

# News Coverage of Libraries in India: An NLP-based Analysis of News Indexed in Google

**Sourav Mazumder**

Research Scholar, Department of Library and Information Science, University of North Bengal, West Bengal, India

E-mail: [smazumderlis91@gmail.com](mailto:smazumderlis91@gmail.com)

**Dr. Tapan Barui**

Asst. Professor, Department of Library and Information Science, University of North Bengal, West Bengal, India

E-mail: [tapanbarui@nbu.ac.in](mailto:tapanbarui@nbu.ac.in)

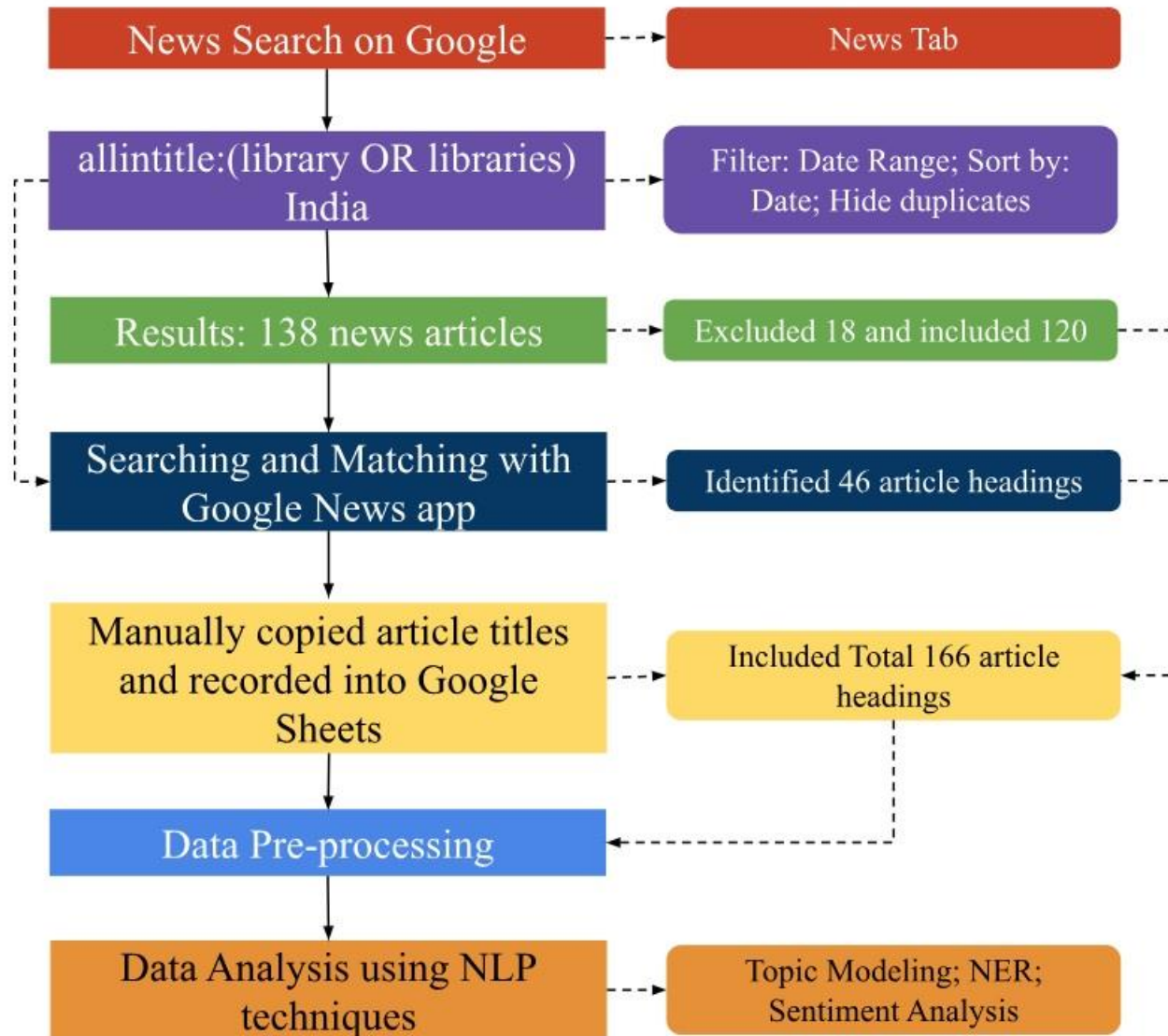
# Introduction

- Libraries play an important role in society for **creating and disseminating knowledge** by providing access to books, digital books, and educational programs.
- Scientific outputs showcase **libraries' activities and innovation**, while **news media** provides information on **library news and initiatives**.
- Online news can be accessed through smartphones, tablets, and computers via the internet.
- Cui and Liu (2017) mentioned three types of online news media: **legacy, explanatory, and citizen news**.
- Google Search engine dominates the market with **84.20 billion monthly visits** and **91.74%** worldwide market share (Online search market worldwide - Statistics & Facts, 2024)
- Google News aggregates news from various sources. In March 2024, it received **375.60 million visits**, primarily from the USA.
- This study uses **Natural Language Processing (NLP)** to analyze news trends, topics, and sentiments from India's library news coverage from **April 2023 to March 2024**,

# Research Objectives

- To analyse the **trend** of news coverage about libraries in India from April 2023 to March 2024
- To **identify topics** discovered through news titles
- To explore the **entities** that exist in the news titles
- To measure the **sentiment** expressed in the news titles

# Methods



- **Data** was collected on 01/04/2024
- **Dataset** pre-processing: removal of stopwords, punctuations, special characters, and URLs.
- **Topic modelling** using BERTopic
- **NER** using spaCy
- **Sentiment analysis** using NLTK's Sentiment IntensityAnalyzer with VADER lexicon.
- **Scores includes:** Compound Score, positive ( $CS > 0$ ), neutral ( $CS \approx 0$ ) and negative ( $CS < 0$ )
- Data analyses and visualizations are created using Python libraries like Pandas and Matplotlib.

Figure 1: Schematic Representation of the Implemented Methods for Analysis

# Results

- *Trend of News Coverage on Libraries in India*

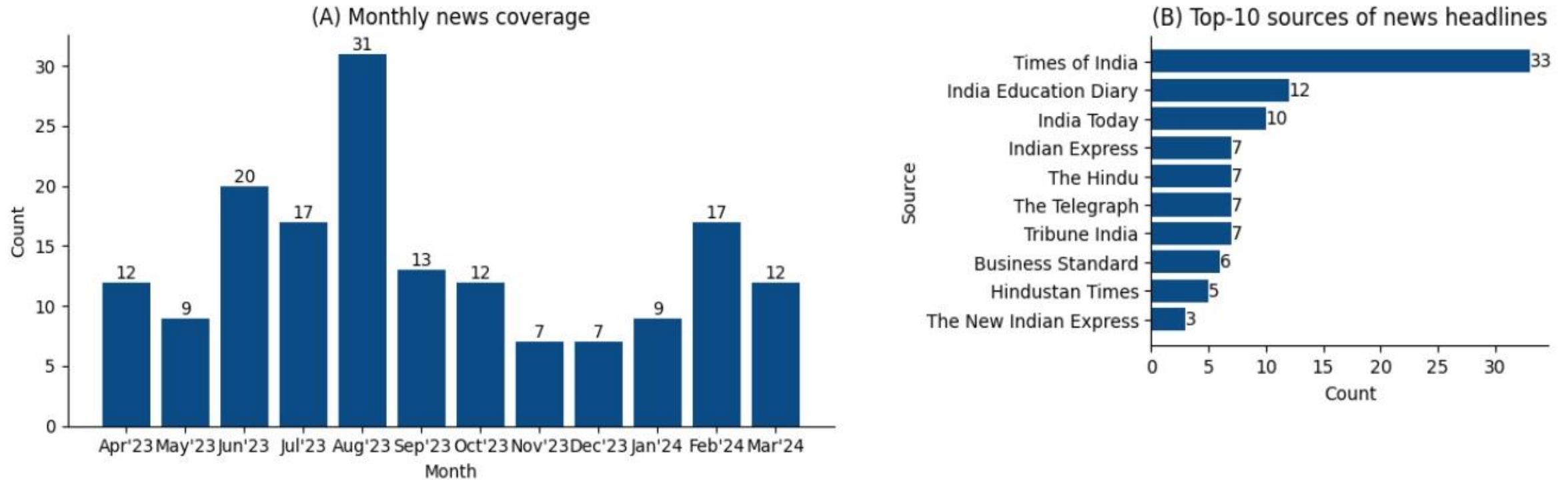


Figure 2 (A): Monthly News Coverage on Libraries and (B) Top-10 Sources of News

# Continued...

- *Topic Modelling*

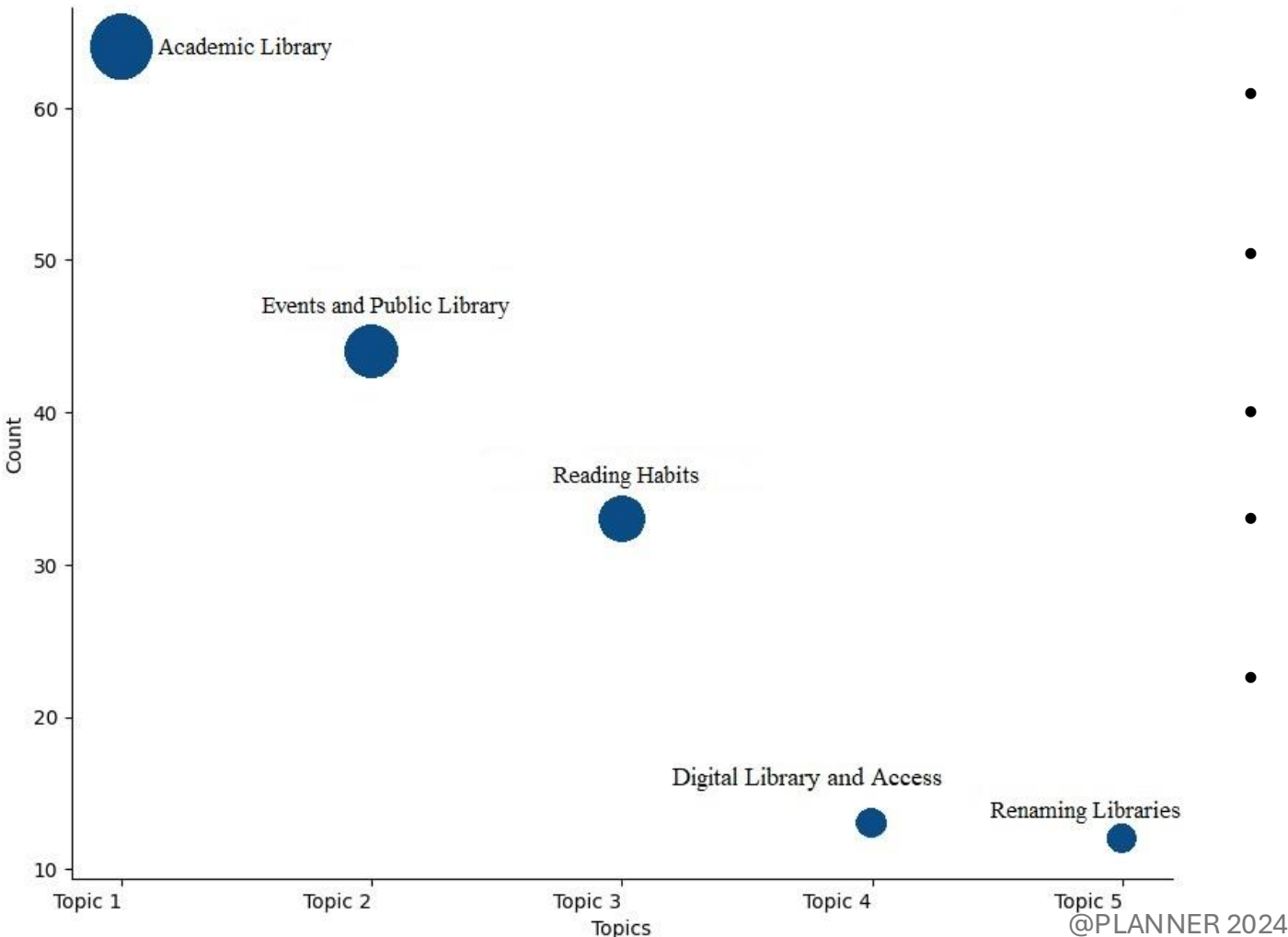


Figure 3: Five Topics Assigned to Topic Numbers

- **Academic Library:** ‘university’, ‘public’, ‘director’, ‘college’, ‘ranchi’, ‘students’, ‘first’, ‘ever’, ‘collapses’, and ‘classical’
- **Events and Public Library:** ‘inaugurates’, ‘hyderabad’, ‘delhi’, ‘list’, ‘biggest’, ‘memory’, ‘chennai’, ‘mansa’, ‘national’, and ‘public’
- **Reading Habit:** ‘book’, ‘beautiful’, ‘reading’, ‘world’, ‘public’, ‘biggest’, ‘digital’, ‘people’, ‘visit’, and ‘old’
- **Digital Library and Access:** ‘digital’, ‘awareness’, ‘indian’, ‘national’, ‘students’, ‘first’, ‘treasure’, ‘subjects’, ‘nlsiu’, and ‘underserved’
- **Renaming Libraries:** ‘museum’, ‘memorial’, ‘prime’, ‘renamed’, ‘ministers’, ‘society’, ‘name’, ‘dropped’, ‘government’, and ‘renames’

# Continued...

- *NER Analysis*

IIT Kharagpur Creates National Digital Library **ORG** ; Treasure Trove Of Knowledge For Students In All Subjects

Public libraries in **India GPE** : Most of the country's public libraries in the **3 CARDINAL** southern states! How many numbers is **Bengali NORP** in the list?

KK Handiqui Library **ORG** of **Gauhati University ORG** declared best in **India GPE**

Pankaj Tripathi **PERSON** inaugurates library in hometown in memory of late Father

**86 CARDINAL** -yr-old who transformed his residence into public library in **Kerala GPE** felicitated with Padma Shri award

Figure 4: Examples of Entities Identified in News Titles

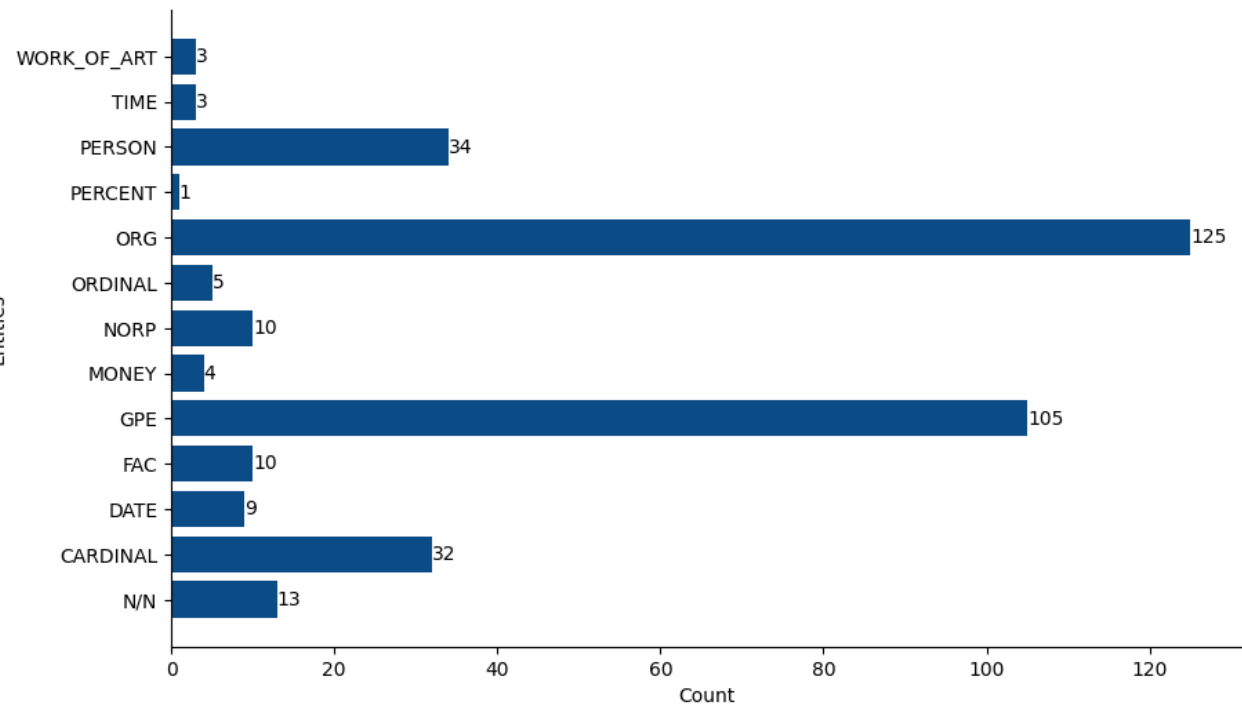


Figure 5: Counts of Entities Identified in News Titles

- 'ORG' (organisations, companies, agencies) is marked in **cyan**
- 'GPE' (Geo-Political Entities) is highlighted in **yellow**, while 'NORP' (Nationalities or Religious or Political Groups) is highlighted in **purple**
- 'PERSON' entities are shown in **light purple**, and 'CARDINAL' (numerals) in **grey**

- **13** out of 18 entity types were found in 153 news titles
- The most frequent entity type was '**ORG**' followed by '**GPE**'
- Other frequent entities include '**PERSON**,' '**CARDINAL**,' '**NORP**,' and '**FAC**,' 'DATE' (n=9), 'ORDINAL', 'MONEY', 'TIME', and 'PERCENT'

# Continued...

- *Sentiment Analysis*

**Table 1: Examples of Text along with Sentiment Score**

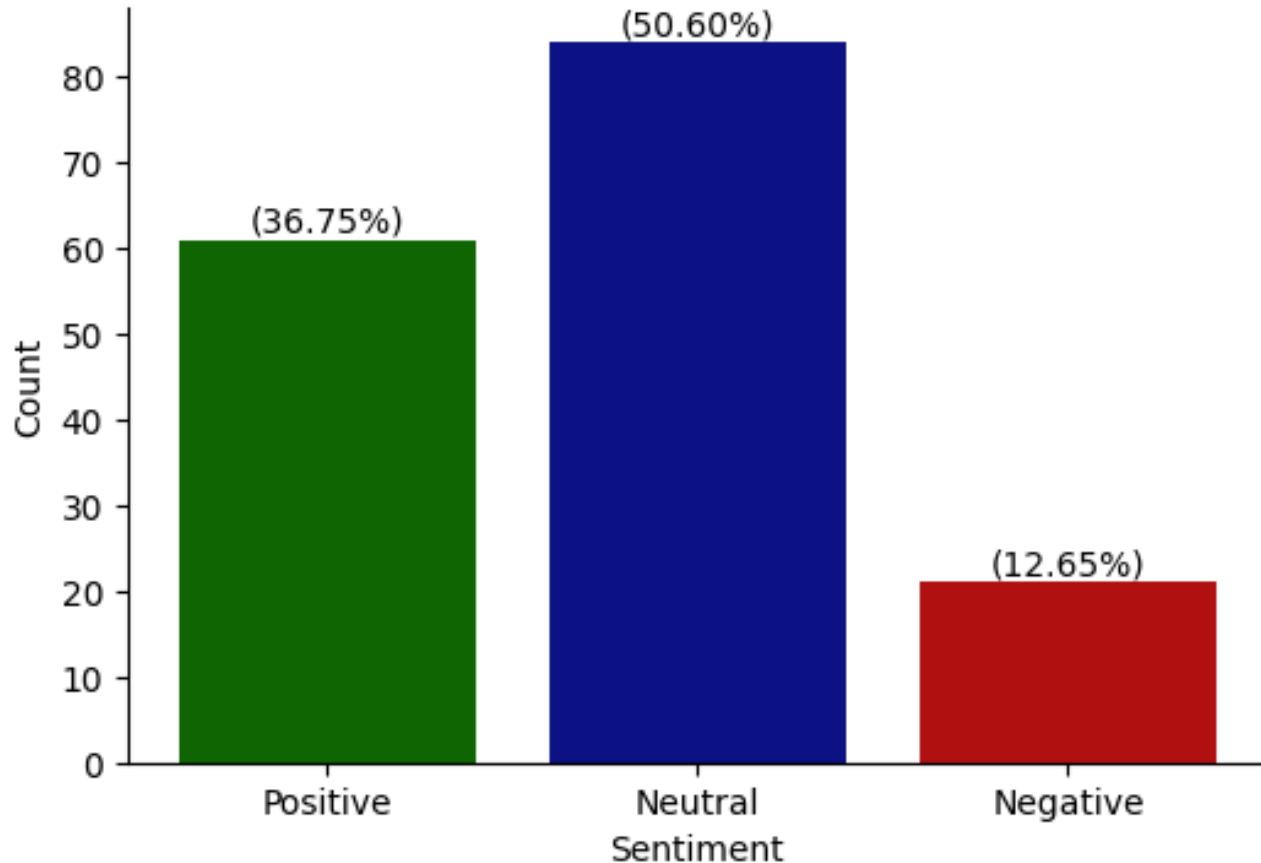
Title	Scores				
	Sentiment	Compound	Positive	Negative	Neutral
Visits to public libraries increase	Positive	0.3182	0.535	0	0.465
Dispute rocks Murty Classical Library of India; editorial board dismissed without explanation	Negative	-0.4019	0	0.278	0.722
Library to accept old book donation: The library at Swapnabhor, the senior citizens' club, is building on its stock with donated books	Positive	0.3818	0.191	0	0.809
Connecting People With Books: The Rise of Free Libraries In India	Positive	0.5106	0.452	0	0.548
Talk at the Library: C. G. Jung and India: A Conversation with Cultural Historian Sulagna Sengupta	Neutral	0	0	0	1
Bangalore man arrested for stealing rare, century-old book from Nilgiri library	Negative	-0.7783	0	0.493	0.507
Why libraries are liberating spaces	Neutral	0	0	0	1

**Table-1** provides a few titles along with their corresponding sentiment scores. It offers a clear understanding of the emotion of a text. Based on the **compound score, the levels of positive, neutral, and negative** can be aggregated for an informative outcome



# Continued...

- *Sentiment Analysis*



- The majority of the sentiments of the news titles are **neutral (n=84, 50.60%)**.
- **Positive sentiments** were expressed in **61 (36.76%)** titles.
- A significant portion of the news emotional tone is **either neutral or pointing a positive stance** on libraries. On the other hand, a total of **negative sentiments** were discovered in **21 (12.65%)** news titles.

Figure 6: Distribution of Sentiments Expressed in the News Titles

# Conclusion

- **August 2023** received the highest media attention.
- **The Times of India** was the top source for library-related news.
- Most prominent topics were ‘**Academic Library,**’ ‘**Events and Public Library,**’ and ‘**Reading Habit.**’
- ‘**ORG**’ and ‘**GPE**’ were the top NER categories.
- Sentiment analysis showed a dominant **neutral sentiment**, with a **significant presence of positive sentiment** and **fewer negative sentiments**.
- Library stakeholders can use the findings to **focus on community engagement**.
- Media professionals can highlight the **interests and issues related to libraries**.
- **In-depth analysis** over a longer period to observe the **evolution of media coverage**.

Key Outcomes

Implications

# Continued...



**Only one year** was taken as a time frame using a **couple of search keywords**, which might restrict the comprehensiveness of the search results related to library news in India.



Search results might be changed based on **ranking over time**, and some entities were **not recognized correctly**.

Limitations



A more **compact and fine-tuned NER** approach could tackle this situation and reduce bias over entity extraction.



The results provide significant outcomes regarding media coverage of libraries in India and can **guide future research prospects for advancing the field** using more titles, including full-text articles.

Future Research

# References

Cui, X., & Liu, Y. (2017). How does online news curate linked sources? A content analysis of three online news media.

Journalism, 18(7), 852–870. <https://doi.org/10.1177/1464884916663621>

Online search market worldwide - Statistics & Facts. (2024, March 8). [https://www.statista.com/topics/](https://www.statista.com/topics/1710/search-engine-usage/#topicOverview)

[1710/search-engine-usage/#topicOverview](https://www.statista.com/topics/1710/search-engine-usage/#topicOverview)

# Thank You