INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Natural Language Processing (NLP)

Dr. Chandrika G

Asst. Professor, Dept of computer science (BCA), Sarvodaya Degree College, TUMKUR- 572101

Abstract: Machines utilize Natural Language Processing (NLP), a subfield of Artificial Intelligence (AI), to comprehend, evaluate, and translate human languages. Due to advancements in information and communication technology over the last ten years, NLP has gained increased awareness and has been the subject of several studies. Understanding the steps required to advance literary understanding is so crucial. The study's objective is to provide an organized review of the literature in NLP research using bibliometric analysis. In the discipline of NLP, the study highlights important research fields, research clusters, cited publications, significant authors, institutions, nations, and publishing patterns. From the Web of Science (WoS) database, 12541 NLP papers were taken out and subjected to further bibliometric analysis The outcome showed that the largest number of NLP publications occurred in 2021, with the first publication being in 1989. The leading journal with the most publications was IEEE Access, and NLP papers have earned the most citations—3174—than any other journal. Liu HF is the most productive author in the field of NLP, and Harward University is the most significant organization. When it comes to the overall amount of publications, the US is the top nation. Researchers looked at the applied sciences in great detail.

I. INTRODUCTION

A subfield of artificial intelligence (AI) called natural language processing (NLP) gives computers the ability to understand, produce, and modify human language. Natural language text or voice can be used to query the data using natural language processing.

It's likely that most customers have dealt with NLP without even recognizing it. For example, the fundamental technology underlying virtual assistants like Alexa, Cortana, Siri, and the Oracle Digital Assistant (ODA) is natural language processing.

Their ability to comprehend the user's request and reply in normal language is attributed to natural language processing (NLP). NLP is applicable to all human languages and may be used with both written and spoken language. Web search, email spam filtering, text or speech translation software, document summarization, sentiment analysis, grammar and spell checking are further examples of NLP-powered technologies.

II. HISTORY

NLP's past Although the name "machine translation" (MT) was not even used yet, the field's research had begun. During this time, research was not entirely regional. Although Chinese was utilized for MT, Russian and English were the most commonly used languages (Booth, 1967). According to the ALPAC study, which indicated that MT is going nowhere, MT/NLP research nearly perished in 1966. However, thereafter, a few MT production systems supplied their clients with output (Hutchins, 1986). By now, efforts had also begun to employ computers for literary and linguistic research. AI-influenced signature work dates back to 1960, when the BASEBALL Q-A systems were introduced (Green et al., 1961). LUNAR (Woods, 1978) and Winograd SHRDLU were these systems' logical successors.

Early in the 1980s, computational grammar theory emerged as a vibrant field of study concerned with meaning logics and knowledge's capacity to address user intents and beliefs as well as functions like themes and emphasis. By the end of the decade, there was a way to handle longer discourse inside the grammatico-logical framework thanks to strong general purpose sentence processors as SRI's Core Language Engine (Alshawi, 1992) [15] and Discourse Representation Theory (Kamp and Reyle, 1993). The community was expanding during this time. Grammars, tools, parsers, and practical resources (such the Alvey Natural Language Tools; Briscoe et al., 1987) become accessible. In addition to the challenges they addressed, the (D)ARPA conferences on speech recognition and message interpretation (information extraction) were notable for their extensive

One focus of the study article was user modeling (Kobsa and Wahlster, 1989), while discourse structure was the other focus area (Cohen et al., 1990). Rhetorical schemas may also be utilized to produce language that is both linguistically coherent and effectively communicates ideas, as demonstrated by McKeown (1985). Certain NLP studies highlighted key areas for the future, such as word meaning disambiguation (Small et al., 1988), probabilistic networks, statistically colored NLP, and lexicon study.

III. PREVIOUS WORK

Associated Work Many researchers developed the mechanisms and technologies that made natural language processing (NLP) what it is today. NLP is an excellent field for research because of tools like Sentiment Analyser, Parts of Speech (POS)Taggers, Chunking, Named Entity Recognition (NER), Emotion detection, and Semantic Role Labelling. A sentiment analyzer (Jeonghee et al., 2003) gathers opinions on a certain subject.

 Sentiment analysis includes sentiment extraction, association via relationship analysis, and feature term extraction tailored to a certain topic. Sentiment analysis uses the sentiment pattern database and the sentiment lexicon as two language resources. It attempts to assign scores ranging from -5 to +5, analyzing the texts for both positive and negative terms. Research is being conducted on parts of speech taggers for languages such as European languages for more languages, such as Hindi (Pradipta Ranjan Ray et al., 2003), Arabic, and Sanskrit (Namrata Tapswi, Suresh Jain, 2012). Words may be effectively tagged and categorized as verbs, adjectives, nouns, etc. Most part-speech processes are effective in European languages, while they are ineffective in Asian or Middle Eastern languages. The Sanskrit portion of the speech tagger specifically uses the treebank approach. Arabic employs a Support Vector Machine (SVM) technique (Mona Diab et al., 2004) to automatically tokenize, annotate basic phrases, and tag bits of speech in Arabic text.

Chunking – it is also known as Shadow Parsing, it works by labeling segments of sentences with syntactic correlated keywords like Noun Phrase and Verb Phrase (NP or VP). Every word has a unique tag often marked as Begin Chunk (B-NP) or Inside Chunk (I-NP). Chunking is usually evaluated using the CoNLL 2000 shared task. CoNLL 2000 provides test data for Chunking. Since then, several systems have arised (Sha and Pereira, 2003; McDonald et al., 2005; Sun et al., 2008), all reporting around 94.3% F1 score. These systems use features composed of words, POS tags, and tags. Usage of Named Entity Recognition in places such as Internet is a problem as people don't use traditional or standard English. This degrades the performance of standard natural language processing tools substantially. By annotating the phrases or tweets and building tools trained on unlabelled, in domain and out domain data (Alan Ritter., 2011). It improves the performance as compared to standard natural language processing tools.

IV. APPLICATIONS

Natural language processing tools are important for businesses that deal with large amounts of unstructured text, whether emails, social media conversations, online chats, survey responses, and many other forms of data.

Let's take a look at 11 of the most interesting applications of natural language processing in business:

A. Sentiment Analysis

Natural language understanding is particularly difficult for machines when it comes to opinions, given that humans often use sarcasm and irony. Sentiment analysis, however, is able to recognize subtle nuances in emotions and opinions – and determine how positive or negative they are.

When you analyze sentiment in real-time, you can monitor mentions on social media (and handle negative comments before they escalate), gauge customer reactions to your latest marketing campaign or product launch, and get an overall sense of how customers feel about your company.

B. Text Classification

Text classification, a text analysis task that also includes sentiment analysis, involves automatically understanding, processing, and categorizing unstructured text.

C. Chatbots & Virtual Assistants

Chatbots and virtual assistants are used for automatic question answering, designed to understand natural language and deliver an appropriate response through natural language generation.

D. Text Extraction

Text extraction, or information extraction, automatically detects specific information in a text, such as names, companies, places, and more. This is also known as named entity recognition. You can also extract keywords within a text, as well as pre-defined features such as product serial numbers and models.

E. Machine Translation

Machine translation (MT) is one of the first applications of natural language processing. Even though Facebooks's translations have been declared superhuman, machine translation still faces the challenge of understanding context.

F. Text Summarization

Automatic summarization is pretty self-explanatory. It summarizes text, by extracting the most important information. Its main goal is to simplify the process of going through vast amounts of data, such as scientific papers, news content, or legal documentation.

G. Market Intelligence

Marketers can benefit from natural language processing to learn more about their customers and use those insights to create more effective strategies.

H. Auto-Correct

Natural Language Processing plays a vital role in grammar checking software and auto-correct functions. Tools like Grammarly, for example, use NLP to help you improve your writing, by detecting grammar, spelling, or sentence structure errors.

V. RESULT

You can build web app that translates news from Arabic to English and Summarizes them, using great python libraries like newspaper, transformers .

A. Lexical Analysis and Morphological

The first phase of NLP is the Lexical Analysis. This phase scans the source code as a stream of characters and converts it into meaningful lexemes. It divides the whole text into paragraphs, sentences, and words.

B. Syntactic Analysis (Parsing)

Syntactic Analysis is used to check grammar, word arrangements, and shows the relationship among the words.
Example: Agra goes to the Poonam
In the real world, Agra goes to the Poonam, does not make any sense, so this sentence is rejected by the Syntactic analyzer.

C. Semantic Analysis
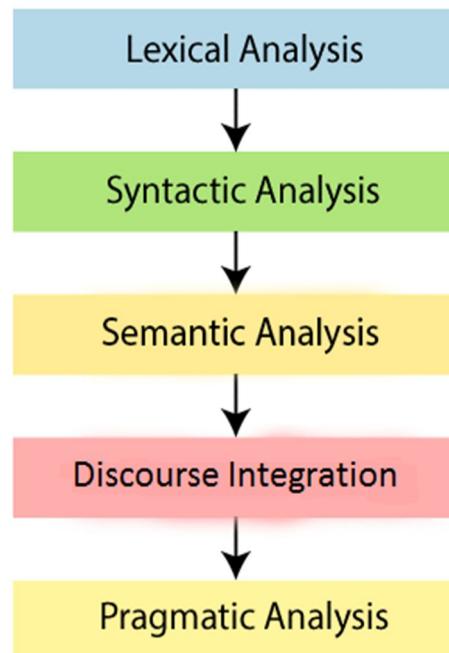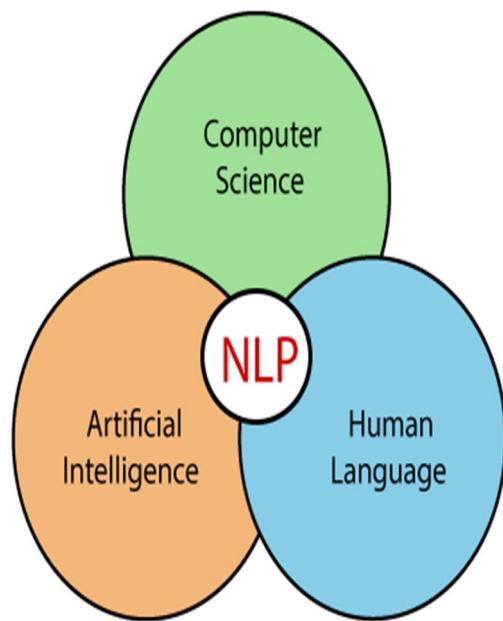
Semantic analysis is concerned with the meaning representation. It mainly focuses on the literal meaning of words, phrases, and sentences.

D. Discourse Integration

Discourse Integration depends upon the sentences that proceeds it and also invokes the meaning of the sentences that follow it.

E. Pragmatic Analysis

Pragmatic is the fifth and last phase of NLP. It helps you to discover the intended effect by applying a set of rules that characterize cooperative dialogues.
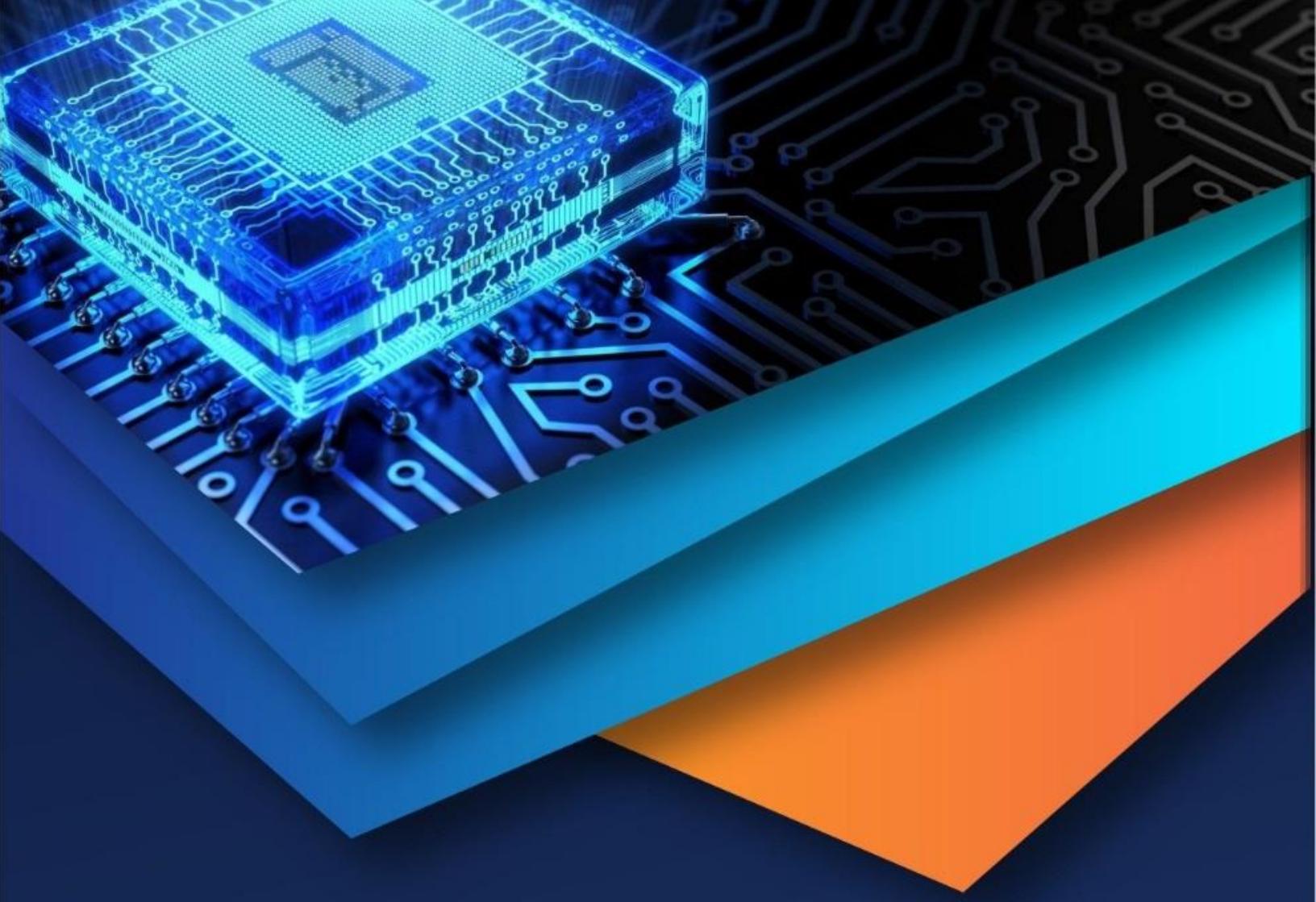
VI.   CONCLUSION

In the end, we think the suggested fix is tasteful and straightforward to apply. The modular design of the program facilitates independent component evolution and replacement, provided that cross-component communication is upheld and updated.

Subsequent efforts will concentrate on enhancing the functionality and communication of every module. By increasing the general purpose crawler's ability to gather pertinent curated content and creating additional specialized crawlers, we want to increase the adaptability of the Crawling component. By incorporating Xhtml into the Semantic Analysis module, the information gathered for project calls may be more accurate. Security settings and query language integration in the API will be improved in the back office and user interface to expedite data gathering for end users.

REFERENCES

[1] Chomsky, Noam, 1965, Aspects of the Theory of Syntax, Cambridge, Massachusetts: MIT Press.
[2] Rospocher, M., van Erp, M., Vossen, P., Fokkens, A., Aldabe,I., Rigau, G., Soroa, A., Ploeger, T., and Bogaard, T.(2016). Building event-centric knowledge graphs from news. Web Semantics: Science, Services and Agents on the World Wide Web, In Press.
[3] Shemtov, H. (1997). Ambiguity management in natural language generation. Stanford University.
[4] Emele, M. C., & Dorna, M. (1998, August). Ambiguity preserving machine translation using packed representations. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊙ (24*7 Support on Whatsapp)