, J. (2021). *Stop words.* https://smltar.com/stopwords.

Ignatow, G., & Mihalcea, R. (2017). Basic text processing. In G. Ignatow, & R. Mihalcea (Eds.), *Text mining: A guidebook for the social sciences* (pp. 52-61). Sage.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., &

Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications,* 78(11), 15169-15211. https://doi.org/10.1007/s11042-018-6894-4.

Kahani, N., Bagherzadeh, M., Dingel, J., & Cordy, J. R. (2016, October 2-7). The problems with eclipse modeling tools: A topic analysis of eclipse forums. In J. DeAntoni (Ed.), *Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems (MODELS' 2016)* (pp. 227-237). ACM.

Kim, Y. B., Lee, J., Park, N., Choo, J., Kim, J. H., & Kim, C. H. (2017). When Bitcoin encounters information in an online forum: Using text mining to analyse user opinions and predict value fluctuation. *PloS One,* 12(5), e0177630. https://doi.org/10.1371/journal.pone.0177630.

Lee, C. F. K. (2004). Written requests in emails sent by adult Chinese learners of English. *Language, Culture and Curriculum,* 17(1), 58-72. https://doi.org/10.1080/07908310408666682.

Lewis, D. D., Yang, Y., Rose, T. G., & Li, F. (2004). RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research,* 5, 361-397. https://www.jmlr.org/papers/volume5/lewis04a/lewis04a.pdf.

Li, X., & Lei, L. (2021). A bibliometric analysis of topic modelling studies (2000–2017). *Journal of Information Science,* 47(2), 161-175. https://doi.org/10.1177/0165551519877049.

McKenna, E., & Thomson, M. (2014). Demand response behaviour of domestic consumers with photovoltaic systems in the UK: An exploratory analysis of an Internet discussion forum. *Energy, Sustainability and Society,* 4, 13. https://doi.org/10.1186/2192-0567-4-13.

Miley, F., & Read, A. (2011). Using word clouds to develop proactive learners. *Journal of the Scholarship of Teaching and Learning,* 11(2), 91-110. https://eric.ed.gov/?id=EJ932148.

Munezero, M., Kojo, T., & Männistö, T. (2017, November 9-10). An exploratory analysis of a hybrid OSS company's forum in search of sales leads. In B. Randall (Ed.), *Proceedings of the 11th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (pp. 442-447). IEEE.

Muñoz-Cañavate, A., Fernández-Falero, M. R., & Hurtado-Guapo, M. A. (2017a, November 1-3). Information capture and knowledge sharing systems in the field of library and information science: The case of MEDLIB-L in medicine. In K. Liu, A. C. Salgado, J. Bernardino, & J. Filipe (Eds.), *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KMIS)* (pp. 181-188). Science and Technology Publications.

Muñoz-Cañavate, A., González, A. C., Hípola, P., & Miranda, E. A. C. (2017b, October 18-20). Mailing lists on the Internet - A collaboration tool that is still alive. The case of the rediris lists. In P. Isaías, & H. Weghorn (Eds.), *Proceedings of the 2017 International Conference on WWW/Internet: Applied Computing* (pp. 261-266). IADIS.

Omidvar, A., Garakani, M., & Safarpour, H. R. (2014). Context based user ranking in forums for expert finding using WordNet dictionary and social network analysis. *Information Technology and Management,* 15(1), 51-63. https://doi.org/10.1007/s10799-013-0173-x.

Özcan-Tok, E., Özmen, M. U., Tok, E., & Yılmaz, T. (2019). The impact of collective action and market prices: Evidence from an online agricultural discussion forum. *Online Information Review,* 43(4), 565-583. https://doi.org/10.1108/OIR-08-2018-0243.

Pandapotan, I. M., Alamsyah, A., & Paryasto, M. (2015, May 27-29). Indonesian music fans group identification using social network analysis in Kaskus forum. In M. A. Bijaksana, D. D. Jatmiko, A. T. Wibowo, Y. Redityamurti, M. Arzaki, & I. Asror (Eds.), *Proceedings of the 3rd International Conference on Information and Communication Technology (ICoICT)* (pp. 322-326). IEEE.

Porter, M. (n.d.). *Snowball.* https://snowballstem.org.

Press, G. (2016). *Cleaning big data: Most time-consuming, least enjoyable data science task, survey says.* https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=28c5b0ee6f63.

Pujar, S. M., Mahesh, G., & Jayakanth, F. (2014). An exploratory analysis of messages on a prominent LIS electronic discussion list from India. *DESIDOC Journal of Library & Information Technology,* 34(1), 23-27. https://doi.org/10.14429/djlit.34.1.5942.

Qian, Y., & Gui, W. (2021). Identifying health information needs of senior online communities users: A text mining approach. *Aslib Journal of Information Management,* 73(1), 5-24. https://doi.org/10.1108/AJIM-02-2020-0057.

Saranya, M. S., & Geetha, P. (2020, July 28-30). Word cloud generation on clothing reviews using topic model. *Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 177-180). IEEE.

Sawant, S., & Sawant, P. (2016). Indian LIS job market and its visibility through portals and mailing lists/forums. *SRELS Journal of Information Management,* 53(5), 387-391. https://doi.org/10.17821/srels/2016/v53i5/96051.

Shukla, A., & Dawngliana, J. M. (2018). Do online professional forums promote professional contents effectively? An analytical study of new millennium LIS professionals (NMLIS). *International Journal of Library and Information Studies,* 8(1), 61-70. https://www.ijlis.org/articles/do-online-professional-forums-promote-professional-contents-effectively-an-analytical-study-of-new-millennium-lis-profes.pdf.

Siddique, N., Shafi Ullah, F., Mahmood, K., & Ajmal Khan, M. (2020). Professional networking with emailing groups: A case of Pakistan Library Automation Group. *Journal of Librarianship and Information Science,* 53(3), 499-509. https://doi.org/10.1177/0961000620965668.

Singh, S., Chauhan, T., Wahi, V., & Meel, P. (2021, April 8-10). Mining tourists' opinions on popular Indian tourism hotspots using sentiment analysis and topic modeling. *Proceedings of the 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1306-1313). IEEE.

The Editors of Encyclopaedia Britannica. (2019). *Indian languages.* https://www.britannica.com/topic/Indian-languages.

Wang, C., & Tang, X. (2016). Stance analysis for debates on traditional Chinese medicine at Tianya forum. In H. Nguyen, & V. Snasel (Eds.), *International Conference on Computational Social Networks. CSoNet 2016: Computational Social Networks* (pp. 321-332). Springer.

Zarra, T., Chiheb, R., Faizi, R., & Afia, A. E. (2018, May 2-5). Student interactions in online discussion forums: Visual analysis with LDA topic models. *Proceedings of the 2018 International Conference on Learning and Optimization Algorithms: Theory and Applications (LOPAL '18)* (pp. 1-5). ACM.