

Sentiment Analysis of Social Network for Information Professionals: A Case Study of LisLinks Discussion Forum

Vinay Anand

Research Scholar, Department of Library & Information Science, University of North Bengal, Raja Rammohunpur, West Bengal, India.

Corresponding Author: rs_vinay@nbu.ac.in

ORCID iD: <https://orcid.org/0000-0002-7748-1223>

Saptarshi Ghosh

Professor, Department of Library & Information Science, University of North Bengal, Raja Rammohunpur, West Bengal, India.

sgosh@nbu.ac.in

ORCID iD: <https://orcid.org/-0000-0001-9567-5868>

Received: 22 June 2022

Accepted: 23 January 2023

Abstract

This study aims to analyze members' sentiments in India's one of the most used Library and Information Science forums, the LisLinks discussion forum. We retrieved 10,420 discussion posts and their five replies to analyze the sentiment: Positive, Negative, and Neutral, using the RapidMiner tool through a lexicon-based approach. The study results show that 64% of the replies have neutral sentiment, 32% have positive sentiment, and only 4% have a negative sentiment. The more neutral sentiment indicates that the forum members are not actively participating in the discussion site. When we compare the sentiment of the five replies, the results show an increment in neutral sentiment from the first to the fifth. The undertone of the research finding pertains to the users' gradual decremental use of this discussion platform. This study may aid administrators of the forum in grasping the feelings and opinions of the members to enhance the quality of the discussion forum.

Keywords: Discussion forum, Sentiment Analysis, Social Network, LisLinks, RapidMiner, Octoparse.

Introduction

Over the past few years, the rapid development of the Internet and World Wide Web (WWW) technologies has led to a remarkable transformation. It contributes to disseminating knowledge and depicts the simplified data exchange procedure. Individuals can easily retrieve information. Social networking has become a new paradigm because of information technology's rapid and significant evolution. Individuals use social media to express their views, beliefs, feelings, fears, perceptions, experiences, and day-to-day actions. They share their perspectives on various topics, products, and services across many sites, including Twitter, Facebook, Quora, Yahoo Answers, Blogs, and other domain-specific forums (Younis, 2015). Those kinds of platforms do not have proprietary content. Much of it contains user-generated content; thus, these knowledge repositories are mostly user-driven to produce data for the specific domain.

Liu (2010) pioneered sentiment analysis. The exploration of opinion and the analysis of subjectivity refer to the method of determining the attitude or polarity. Sentiment analysis is a Natural Language Processing and Information Extraction work that examines many documents

to ascertain the writer's emotions represented in positive or negative remarks, queries, and requests. In general, sentiment analysis determines the overall functioning of a speaker, an author's point of view, or a publication. Sentiment Analysis uses a variety of categorization approaches to determine the tone of a piece of text. It shows the text's positivity, negativity, or neutrality. This analysis may be performed on massive data sets, and the resulting information may be used in various scenarios (Bhonde, Bhagwat, Ingulkar & Pande, 2015). Sentiment analysis is widely used for the detection of the polarity of opinions. It is always hard to find views because they are inconsequential. It is hidden in extensive forums and blog posts. The need for automated systems for exploring and synthesizing opinions calls for creating these computerized systems. This is where sentiment analysis (opinion mining) is born (Liu, 2010).

Lislinks.com is a domain-specific website dedicated to library and information science, established in 2008, and has become one of the essential portals among library practitioners in India (Shaveta, 2016). LisLinks hosts a community website at *www.lislinks.com/forum*, where library professionals and students can discuss various topics relevant to their careers and education. It is an area in which participants share their views and ideas among the group members through the website to clarify fuzzy information. These responses convey the responder's emotions in three border categories, i.e., Positive, Negative, and Neutral. It demonstrates the respondent's satisfaction with the response or specific facts.

The study aims to analyze the emotions of the members of the LisLinks discussion forum using the lexicon-based machine learning method. Henceforth, the objectives of the study include the following:

- a. To identify the most accessible method of obtaining replies to inquiries is through discussion forums.
- b. To trace the compatibility of radical beliefs and sentiments of the members.
- c. To explore whether forum replies are purposive to the knowledge gaps of the users of this forum.
- d. To assess the interest of members in this discussion forum.

Sentiment analysis Techniques

From a technical point of view, machine learning, lexicon-based, statistical, and rule-based approaches are the most popular techniques for sentiment analysis (Collomb, Joyeux, Hasan & Brunie, 2014). Figure 1 presents a detailed classification of sentiment analysis techniques. In this study, we used a lexicon-based approach for sentiment analysis.

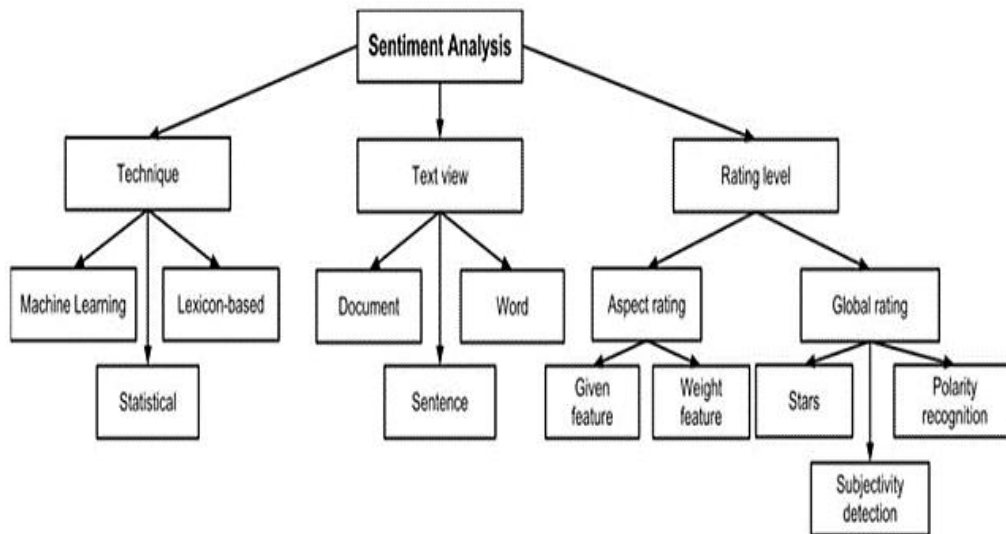


Figure 1: Sentiment Analysis Techniques (Collomb et al., 2014)

Literature Review

A quantitative analysis of the LisLinks discussion forum was conducted to obtain the common theme. Using an R-coded application, ten thousand two hundred eighty-six posts were extracted from the LisLinks discussion forum. The analysis found that most posts were under the "Other Discussion" category. The results also revealed the adverse interest in the forum. To verify this result, we formulated one of the objectives to assess the interest of the forum members (Garg & Kanjilal, 2021). Another study on the LisLinks discussion forum presented a framework for data harvesting and pre-processing for the LisLinks discussion forum. The framework was implemented using R statistical programming language (Garg & Kanjilal, 2019). In our study, we provide a simple method to extract the data automatically from the LisLinks discussion forum. An exploratory analysis of email postings, such as the number of posts, contributors, and discussion topics on a prominent LIS electronic discussion list of India, 'LIS-Forum,' was carried out between 2006 and 2011. The primary objective was to find the relevancy and popularity of LIS-Forum. The results found that during the period of study, an average of 1600 postings per year are being circulated on the LIS-Forum, out of which the 'announcements' category has significant postings (Pujar, Mahesh & Jayakanth, 2014).

People's opinions play a vital role in formulating government policies. Opinion mining was done on tweets regarding government projects using a lexicon-based approach. The exact process is followed in our study. The results showed that most people appreciate and support this project (Verma, Khanday, Rabani, Mir & Jamwal, 2019). The sentiment analysis technique was used to understand people by gender, especially those suffering from an acute disease. The study was done on medical web forums to ascertain patients' emotions, needs, and sentiments to facilitate patient information exchange (Park & Woo, 2019). An open approach was presented to analyze the sentiments of tweets using a statistical tool, R programming. Three thousand tweets were extracted and analyzed. The results show that E-healthcare people have positive sentiment as its positive sentiment score (477) is higher than the negative sentiment score (304) (Saini, Punhani, Bathla & Shukla, 2019). An analysis was conducted on the Twitter post on the conflict in Colombia, which began with the agreement signed in 2016. Two hundred fifty tweets from Colombians and 250 tweets from foreigners were collected.

The data was evaluated in "TEXT2DATA", an Excel Extension. The results found that the tweets of foreigners have more positive feelings than Colombian tweets. The comparison of two different models was carried out for sentiment analysis of movie review posts. The first model of summarization is built using the Aylien Text Analysis extension. The second model is built using the Text processing extension. Both extensions were used in the RapidMiner tool. We used the same device for our study. The results found an accuracy of 90% in the first model and 96% in the second model (Alsaqer & Sasi, 2017).

The literature review finds several studies have been undertaken across various disciplines, but few are in the library and information science domain. The emotions of many libraries' focus groups need to be evaluated. This study is carried out with a quantitative analysis of the members' sentiments on the LisLinks discussion site.

Materials and Methods

The steps involved in the methodology are described as follows.

- I. **Data Access:** - We used Octoparse, a free, open-source web scraping application for extracting data from websites and saving it to Excel, HTML, Text, or directly to a database, to crawl LisLinks' online forum.
- II. **Data Pre-processing:** - We used the Rapid Miner tool for data pre-processing. It provides an integrated environment for machine learning, data preparation, text mining, model deployment, business analytics, and predictive analytics. We removed HTML tags, punctuation, and stop words. Using the Rapid Miner tool, we built a model for sentiment analysis of replies in the LisLinks discussion forum.
- III. **Data Analysis:** - We used the 'Vader Sentiment Extract' operator in the Rapid Miner tool to analyze the replies' sentiments. This package is a lexicon-based technique that utilizes a collection of positive and negative terms. It performs analysis tasks such as discovering associative laws, determining which terms are more frequent, and performing sentiment analysis using the hierarchical representation created in the preceding process. A scoring feature assigns a score to each reply in the LisLinks forum and then separates sentiments (Elbagir & Yang, 2019).

Figure 2 shows the detailed step-wise adopting framework from data access to data analysis for sentiment analysis in the LisLinks discussion forum.

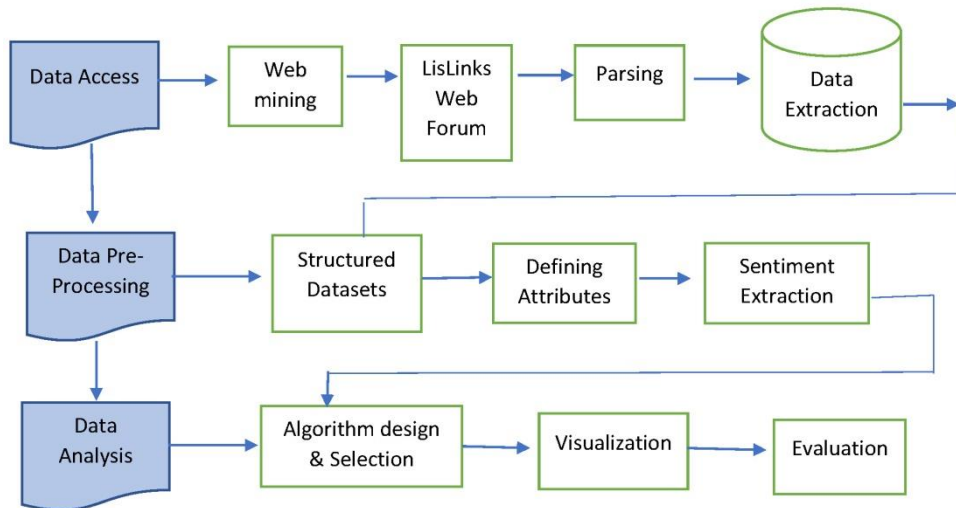


Figure 2: Framework for Sentiment Analysis for LisLinks Forum

Data Collection

Various APIs (Application Programming Interfaces) written in a high-level programming language can access and extract user-generated data from discussion forums and social media platforms such as Quora, Twitter, and Facebook. The LisLinks discussion platform lacks a specific API (Garg & Kanjilal, 2021). As a result, extracting the essential structured data from the LisLinks forum is extremely difficult. To collect data from the LisLinks discussion site we used Octoparse 8.2, a publicly accessible tool with limited features. Figure 3 illustrates the data extraction procedure captured from the Octoparse web scraping tools.

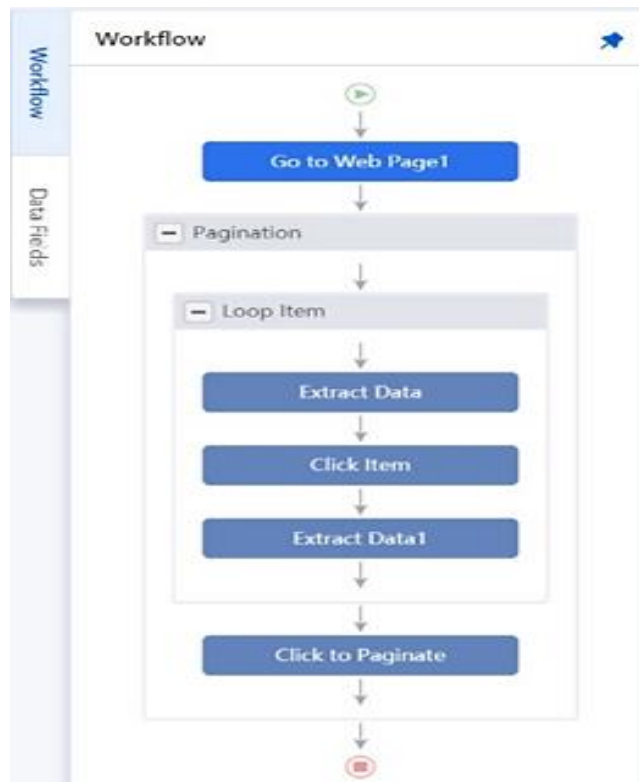


Figure 3: Workflow of Data Extraction of LisLinks

Once the web was crawled, data was collected into the XML file. The total ten thousand

four hundred twenty posts with a total of 48,275 responses have been retrieved in Table 1. We extracted five responses from each post to make the data more understandable and organized. Each post is assumed to have at least five responses. Now, rounding up, the total number of responses becomes 52,100. The data set has responses ranging from 0 to 5.

Table 1

Web Scraping result of LisLinks.com using Octoparse

S.No	Post Title	Replies	Reply1	Reply2	Reply3	Reply4	Reply5	Views
	Hello World.	2	Very glad to know...	Madam ji...	?	?	?	248
	Installing Koha...	10	The pacakage...	Thanks for...	If you can give...	Ubuntu 20.04...	I tried also...	455
	How to...	1	Thank you sir...	?	?	?	?	287
	Requesting...	4	Dear Mr.	Respected sir...	This link is...	Dear sir...	?	566
	Application the...	3	Firstly, you	You just try...	Respected...	?	?	565
	An online...	0	?	?	?	?	?	136
	For print journal...	4	Dear Brother...	UGC approved...	Hello madam...	UGC approved...	?	279
	Swayam portal...	3	Time...	Please support...	Go to webste...	?	?	835
	Latest Live...	2	The latest live...	Sir...	?	?	?	635
	Annual increment...	1	Hey you can...	?	?	?	?	432
	Call for paper...	1	Sir...	?	?	?	?	639
	How to create...	4	IRINS is web...	How these...	Thank you...	Very first institute...	?	583
	NDL member...	7	For NDL...	How to become...	Please click...	After submission...	?	791
	Scope of ...	3	This scope..	Regular mod...	I would support...	?	Send me your...	372
	Free library...	2	Kindly visit...	Thanks sir...	?	?	?	685
	Data from...	1	For resolve...	?	?	?	?	109
	AICTE norms...	1	How we can...	?	?	?	?	135
	How to prepare...	7	So many...	Dear Devendra...	Thanks sir...	You can prepare...	?	5144
	Call for books...	0	?	Thank you...	?	?	You can prepare...	368
	Useful videos...	9	Very useful with	?	Useful...	Thank you...	?	765

Data Pre-processing

Data pre-processing is performed to ensure the data is readable. We derived raw data from replies. It includes extraneous data that is not appropriate for the study. Before performing

sentiment analysis on extracted replies to datasets, we must delete redundant data such as HTML connections, punctuation, and stop words. Rapid miner, an open-source data mining application, is used to perform these operations. It integrates data planning, machine learning, deep learning, text mining, and predictive analytics into a single environment. For sentiment extraction, the "Vader algorithm" has been used. VADER is a Valence Aware Dictionary and sentiment Reasoner (VADER) application. VADER is a vocabulary and library for sentiment analysis based on rules. Even if it does not need negative and positive text samples to train a classifier or develop code to search for terms in a sentiment lexicon, VADER allows it to get up and run quickly with sentiment classification performance. Compared to other Machine Learning and Deep Learning methodologies, VADER is also computationally efficient. Because the VADER method has a built-in training library, There is no need to train our model individually (Yadav, Raskar, Waman& Chaudhari, 2019). Figure 4 illustrates the data pre-processing and sentiment analysis model captured from the RapidMiner tool.

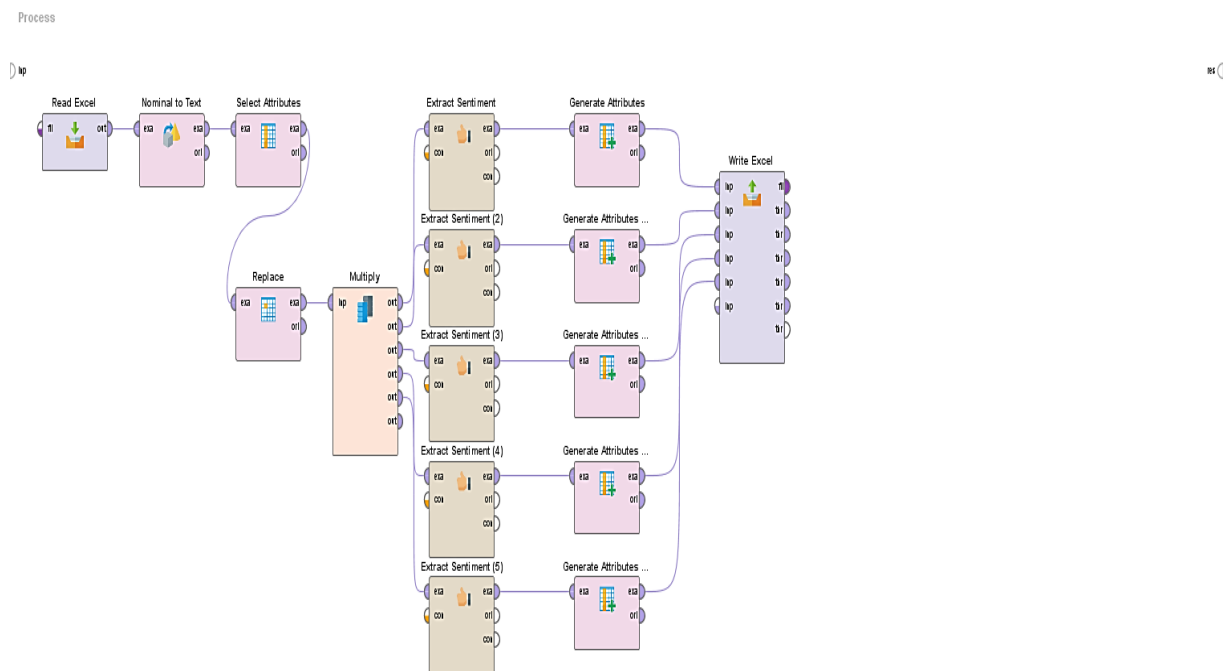


Figure 4: Model for Data Pre-processing and Sentiment Analysis of LisLinks Discussion Forum

Steps in Data Pre-Processing

- I. Data extraction using Octoparse, a web scraping service, and exporting as a Microsoft excel file.
- II. Import of Excel file in RapidMiner for processing.
- III. Conversion of Excel file into text format for further operations.
- IV. Selection of required attributes for sentiment analysis.
- V. Removing of unnecessary data using remove operator, with RegEx string `https?://[-a-zA-Z0-9+&@#/%?~_!:,;]*[-a-zA-Z0-9+&@#/%~_]`
- VI. Defining and selecting the attributes for sentiment analysis. We categorized replies under five excel columns as Reply1, Reply2, Reply3, Reply4, and Reply5.
- VII. Sentiment extraction through the 'Extract Sentiment' operator. The sentiment was extracted using the 'Vader' method for all supplied characteristics.

VIII. Creation of a property for each reply (Sentiment1 for Reply1, Sentiment2 for Reply2, Sentiment3 for Reply3, Sentiment4 for Reply4, and Sentiment5 for Reply5) to indicate the text's polarity; positive, negative, and neutral feelings. Here is a logical phrase for determining the text's positive, negative, and neutral polarity.

if (Score > 0, "positive", if (Score < 0, "negative", if (Score == 0, "neutral", "")))

IX. 'Write Excel' operator for getting results and further analysis on an Excel sheet.

Results

Once the output was obtained, each text was assigned a separate score. The positive score indicates that the text has a positive sentiment, the negative score shows that the text has a negative emotion, and the null value shows that the text has a neutral sentiment. Six additional attributes were added: *the Score, the Scoring string, the Negativity, the Positivity, the Uncovered Tokens, and the Total Tokens*. The score attribute indicates the sentimental worth, which is determined by the difference between the Positivity and Negativity attributes. A number less than zero indicates Negative Sentiment, a score greater than zero indicates Positive Sentiment, and a score equal to zero indicates Neutral Sentiment. Out of 52,100 replies, the Vader algorithm detects 33,121 Neutral replies, 16,784 positive replies, and 2,195 negative replies.

Figure 5 shows the comparative frequency distribution of sentiment polarity for all the replies. It helps to examine the polarity pattern of responses from Reply 1 to Reply 5. Reply1, i.e., The first reply of each post, gets the highest count of positive sentiment. While next, all the replies get more neutral counts than positive and negative. The neutral polarity of responses is growing, while the positive and negative polarities are declining.

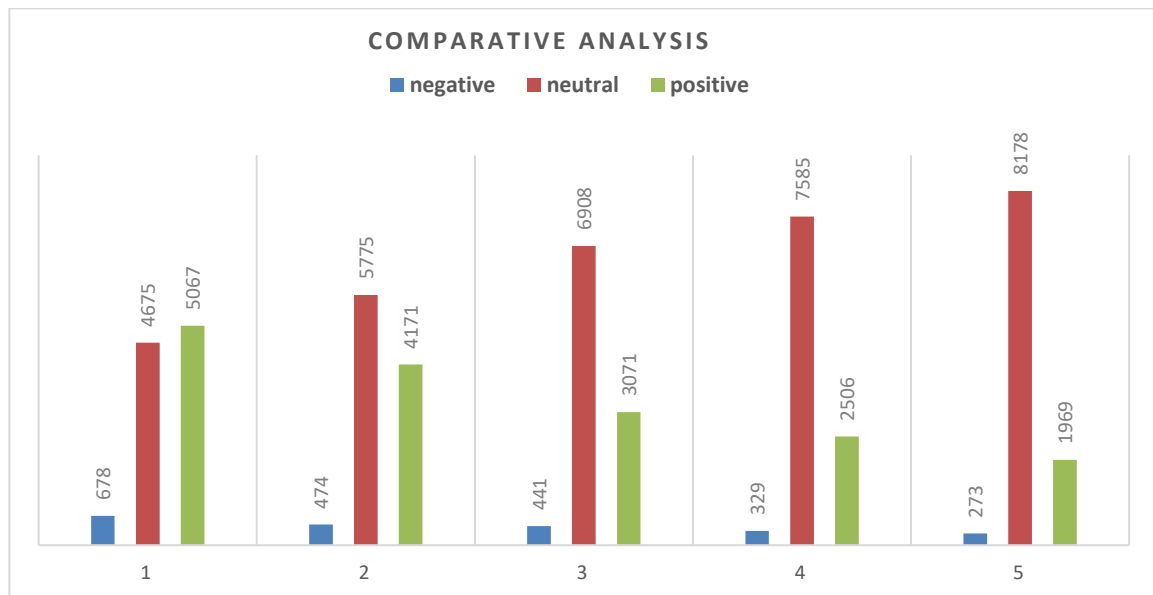


Figure 5: Comparative Analysis of Sentiment Count of Reply1, Reply2, Reply3, Reply4, and Reply5

Figure 6 displays the percentage of the polarity of sentiments of all five replies in the extracted dataset of the LisLinks forum. Based on the results shown in Figure 6, most of the sentiment of replies in our datasets were neutral. However, interestingly, 4% of the replies expressed negative sentiments. 64% of the replies expressed neutral sentiments, and 32% expressed positive sentiments. The neutral percentage is the highest percentage among all other

classes.

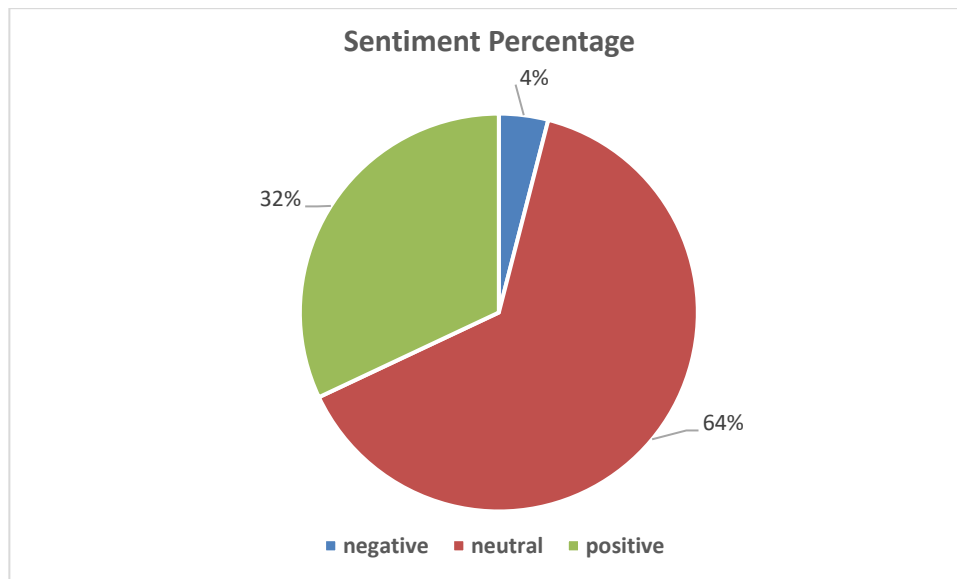


Figure 6: Sentiment percentage

Discussion

Discussion forums have become a knowledge reservoir due to the quick growth in their use. Examining this information can aid in our understanding of the community's attitudes about a certain issue. However, how to collect and analyze this unstructured kind of data is the most difficult part of assessing these repositories. A framework for extracting data from LisLinks Forums and analyzing it in R is presented in the paper by Garg & Kanjilal (2019). However, adopting the framework is problematic because it necessitates understanding the R programming environment and is difficult for most individuals to grasp. In this study, we found the most accessible tool, 'The Octoparse tool,' to extract the unstructured data from the LisLinks discussion forum. We also provided a straightforward drag-and-drop way to extract data from discussion forums.

The study's finding reveals that 64% of sentiment polarities are neutral. However, only 32% of sentiment polarities are positive, and 4% are negative. We found that the neutral sentiment percentage is higher than positive and negative sentiment. The most significant reason is that members are not actively participating in the discussion forum. The same reason is addressed by Garg & Kanjilal (2021) that the interest of the members of LisLinks is declining. The members are losing interest in participating in the forum discussion. Alsaqer and Sasi (2017) evaluated the sentiments of movie review posts using the Rapid Miner tool. Our study followed the same tool to build a model to evaluate the members' sentiments in the discussion forum.

The study by Vera, Suárez and Lopera (2018) analyzed the sentiments of Twitter posts on the conflict in Columbia, and the results found that foreigners have more positive feelings than Columbian tweets. Our study unequivocally shows that LisLinks forum members are satisfied. However, the more neutral sentiments revealed a waning level of members' interest. The interaction among the members in this forum is declining. The same findings were made by Garg & Kanjilal (2019) in their qualitative examination of the LisLinks forum, where they concluded that the number of discussions on LisLinks has decreased. "Despite the decrease in the number of posts, some of the discussions have been viewed by more than twice the number

of registered members".

A comparative analysis of the sentiment count of each five replies indicates that the neutral polarity of replies is growing while the positive and negative polarities are declining. The primary reason for this discrepancy is that this forum has few members willing to respond to the inquiry. That is why, from Reply1 to Reply5, neutral polarity increases. This increment shows the chances of getting more replies to each post are very low.

Conclusion

In this paper, we attempted to present the sentiments of the members of the LisLinks discussion forum using the lexicon-based approach with the help of the Rapid Miner tool. This study calculated the sentiment score and counted it as the number of positive, negative, and neutral sentiment replies—all the objectives we framed for focusing upon the forum members. The findings show that neutral sentiment is the most prevalent, while negative sentiment is the least prevalent. The more neutral sentiment indicates the lower interest of the members in the discussion forum. Members are not active to take participating in the discussion forum. This study will aid LisLinks' administrator in grasping users' viewpoints, feelings, and opinions. As a result, they would enhance the quality of the discussion forum. This study is limited to only the lexicon-based approach of sentiment analysis. Future studies may be carried out with various approaches to sentiment analysis.

References

- Alsager, A. F. & Sasi, S. (2017, July). Movie review summarization and sentiment analysis using rapidminer. In *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)* (pp. 329-335). IEEE.
- Bhonde, R., Bhagwat, B., Ingulkar, S. & Pande, A. (2015). Sentiment analysis based on dictionary approach. *International Journal of Emerging Engineering Research and Technology*, 3(1), 51-55. Retrieved from <https://www.ijeert.org/pdf/v3-i1/9.pdf>
- Collomb, A. A., Joyeux, D., Hasan, O. & Brunie, L. (2014). A study and comparison of sentiment analysis methods for reputation evaluation. *Rapport de Recherche RR-LIRIS-2014-002*, 1–10. Retrieved from <https://liris.cnrs.fr/Documents/Liris-6508.pdf>
- Elbagir, S. & Yang, J. (2019, March). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 122, p. 16).
- Garg, M. & Kanjilal, U. (2019). A framework to process text data of web discussion forums: A study of LIS links. *DESIDOC Journal of Library and Information Technology*, 39(6), 315–321. <https://doi.org/10.14429/djlit.39.06.15145>
- Garg, M., & Kanjilal, U. (2021). A quantitative analysis of lis links discussion forum. *Annals of Library and Information Studies*, 68(1), 75–81. <http://op.niscair.res.in/index.php/ALIS/article/view/40396>
- Liu, B. (2010). Sentiment analysis and subjectivity. In N. Indurkha and F. J. Damerau (Eds) *Handbook of Natural Language Processing, Second Edition*, (pp.627–666).
- Park, S. & Woo, J. (2019). Gender classification using sentiment analysis and deep learning in a health web forum. *Applied Sciences*, 9(6), 1249. <https://doi.org/10.3390/app9061249>

- 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). <https://doi.org/10.14429/djlit.34.1.5942>
- Saini, S., Punhani, R., Bathla, R., & Shukla, V. K. (2019, April). Sentiment analysis on twitter data using R. In *2019 International Conference on Automation, Computational and Technology Management (ICACTM)* (pp. 68-72). IEEE. <https://doi.org/10.1109/ICACTM.2019.8776685>
- Shaveta. (2016). Library and information links: A great platform for lis professionals. *International Journal of Library and Information Studies*, 6(3), 25-30. Retrieved from <https://www.ijlis.org/articles/library-and-information-links-a-great-platform-for-lis-professionals.pdf>
- Vera, V. D. G., Suárez, L. M. M., & Lopera, I. C. P. (2018). Sentiment analysis on post conflict in Colombia: A text mining approach. *Asian Journal of Applied Sciences*, 6(2), 53-59. [file:///C:/Users/drghane/Downloads/admin,+1_Sentiment+Analysis+on+Post+conflict+in+Colombia+A+Text+Mining%20\(3\).pdf](file:///C:/Users/drghane/Downloads/admin,+1_Sentiment+Analysis+on+Post+conflict+in+Colombia+A+Text+Mining%20(3).pdf)
- Verma, P., Khanday, A. M. U. D., Rabani, S. T., Mir, M. H. & Jamwal, S. (2019). *International Journal of Recent Technology and Engineering*, 8(3), 8338–8341. <https://doi.org/10.35940/ijrte.C6612.098319>
- Yadav, M., Raskar, S., Waman, V. & Chaudhari, S. B. (2019). *Sentimental analysis on audio and video using vader algorithm. International Research Journal of Engineering and Technology (IRJET)*, 6(6), 190-195. Retrieved from <https://www.irjet.net/archives/V6/i6/IRJET-V6I641.pdf>
- Younis, E. M. G. (2015). Sentiment analysis and text mining for social media microblogs using open source tools: An empirical study. *International Journal of Computer Applications*, 112(5), 44-48. <https://doi.org/10.5120/19665-1366>