Seminar Report on Natural Language Processing

Submitted in partial fulfillment for the requirement of the degree of

Bachelor of Technology
In
Information Technology

Submitted By
Vivek Kejriwal
Regd. No. - 1802081033

Under the supervision of
Dr. H. S. Behera

(Professor)

DEPARTMENT OF INFORMATION TECHNOLOGY,
VEER SURENDRA SAI UNIVERSITY OF TECHNOLOGY,
BURLA, SAMBALPUR - 768018
[2021-22]

Department of Information Technology

Veer Surendra Sai University of Technology, Burla, Odisha

CERTIFICATE

This is to certify that the seminar entitled Seminar Report on "Natural Language Processing" is being submitted by Vivek Kejriwal, Reg.no 1802081033, to the Department of Information Technology, Veer Surendra Sai University of Technology, Burla, in partial fulfillment of the requirements for the degree of Bachelor of Technology in Information Technology.

Dr. H. S. Behera
Supervisor
VSSUT, Burla

ACKNOWLEDGEMENT

I would like to thank my supervisor **Dr. H. S. Behera**, for giving me the opportunity to present this seminar. I am extremely grateful for his insightful guidance, timely advice, continuous support and motivation throughout our candidature. I especially thank him for his kindness and excellent contributions.

I would like to thank all the faculty and staff members of the Department of Information Technology at the Veer Surendra Sai University of Technology, Burla, India, for their help and support. I am deeply thankful to **Prof. Banshidhar Majhi**, Vice-Chancellor, and **Dr. M. R. Senapati**, Head of the Department, Department of Information Technology, for their moral support and providing us with all the necessary facilities during our candidature. I would like to thank our family members for their unconditional love and support.

Vivek Kejriwal

Regd.No:1802081033

ABSTRACT

Natural language processing is a branch of computer science and artificial intelligence which is concerned with interaction between computers and human languages. Natural language processing is the study of mathematical and computational modeling of various aspects of language and the development of a wide range of systems. These includes the spoken language systems that integrate speech and natural language. Natural language processing has a role in computer science because many aspects of the field deal with linguistic features of computation. Natural language processing is an area of research and application that explores how computers can be used to understand and manipulates natural language text or speech to do useful things. The applications of Natural language processing include fields of study, such as machine translation, natural language text processing and summarization, user interfaces, multilingual and cross language information retrieval (CLIR), speech recognition, artificial intelligence (AI) and expert systems.

CONTENTS

1. INTRODUCTION

NLP is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation. “Apart from common word processor operations that treat text like a mere sequence of symbols, NLP considers the hierarchical structure of language: several words make a phrase, several phrases make a sentence and, ultimately, sentences convey ideas,” John Rehling, an NLP expert at Meltwater Group, said in How Natural Language Processing Helps Uncover Social Media Sentiment. “By analyzing language for its meaning, NLP systems have long filled useful roles, such as correcting grammar, converting speech to text and automatically translating between languages.” NLP is used to analyze text, allowing machines to understand how humans speak. This human-computer interaction enables real-world applications like automatic text summarization, sentiment analysis, topic extraction, named entity recognition, parts-of-speech tagging, relationship extraction, stemming, and more. NLP is commonly used for text mining, machine translation, and automated question answering. NLP is characterized as a difficult problem in computer science. Human language is rarely precise, or plainly spoken. To understand human language is to understand not only the words, but the concepts and how they’re linked together to create meaning. Despite language being one of the easiest things for the human mind to learn, the ambiguity of language is what makes natural language processing a difficult problem for computers to master.

2. PREAMBLE OF THE PROBLEM

The essence of Natural Language Processing lies in making computers understand the natural language. That's not an easy task though. Computers can understand the structured form of data like spreadsheets and the tables in the database, but human languages, texts, and voices form an unstructured category of data, and it gets difficult for the computer to understand it, and there arises the need for Natural Language Processing. There's a lot of natural language data out there in various forms and it would get very easy if computers can understand and process that data. We can train the models in accordance with expected output in different ways. Humans have been writing for thousands of years, there are a lot of literature pieces available, and it would be great if we make computers understand that. But the task is never going to be easy. There are various challenges floating out there like understanding the correct meaning of the sentence, correct Named-Entity Recognition (NER), correct prediction of various parts of speech, co-reference resolution (the most challenging thing in my opinion). Computers can't truly understand the human language.

If we feed enough data and train a model properly, it can distinguish and try categorizing various parts of speech (noun, verb, adjective, supporter, etc.) based on previously fed data and experiences. If it encounters a new word, it tried making the nearest guess which can be embarrassingly wrong few times. It's very difficult for a computer to extract the exact meaning from a sentence. For example – The boy radiated fire like vibes. The boy had a very motivating personality or he actually radiated fire? As you see over here, parsing English with a computer is going to be complicated. There are various stages involved in training a model. Solving a complex problem in Machine Learning means building a pipeline. In simple terms, it means breaking a complex problem into a number of small problems, making models for each of them and then integrating these models. A similar thing is done in NLP. We can break down the process of understanding English for a number of small pieces.

3. HOW DOES NLP WORK?

It doesn't matter whether it's processing an automatic translation or a conversation with a chat bot: all natural language processing methods are the same in that they all involve understanding the hierarchies that dictate interplay between individual words. But this isn't easy – many words have double meanings. 'Pass' for example can mean a physical handover of something, a decision not to partake in something, and a measure of success in an exam or another test format. It also operates in the same conjugation as both a verb and a noun. The difference in meaning comes from the words that surround 'pass' within the sentence or phrase (I passed the butter/on the opportunity/the exam).

These difficulties are the main reason that natural language processing is seen as one of the most complicated topics in computer science. Language is often littered with double meanings, so understanding the differences requires an extensive knowledge of the content in which the different meanings are used. Many users have first-hand experience of failed communication with chat bots due to their continued use as replacements for live chat support in customer service.

But despite these difficulties, computers are improving their understanding of human language and its intricacies. To help speed this process up, computer linguists rely on the knowledge of various traditional linguistic fields:

Steps in NLP:

- **Lexical Analysis** − It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.

- **Syntactic Analysis (Parsing)** − It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as "The school goes to boy" is rejected by English syntactic analyzer.

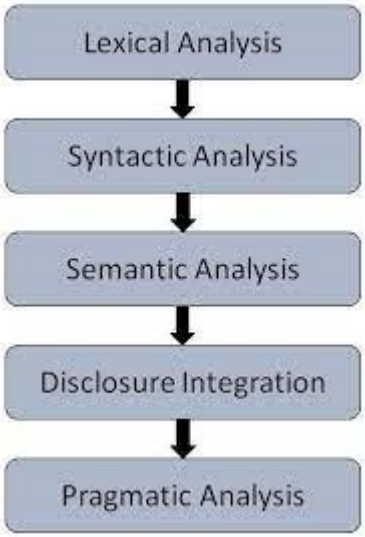

Figure 1: Sequential Steps to break NLP

• **Semantic Analysis** − It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as “hot ice-cream”.

• **Disclosure Integration** −The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.

• **Pragmatic Analysis** −During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

4. APPLICATIONS OF NLP

- **Sentiment Analysis**:

The goal of sentiment analysis is to identify sentiment among several posts or even in the same post where emotion is not always explicitly expressed. Companies use natural language processing applications, such as sentiment analysis, to identify opinions and sentiment online to help them understand what customers think about their products and services (i.e., "I love the new iPhone" and, a few lines later "But sometimes it doesn't work well" where the person is still talking about the iPhone) and overall indicators of their reputation. Beyond determining simple polarity, sentiment analysis understands sentiment in context to help you better understand what's behind an expressed opinion, which can be extremely relevant in understanding and driving purchasing decisions.

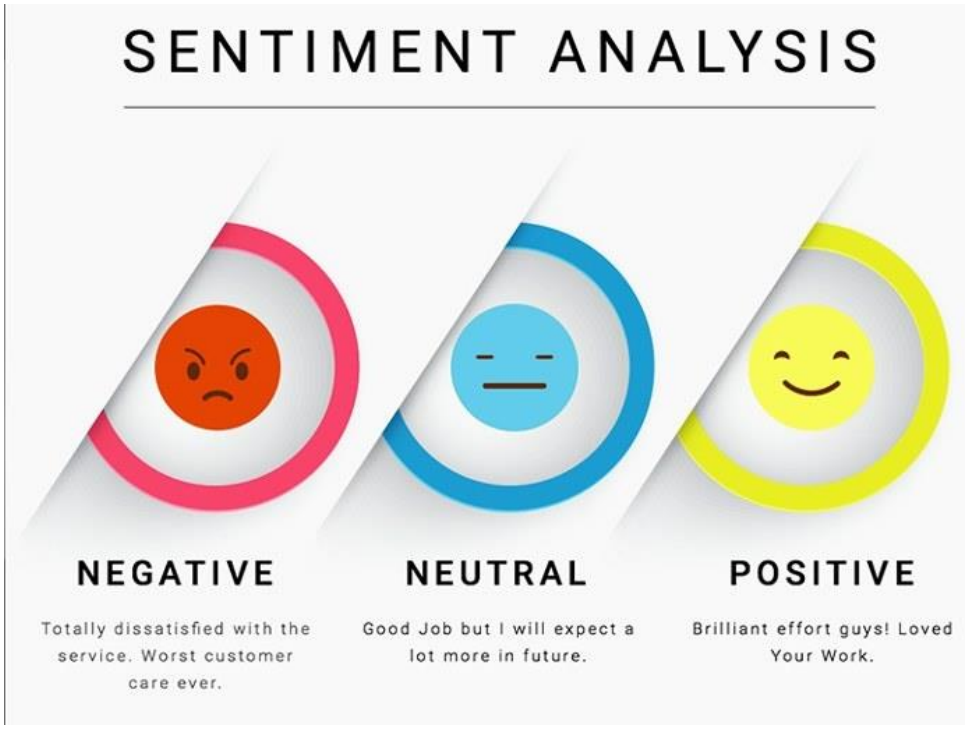

Figure 2: Negative, positive and neutral sentiments

- **Machine Translation**:

As the amount of information available online is growing, the need to access it becomes increasingly important and the value of natural language processing applications becomes clear. Machine translation helps us conquer language barriers that we often encounter by translating technical manuals, support content or catalogs at a significantly reduced cost. The challenge with machine translation technologies is not in translating words, but in understanding the meaning of sentences to provide a true translation.

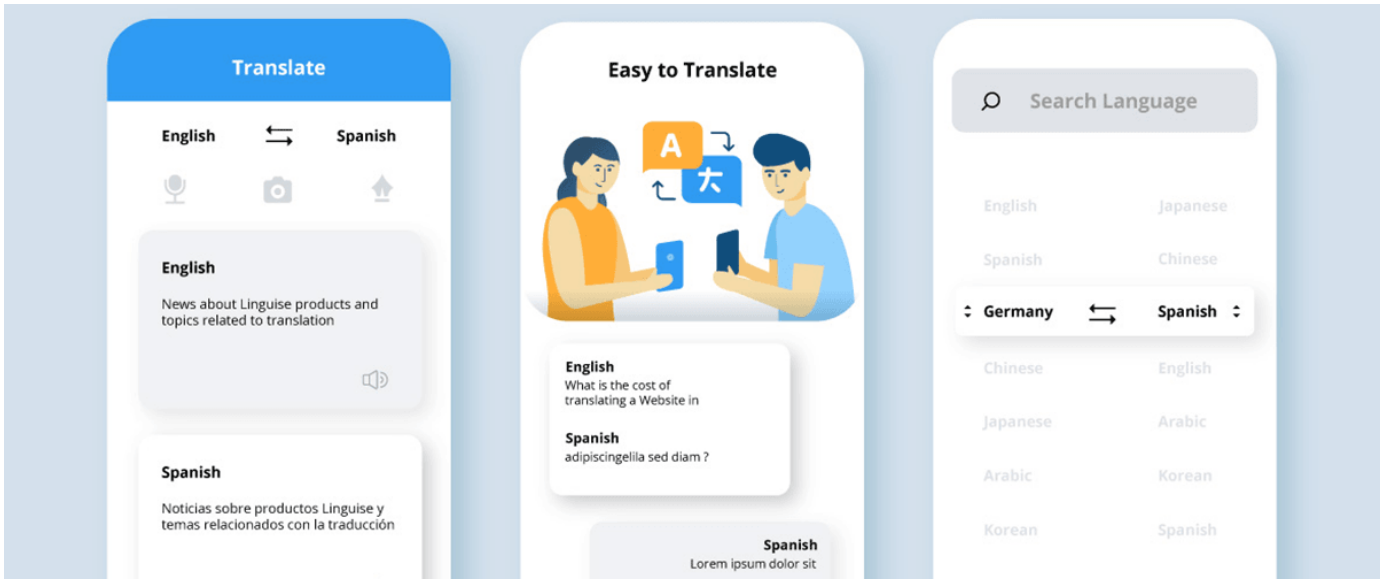

Figure 3: Machine Translation in different languages

- **Speech Recognition**

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. Rudimentary speech recognition software has a limited vocabulary of words and phrases, and it may only identify these if they are spoken very clearly. The first component of speech recognition is, of course, speech. Speech must be converted from physical sound to an electrical signal with a microphone, and then to digital data with an analog to digital converter. Once digitized, several models can be used to transcribe the audio to text.

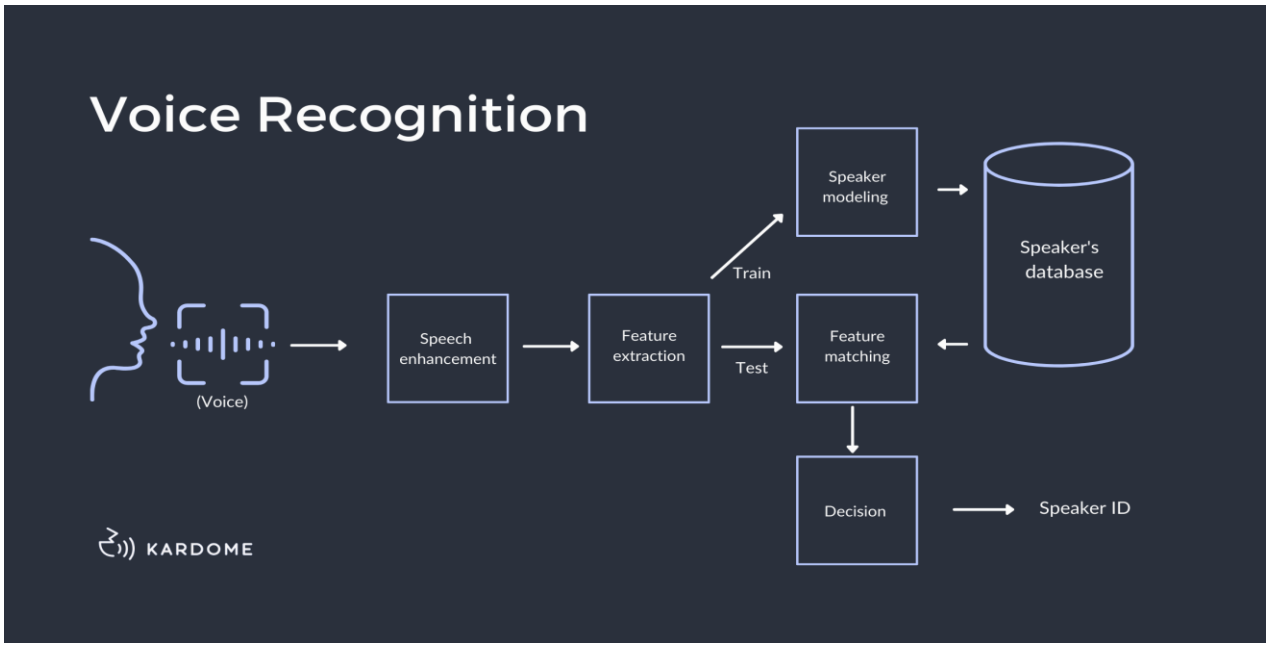

Figure 4: Voice recognition mechanism

- **Spell Checking**

In order to correct the error, a spell checker searches the dictionary for words that resemble the erroneous word most. These words are then suggested to the user who chooses the word that was intended. Spelling checking in used in various application like machine translation, search, information retrieval and etc.

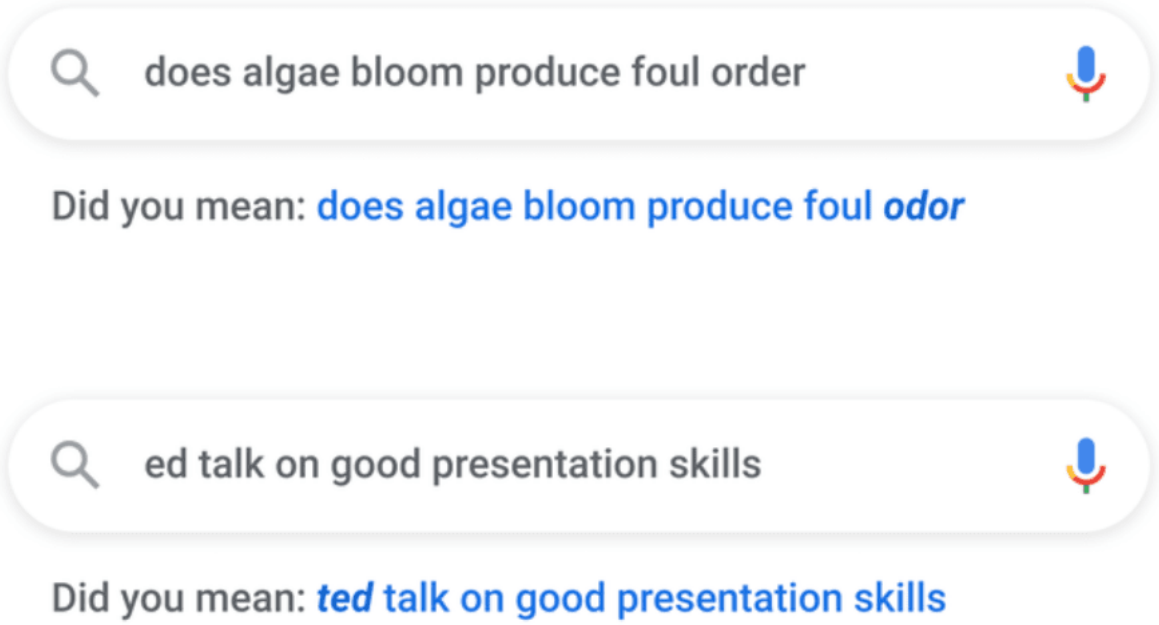

Figure 5: Spelling Checker enabled on google search engine

5. CONCLUSION

NLP supposedly makes the job easier but still demands a human interference. People and the industry fear NLP would start a trend of job snatching which is true to a certain sense but it certainly cannot function the way it does without human inputs. The will to work and cater to the loopholes or bugs in a machine is the task of a human who is handling it. Notwithstanding, the advantages of NLP may anger in the arena of jobs but right now it is the knight in the shining armor of the industry. While NLP is a relatively recent area of research and application, as compared to other information technology approaches, there have been sufficient successes to date that suggest that NLP-based information access technologies will continue to be a major area of research and development in information systems now and far into the future.

6. FUTURE SCOPE

The bots:

Frequently used in customer service, especially in banking, retail and hospitality, chatbots help customers get right to the point without the wait, answering customer questions and directing them to relevant resources and products at any hour, any day of the week. To be effective, chatbots must be fast, smart and easy to use, especially in a customer service context where users have high expectations (and sometimes low patience) for being understood. To accomplish this, chatbots employ NLP to understand language, usually over text or voice-recognition interactions, where users communicate in their own words, as they would speak to an agent. Integration with semantic and other cognitive technologies that enable a deeper understanding of human language will allow chatbots to get even better at understanding and replying to more complex and longer-form requests and functions in more than a single context, all in real time.

This expanded functionality will also benefit other types of bots to make them more effective and intuitive over time, from virtual assistants like Siri and Amazon's Alexa to bot platforms that are more automation or task oriented. These bots will increasingly use NLP to understand text and perform actions such as sharing geo information, retrieving links and images or execute other more complex actions for us, and you can see some of these at work in some of the most common messaging applications.

REFERENCES

- Allen, J. Natural Language Understanding. Benjamin/Cummings, Menlo Park, CA, 1987.

- https://data-flair.training/blogs/ai-natural-language-processing/