Integrating Natural Language Processing (NLP) with Existing Library Framework in Enhancing Level of Users' Satisfaction

Rajesh Chutia , Mukut Sarmah, Mousum Handique & Jose Rodolfo Hernandez-Carrion

Rajesh Chutia
Research Scholar (Ph.D
Dept of Lib and Inf Science
Assam University
rajesh.chutia@aus.ac.in

Dr Mukut Sarmah
Associate Professor
Dept of Lib & Inf Science
Assam University
drmsarmah@gmail.com

Dr Mousum Handique
AssistantProfessor
Dept of Computer Sc. & Engg
Assam University
mousum.smit@gmail.com

Dr Jose Rodolfo Hernandez-Carrion
Professor
Dept. of Applied Economics,
University of Valencia, Spain
rodolfo.hernandez@uv.es

ABSTRACT:

Introduction Libraries are the storehouse of information instilling knowledge into human minds. The easy availability of information from online sources has barred users from visiting the library physically. To keep the users intact within the physical confines of a library, there is a need for libraries to evolve through the implementation of state-of-the-art techniques.

Purpose- The purpose of this study is to develop a theoretical framework and conceptual background for applying natural language processing (NLP) in library and information science (LIS) to improve library services and related research. This study also aims to identify the LIS areas where NLP can use and implement one such area to provide efficient services.

Research Problem-This study integrates natural language processing with the existing library framework to enhance the satisfaction level of the users.

The objective-The primary objective of this study is to amalgamate natural language processing in library libraries by identifying the areas of improvement.

Methodology- This study uses a conceptual paradigm by systematic observation of NLP facts in accentuating library services. This study aims to make libraries more efficient in providing dedicated resources and services to users by implementing NLP techniques. The data for this study were collected from secondary sources and analyzed to obtain meaningful insights.

Findings- Considering the staggering amount of informative data that a library has to deal with, NLP can be brought to mainstream libraries to develop numerous models to provide better services to its users. NLP not only supplements library services but also helps make research more inclusive of applications such as keyword extraction, translation, and summarization. The library is a service-oriented organization, and to provide better services, NLP should be taught and implemented in information science education.

KEYWORDS:

.

Introduction

The origins of libraries differ depending on the literary evidence found in various geographical locations. The invention of the earliest literature dates back to the Mesopotamian time, and libraries emerged alongside the invention of writing. Libraries, considered storehouses, gradually developed into information centres over time. The library remained accessible only to privileged and upper-class citizens in the medieval period, which is, to kings and ministers. Today, these dynamics have drastically widened the use of library services for the masses. Simultaneously, the roles of libraries and librarians' have gradually changed from record-keeping and classification to the implementation of information communication technologies. With this changing scenario, development is necessary to cope with the convergence of technology with libraries. Such interdependence between digitalization and libraries is evident in the information retrieval behaviour of people. The implementation of information communication technologies in a library is just a nascent thing, where computers have evolved from an abacus to artificial intelligence and robotics.

On the one hand, many libraries are yet to be automated, while on the other hand, artificial intelligence has seized the world by making peoples' lives easier. Libraries that do not implement any information communication applications are no longer regarded as libraries. This study considers the higher stage of ICT, where natural language processing bolsters libraries to extend their functions significantly. Natural language processing makes computers friendlier to humans so that they can communicate with computers as they communicate with one another. For effective implementation of Natural Language Processing, the library should be well equipped with computer facilities and competent library staff, both well-versed and skilled in natural language processing.

Objectives

- To amalgamate Natural Language Processing in an ICT-based framework for Libraries.
- To discuss areas that can be improved and enhanced using natural language processing applications.
- To create a user-friendly library environment where users can enjoy library facilities with minimal human intervention.

Literature Review

The literature review discusses libraries from their inception and evolution. It then proceeds towards implementing ICT and uses natural language processing in libraries. Zarifeh (2000) corroborated the origin of writing with the emergence of Mesopotamian cities. According to him, the concept of libraries has begun to take shape. Taskin and Al (2019) supported natural language processing in making tasks easier in library and information science. They examined 6,607 publications to advance their research. Jeevitha and Kavitha (2019) opined that integrating natural language processing with existing library services increases the efficiency of librarians.

Methodology

Research Method: A conceptual-based research method was used in this study. The idea behind this research is solely to make libraries more efficient in providing dedicated resources and services to their users.

Data Collection: The data for this study were mainly collected from secondary sources. Journal articles, visual data, websites, blogs, research papers, critical observation reports from experts, newspaper columns, magazines, and other relevant sources were also collected.

Natural Language Processing

Natural language processing is a branch of artificial intelligence that allows computers to understand human speech and text as human beings write and speak naturally. Humans' ability to speak, write, and understand establishes communication because they can share and gather information. Likewise, computer language has evolved from 0 and 1 to a high-level language, enabling computers to interact with humans as they speak and write.

Initially, computers were used to deal with low-level languages, also known as machine languages (0 and 1). The problem with machine language is that computers only receive and deliver input and output in machine language, that is, 0 and 1. The user must decode the binary digits into meaningful results and code the input into the machine language. The coding and decoding of machine language require a programmer, and it is not easy for normal people to use computers. Later, computer language evolved into an assembly language, where special characters could also be used as input and output. This helped the computer users a bit where they could use mnemonics in the input and output processes. The revolution in computer language has been brought about by high-level languages, similar to human language. High-level language enables computers and humans to interact without any difficulty.

Natural language is a subfield of artificial intelligence that makes life easier by enabling users to communicate with computers as they communicate with their peers. Its ability to understand human language has helped develop applications that benefit human beings.

a. Levels of Natural Language Processing

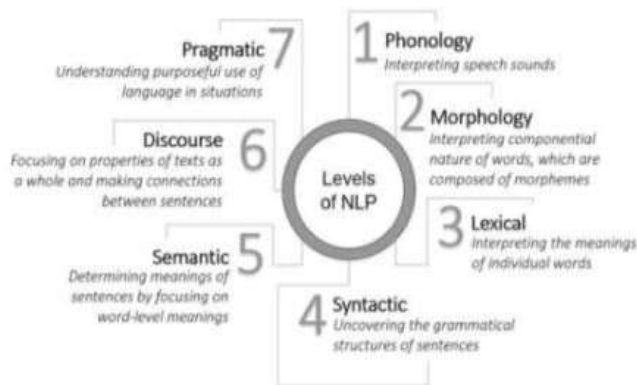Feldman (1999) and Liddy (2010) divided NLPlevels (Figure 2) in the following manner:

Figure 1 Levels of NLP
Source: Feldman 1999 (p. 62-64) and Liddy 2010 (p. 3867-3868)

Application of Natural Language Processing in Library and Information Science

Following are the ways where the application of natural language processing can be used in library and information science:

a. Sentiment Analysis:

People's behaviour is not always based on their inborn and instinctive tendencies. Of course, such innate dispositions, along with the social environment, play a strong role in the formation of certain repetitive actions. These are more stable, as they habituate over time and become integral aspects of personality, known as sentiments. A complex of experiences garnered over time, prompted by feelings, thoughts, and emotions, becomes sentiment. Thus, sentiments are acquired, but deeper dispositions involve several affective and conative factors. Sentiments are attitudes and judgments that constitute a permanent part of our mental structures. These are directed towards certain objects, things, ideas, or concepts, and are usually expressed through emotions.

Sentiment analysis is a mechanism that deals with people's sentiments by finding context out of an unstructured text and also by segregating them into different categories. It is also known as emotional artificial intelligence, in which emotions are analyzed. This is how computers are programmed to understand human emotions. Sentiment analysis or opinion analysis uses natural language processing to analyze and identify the exact data. Sentiment analysis is commonly used for messaging applications. They are also used in customer reviews of goods and products, social media, and survey responses.

a1. Sentiment Analysis in Library and Information Science:

The application of sentiment analysis can provide a different dimension to the library and information science. The library is a service-oriented user-based organization. The libraries are based on satisfying user needs. To this end, the library collects feedback from users to improve their services.

Areas of implementation
- Online Feedback form
- Student Peer Feedback
- Student self-Feedback

b. Chatbot

Application software is used to replace human interventions in audio and text chats. It uses pre-recorded videos or text messages to reply to customers. It is mainly used in flight books, travel books, and telecom applications. It uses natural language processing by matching text to answer queries.

b1. Implementation in Libraries

Chatbot applications can be used in libraries to satisfy user requirements. It can be implemented in reference services by posing questions to the librarians. For instance, users connect with reference librarians during times of difficulty in the operation of services. Usually, users communicate with a reference librarian by sending messages through email, WhatsApp, or text messages. The librarian often replies to all queries, but after all, being a human being comes with limited accessibility. It is challenging for him to stay connected every time, while it is quite impossible to employ a dedicated staff member such as a customer care executive to solve users' queries. Therefore, having a chatbot mechanism in the library maximizes its efficiency. The librarian should maintain a dataset of frequently asked questions to make the application user-friendly.

b1.1. Chatbot Script:

The imaginary scenario mentioned by Jeevitha and Kavitha (2019)

User: "Hey, can you help me out?"
Chatbot: "Yes, Sure; please tell me what you need?"
User: "Is the book Selfish Gene available."
Chatbot: "Please wait. I will come back to you."
Chatbot: "Do you mean The Selfish Gene by Richard Dawkins?"
User: "Yes"
Chatbot: "We have it. Presently it is out on lending and will be returned by the 30th of this month, should the current user choose not to renew it."
User: "Thanks for the response."
Chatbot: "Thank You"

c. Speech Recognition

Speech recognition is an application software whose input and output are loudly spoken. Users provide input in voice messages, and the application answers invoice messages in the same way. Speech-recognition technologies have revolutionized the world and interact with their devices. Commanding devices has become easier because humans can input devices as they interact with other humans. Instead of providing input in text, humans can provide an input invoice form. Currently, voice recognition technology is used in health, education, customer service, disability assistance, emotion recognition, and hands-free communication. Examples of speech recognition technology include Google Assistant, Siri, Alexa, and Cortana.

c1. Implementation in Libraries

It is also an application of NLP that can be implemented in libraries. It is similar to a chatbot, but its only difference is that it uses voice instead of text. Voice recognition can replace reference librarians. If the voice recognition application is enabled with all the information in the library, it can entirely replace the reference librarian and satisfy library users, thus working as a library assistant.

d. Machine Translation

Machine translation is related to the automatic translation of languages. In retro times translation could only be done through the help of a dictionary. However, owing to the change in time and upgradation in the computer, a language translation model can be developed with the help of natural language processing. It uses a bilingual dataset and other language assets (grammar, syntax (sentence structure), semantics (meanings)) to build language and phrases to translate text.

d. Some other example applications

a) **Natural language interfaces to databases:** This makes information retrieval easier where users can communicate in natural language, and the result will be in natural language.

b) **Natural language interface to computers:** In the case of a problem, users can seek online assistance. For example, "*UNIX consultant*."

c) **Question answering systems:** LUNAR by Woods (1977) is a research vehicle developed to conduct sophisticated analysis of quantification in natural language sentences.

d) **Story understanding:** Given a story in the specified domain, the question-answering system answers questions about it. For example, work was carried out at Yale under

Schank and Abelson (1977).

e) **Spell Checking:** Spell-checking is a mechanism by which spelling errors can be automatically checked.

f) **Information Extraction:** Important information from a text can be automatically extracted with the help of information extraction, for example, the name of an entity and its recognition.

Process of Natural Language Processing

In some instances, the processes involved in natural language processing also act as steps of natural language processing. The following sections discuss the different processes of natural language processing.

a. Tokenization

Tokenization is the process of breaking complex sentences into smaller pieces. It breaks sentences into small bits, emphasizes every word, and finally provides the output to process further.

b. Word Count

Word count is the process of determining the frequency of words in a sentence. It starts with space-separated words but also includes special characters.

c. Stemming

Stemming is the process of converting words into their root forms. Hence, 'likes,' 'liking,' 'liked,' and 'likely' are reduced to the root word or base word like.

d. Lemmatization

Lemmatization is similar to stemming and normalizing words to their root form. The difference between the two is that stemming removes the last few alphabets from the word, whereas lemmatization considers the meaning and context to convert a word into its root form. Example: lemmatization convert the word 'communicating' to 'communicate' but on the other hand, stemming can only convert the word 'communicating' into 'communicate,' just removing ing from the word which does not have any meaning.

e. Stop Words

Stop words are used to build up a sentence i.e., "the," "a," "at," "for," "above," "on," "is," and "all". Stop words are also known as prepositions that are removed in natural language processing to process the concept further by giving meaning context to it.

f. Part Of Speech Tagging (Pos)

Part-of-speech tagging assigns grammatical elements, which are parts of speech sentences, to form meaningful contexts. The parts of speech are nouns, verbs, pronouns, adverbs, conjunctions, adjectives, and interjections, which are used depending on the context of the sentences.

Why	Not	Tell	Someone	?
Adverb	*Adverb*	*Verb*	*Noun*	*Punctuation mark/sentence closure*

In the above sentence, it can be seen that each word has a lexical term written underneath. However, this becomes cumbersome as the corpus grows. Hence, a short representation represents categories referred to as tags. The following figure shows different types of lexical terms and their tags.

Lexical Term	Tag	Example
Noun	NN	Paris, France, Someone, Kurtis
Verb	VB	work, train, learn, run, skip
Determiner	DT	the, a

Why	not	tell	someone	?
WRB	RB	VB	NN	.

Figure 2 Different types of lexical terms and their tags

a) *POS tagging applications:*
- Name Entity Recognition
- CO-reference Resolution
- Speech Recognition

g. *Named Entity Recognition*
Named entity recognition identifies entities such as personal name, location, monetary value, and event.

h. *Chunking*
Chunking is a process of extracting bits and pieces of information from unstructured text, grouping them into more significant pieces, and helping to identify the constituents (noun groups, verbs, verb groups, etc.)

This process states that an NP chunk should be formed whenever the chunker finds an optional determiner (DT), followed by any number of adjectives (JJ) and then a noun (NN).

Implementation of Natural Language Processing Applications in LIS Research

Research is an on-going process that is needed to improve and advance a particular field. Similarly, research in the library and information science fields plays a vital role in making libraries more efficient and effective. Natural language processing applications can make research more efficient and result-oriented.

Following are the technologies of natural language processing to make research more competent:

a. **Keyword Extraction**
Keywords are the most important words that capture the essence of a research paper. In other words, it can be said that most repeated words other than the stop words can be considered keywords. In the context of an electronic document, keywords are words that help the indexer to make the research paper searchable by interested users. The list of keywords is called the index. The more keywords included in the index, the more searchable the document becomes; however, the author sometimes becomes confused in selecting the appropriate words as keywords. Word count is the process of natural language processing used to count the frequency of words in a sentence. It starts with space-separated words but also includes special characters. The most frequent words, other than stop words, can be considered keywords.

b. *Summarization of Text*
Every researcher must read the exhaustive literature while conducting the research. However, this problem arises with an abundance of available literature. Currently, the use of natural and the problem of text summarization can be solved, which can be used in the literature review.

c. *Data Analysis*
There are many simple to complex application software for data analysis, such as MS Excel, SPSS, and R programming. Similarly, all the above application software can be replaced using the natural language toolkit library of Python to perform statistical calculations, including figures, tables, and graphs.

d. *Spell Checking*
Spell checking is an application of natural language processing that rectifies spelling errors in documents. This can be used to proofread this research paper.

e. *Machine Translation*
Accessibility beyond boundaries has brought goods and services from around the globe to users' doorsteps. Likewise, it has made it possible to access information in research papers and through other means. This information can be of any language; therefore, machine translation can help translate the text into a familiar language.

Conclusion
NLP has brought augmentation to computer languages as it understands the semantics and connotations of human languages in the same way that humans speak and write. This is a component of artificial intelligence that enables text simplification. Today, human society is data-driven, and henceforth NLP has become more imperative, especially when we consider the library as the core of the information science network. Considering the staggering amount of informative data that a library has to deal with, natural language processing can be brought to mainstream libraries to develop numerous models to provide better services to its users. NLP not only supplements library services but also helps make research more inclusive of applications, such as keyword extraction, translation, and summarization. The library is an information-based field. Natural

language processing should be taught and implemented in information science education to make librarians more efficient in delivering effective services to users.

References

Berger, A., Della Pietra, S.A. & Della Pietra, V.J. (1996). A maximum entropy approach to natural language processing. Computational linguistics. 22(1): 39-71.

Feldman, S. (1999). NLP meets the jabberwocky: Natural language processing in information retrieval. Online. 23: 62-72.

Gupta, N. & Mathur, P. (2012). Spell Checking Techniques in NLP: A Survey. International Journal of Advanced Research in Computer Science and Software Engineering. 2(12): 217–221.

Hovy, E. & Lin, C.Y. (1999). Automated text summarization in SUMMARIST. Advances in automatic text summarization. 14: 81-94.

Jeevitha, V. &Kaviatha, E. S. (2019). A Study on Adapting Natural Language Processing for Library Services Delivery. Library Philosophy and Practice: 1-10.

Liddy, E. D. (1998). Enhanced text retrieval using natural language processing. Bulletin of the American Society for Information Science and Technology. 24(4): 14-16.

Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S. &McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In Kristina Toutanova, &Hua Wu (Eds.) Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, Baltimore, Maryland. Retrieved from https://aclanthology.org/P14-5010/

Natarajan, M. (2005). Role of text mining in information extraction and information management. DESIDOC Journal of Library & Information Technology. 25(4): 31–38.

Sarkar, P. &Purkayastha, B.S. (2016) . A Study of the Natural Language Processing Tasks to Address Semantics Ambiguities. International Journal. 6(10).

Taskin, Z. & Al, U. (2019). Natural language processing applications in library and information science. Online Information Review: 1-19.

Zarifeh, M. (2000). Libraries of the ancient world. ANZT

About Authors

Mr. Rajesh Chutia is a Ph.D. scholar at the Department of Library and Information Science, Assam University, and a Librarian at Govt. Model College, Borkhola, Cachar, Assam. He has studied at the prestigious "Jawahar Navodaya Vidyalaya (JNV)." He has a degree in "Bachelor of Business Administration" and a master's degree in LIS from Assam University. He has also completed his M.Phil. and pursuing a Ph.D. from Assam University.

He has published research papers in Scopus Indexed Journals and showcased his research works on international platforms, the recent being the National Taiwan Normal University, Taipei. He has initiated collaborative research projects with scholars from Spain, Taiwan, Japan, Indonesia, and the USA. His research interest encompasses "Library Anxiety, Biblio-Tourism, Natural Language Processing and any ICT application in Library and Information Science."

Dr Mukut Sarmah is working as an Associate Professor (HOD) in Department of Library and Information Science at Assam University. He has participated in various seminars, conferences, and workshops at national and international levels. He has contributed more than 80 research papers in different seminar/conference proceedings, journals, and edited books. He has completed two Research Projects sponsored by UGC. He visited University of Colombo, Sri Lanka for delivering a theme paper in International Research Symposium in 2012. He also visited National Taiwan Normal University, Taiwan for delivering a research paper in ICLIS "International conference on Library and Information Science" His fields of interest are: knowledge organisation, IT applications in libraries, library automation, information retrieval, web applications, bibliometric analysis, blended learning, e-publishing and digital library.

Dr Mousum Handique is working as an Assistant Professor at the Department of Computer Science and Engineering, Assam University, Silchar, Assam, India. He received his Bachelor of Engineering (B.E) in Computer Technology from Nagpur University, in 2002 and Master of Technology (M.Tech) degree from Tezpur University, Tezpur, Assam in 2005. He has received his Ph.D. degree in the year 2020 in the field of VLSI Testing and Reversible Computing from India Institute of Technology Guwahati (IITG), Guwahati 781039, Assam, India. His current research interests include Testing and Synthesis of Reversible and Quantum Circuits, Formal System Verification, Machine Intelligence, Workflow Automation, and Queueing Theory.

Dr Jose Rodolfo Hernandez-Carrion is a Professor at the Department of Applied Economics at the University of Valencia (Spain). He has been teaching Economics since 1989 in different centers and various languages and publishing in international journals with an open and critical multidisciplinary approach. He served in the governing board of Spanish Association of Regional Science (AECR) and Spanish Society of Systems Science (SESGE), where he was Vice-President and Editor in Chief of RIS (Revista Internacional de Sistemas – International Journal of Systems). He was also Visiting Professor at North-East Normal University (NENU) in Changchun, teaching regular courses.

RBU JOURNAL OF
Library and information science

Volume 24, 2022

UGC CARE
PEER REVIEWED

Department of Library and Information Science
Rabindra Bharati University
Kolkata

RBU JOURNAL OF LIBRARY & INFORMATION SCIENCE

[Official organ of the Department of Library & Information Science]

Editor
Dr Sudip Ranjan Hatua

About the Journal

The RBU Journal of Library & Information science is a scholarly communication for education, research and development of the Library & Information science field. It is published annually in print format only. This publication is fully funded by Rabindra Bharati University; therefore it never asked any kind of charges or publication fees or donations from the author. The first volume was published in the year 1997. It received ISSN (0972-2750) in its 5th volume in the year 2001. From its 17th Volume published in the year 2015 the journal become peer-reviewed (follows blind peer review process) by the eminent experts across the country. This Journal was enlisted under *UGC List of Journals No. 45237, Sl. No. 2023 when UGC published a list of research journals published across the country in its website. Later this journal enlisted under UGC-CARE List w.e.f. 14.6.2019*. Present publication is its 24th Volume published in the year 2022.

The RBU Journal of Library and Information Science was published under following Editors-

- *Professor Pinakinath Mukhopadhyay : Vol.1 to Vol. 11*
- *Shri Salil Chandra Khan: Vol. 12 to Vol. 14*
- *Dr Sudip Ranjan Hatua: Vol. 15 to Vol. 24 (continuing...)*

For details please visit: http://rbu.ac.in/home/page/102

To view the content of the previous volumes and the full text e-copy of VOLUME 22, 23 & 24 please visit:
https://lisrbu.wixsite.com/dlis/rbu-journal-of-lis

Online Content of previous volumes (Full Text of Volume 22,23,24) available at:
https://lisrbu.wixsite.com/dlis/rbu-journal-of-lis

CONTENTS