NLP : A study on the topic of Natural Language Processing

Monesa Thoguluva Janardhanan
Computer Science Department
University of North Carolina at Charlotte
Charlotte, United States of America
mthogulu@uncc.edu

***Abstract*—The main aim of this paper is to present the ideas grasped from the study of Natural Language Processing. NLP is a branch of Artificial intelligence that deals with the ability to communicate and share information [2]. This paper talks about what is NLP and why is it used, their applications, techniques, methodologies, state of art research conducted. This paper also discusses about the steps or the stages involved in NLP for building a system. It briefly touches about the architecture of NLP classification. Also 12 research papers were randomly selected for this study which uses NLP techniques for their research. A brief explanation of their findings, algorithms used are mentioned in the Literature Review of this paper.**

Keywords—Natural Language Processing (NLP), Natural Language Understanding (NLU), Natural Language Generation (NLG), NLP techniques, Linguistics, Knowledge Base, Lexicalization.

I. Introduction

Machine Learning Techniques are numerously used in every field as this empowers the software to learn, explore and analyses any given data without any human intervention [1]. Natural Language Processing is a part of computer science that deals with Artificial intelligence and Linguistics. NLP has the ability to make computers understand the statement or generate a new statement in the human language [3]. A language can be defined as a set of rules or set of symbol. Symbols are combined and used for conveying information or broadcasting the information. Symbols are tyrannized by the Rules[4]. Fig. 1 describes about the branches involved in NLP [3].

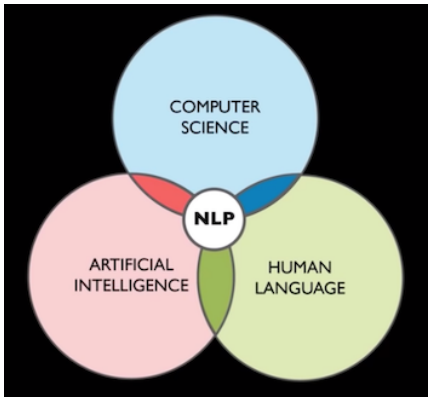

Figure 1: Branches involved in NLP [3]

Some of the applications of NLP are Sentiment Analysis, Chatbot, Speech Recognition and Machine Translation [3]. NLP is broadly classified into two categories - Natural Language Understanding and Natural Language Generation. The researched tasks of NLP involve Automatic Summarization, Co-Reference Resolution, Discourse Analysis, Machine Translation, Morphological Segmentation, Named Entity Recognition, Optical Character Recognition, Part of Speech Tagging etc [3][4]. Automatic summarization produces an understandable summary of a set of given text. Co-reference resolution refers to a sentence or larger set of text that determines which words are similar. Discourse analysis refers to the task of identifying the structure of connected text [4]. Machine translation refers to automatic translation of text from one human language to another. Morphological segmentation refers to separate word into individual morphemes that identifies the class of the morphemes [4]. Named entity recognition (NER) it describes a stream of text, determine which items in the text relates to proper names [3]. Optical character recognition (OCR) it gives an image representing printed text, which help in determining the corresponding or related text [4]. Part of speech tagging it describes a sentence, determines the part of speech for each word. The goal of Natural Language Processing is to accommodate one or more specialties of an algorithm or system [4]. Fig. 2 describes about the broad classification of NLP [4]. The next sections of this paper talks about NLP Techniques and Methodologies, Terminologies, NLP Applications and Literature Review.

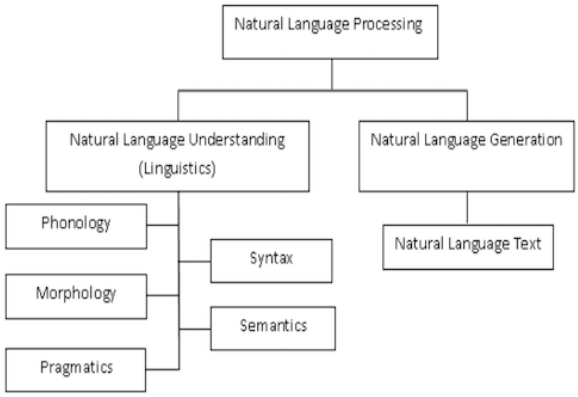

Figure 2: Broad Classification of NLP [4]

II. Techniques and Methodologies of NLP

The following are the techniques which are widely used in NLP.

A. Preprocessing or Data Cleaning

- **Tokenization**: Tokenization means splitting the words into Tokens. There are different types of tokenization like line tokenization, word tokenization, Non-English tokenization [5]. Line tokenization splits the text into lines. e.g. Nice to meet you. What is your name? will be split into two. Word tokenization splits the sentence into words. e.g. what is your name will be split into 'what', 'is', 'your', 'name'. Non-English tokenization splits the text into tokens which are of non-English

origin. For e.g. geht es Ihnen? Gut, danke.' will be split into 'Wie geht es Ihnen?', 'Gut, danke.' [5]

- **Stemming:** Stemming is the process of producing morphological variants of a root word. Stemming programs are referred to as stemming algorithms or stemmers [5]. A stemming algorithm reduces the words “chocolates”, “chocolatey”, “choco” to the root word, “chocolate” and “retrieval”, “retrieved”, “retrieves” reduce to the stem “retrieve” [6]. There are mainly two errors in stemming – Overstemming and Understemming. Overstemming occurs when two words are stemmed to same root that are of different stems. Under-stemming occurs when two words are stemmed to same root that are not of different stems. Figure 3 shows the word ending with “ies” and word ending with “s” are converts to the root form with y and removal of last letter respectively [13].

```
string do_stemmer(string word)
{
    string base_form = null ;
    if (word.Length > 3)
    {
        if (word.EndsWith("ies"))
        {
            base_form = word.Substring(0, word.Length - 3);
            base_form = base_form + "y";
            if (is_noune(base_form))
                return base_form;
        }

        if (word.EndsWith("s"))
        {
            base_form = word.Substring(0, word.Length - 1);
            if (isnoune(base_form))
                return base_form;
        }
    }
    return word;
```

Figure 3: Stemming Algorithm taken from Reference [13]

- **Lemmatization:** Lemmatization groups together different inflected forms of word or otherwise called Lemma. It is somehow similar to stemming as it maps several words to a common root word [3]. The output of Lemmatization is a proper word. For example, gone, went, going comes to go.

- **POS Tagging:** The process of classifying words into their parts of speech is called POS Tagging. Parts of speech are also known as word classes or lexical categories [7]. For example let’s take the sentence "And now for something completely different". The output of POS tagging will look something like ('And', 'CC'), ('now', 'RB'), ('for', 'IN'), ('something', 'NN'), ('completely', 'RB'), ('different', 'JJ') [7].

- **Named Entity Organization:** Named Entity Organization is the main step needed when it comes to information extraction. It classifies named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages [8]. For example, "Google's CEO Sundar Pitchai introduced new pixel 3 in New York central mall". The output of this will look some like ('Google' , 'Organization') , ('Sundar Pitchai' , 'Person') , ('New York' , 'Location') , ('Central Mall' ,' Organization') [3].

- **Chunking:** Chunking is the process of grouping similar words together based on the nature of the word. It is also called as Shadow Parsing. A grammar should be defined for which the chunk will be generated [9]. It is like the sequence of the phrases like nouns, adjectives, adverb etc. These rules will be followed and the chunk is created.

There are many more techniques used in Data Cleaning like removing stopwords, word replacement, chunk classification, creation of document term matrix etc [5][10]. It depends on the text that we are using and depends on what kind of output is expected.

B. Sentiment Analysis

Sentiment Analysis comes under Natural Language Processing and it is widely used. Basically sentiment analysis indicates whether the statement said is positive or negative. So the input of the sentiment Analysis is a corpus [10]. For example, great is positive and not great is negative. If the sentence has the word great then it declares that it is positive sentence. For each sentence there will be a sentiment score, if the score is more than the threshold then it is considered positive or negative. Texblob is a Python library which is built in the top of NLTK is widely used library for Sentiment Analysis. TextBlob finds all the words and phrases that it can assign polarity and subjectivity to, and average all of them together [10].

C. Topic Modeling

Latent Dirichlet Allocation (LDA) is one of the most used algorithms in Topic Modeling [11]. Topic Modeling is nothing but building a cluster of common words and find the theme [12]. The first step is to choose the number of words that we think are there in the corpus. Then it randomly assigns each word in each word in the document to the topics. Then it goes through every word and its topic assignment in each document. It looks at how often the topic occurs in the document and how often the word occurs in the topic overall. Based on this info it assigns the word to a new topic [11]. There are other matrix factorization technique for topic modeling Latent Semantic Indexing(LSI) and Negative Matrix Factorization (NMF) [12].

D. Text Generation

Text generation is the process by which the model generates a new text based on the previous text. The main important part of Text Generation is preserving the order of text and especially punctuation. Markov Chains comes under Text Generation [10]. The first step is to create a dictionary corpus where the keys are the current state and the values are the options for the next state. Then to write a function to randomly generate next term [11]. This is a much simpler process. A complex process would be to use deep learning and Long-Short term Memory (LSTM) because this not only takes account of previous word. But it takes account of the words used before the previous word and the words which has been already generated [10].

III. Applications and NLP in Talk

Natural Language Processing can be applied into various areas like Machine Translation, Text Categorization, Email Spam filtering, Information Extraction, Summarization, Question Answering, Dialogue System, Medicine etc [3][4].

The following mentions the recent developments in the NLP projects implemented by various companies:

- ACE Powered GDPR Robot Launched by RAVN Systems [4].

- Eno A Natural Language Chatbot Launched by Capital One [4].
- Future of BI in Natural Language Processing [4].
- Using Natural Language Processing and Network Analysis to develop a conceptual framework for Medicine Therapy Management Research [4].
- Meet the pilot, world's first translating earbuds [4].

IV. Literature Review

A. Terminologies of Natural Language Processing[4]

Phonology: Phonology comes under the branch of linguistics which refers to the systematic arrangement of sound. The term phonology comes from Ancient Greek and the term phono- which means voice or sound, and the suffix –logy refers to word or speech [4]. In 1993 Nikolai Trubetzkoy stated that Phonology is "the study of sound pertaining to the system of language". Whereas Lass in 1998 wrote that phonology refers broadly with the sounds of language, concerned with the to lathe sub discipline of linguistics, whereas it could be explained as, "phonology proper is concerned with the function, behavior and organization of sounds as linguistic items. Phonology include semantic use of sound to encode meaning of any Human language (Clark et. al) [13].

Morphology: Morphemes are the smallest units of meanings represented by different parts of word. Morphology is the nature of words which are initiated by morphemes. Prefix, suffix, tense comes under morpheme. Any word can be broken into morpheme which represents some meaning of the word. The different parts of the word represent the smallest units of meaning known as Morphemes [4]. For example, suffix -ed indicates that the action took place in the past. The words that cannot be divided and have meaning by themselves are called Lexical morpheme (e.g. Bottle, Computer) The words (e.g. -ed, -ing, -est, -ly, -ful) that are combined with the lexical morpheme are known as Grammatical morphemes (e.g. Largest, Likely, Played, Studied, Ending). Those grammatical morphemes that occurs in combination called bound morphemes(e.g. -ed, -ing) Grammatical morphemes can be divided into bound morphemes and derivational morphemes [4].

Lexical: Lexical is nothing but interpreting the meaning of individual words. The first step in this is to identify the parts of speech of each word. If a word comes under multiple parts of speech then the most probable part of speech will be tagged to it [4]. At this level words can replaced with common words that have same meaning.

Semantic: The meaning of the sentence differ from how it is said, where it is said and also depends on the punctuation marks. Semantic processing tries to determine the possible meanings of the sentence by connecting the different words [4]. For example a note can be interpreted as a note book or a note or the verb as note. Semantics milieu that most words have more than one elucidation but that we can spot the appropriate one by looking at the rest of the sentence [14].

Discourse: Syntax and Semantics deals with words and a single sentence while discourse in NLP deals with units longer than the sentence. i.e it does not interpret multi sentence texts as just sequence sentences, a piece of which can be elucidated singly. Rather it focuses on the properties of the text as a whole which convey the meaning by making connections (Elizebeth D. Liddy, 2001) [15][4]. The two of the most common levels are Anaphora Resolution and Discourse/Text Structure recognition. Anaphora resolution is the replacing of words such as pronouns, which stands separately and replacing it with the previous entity. Discourse/Text Structure recognition sway the functions of sentences in the text, which, in turn, adds to the meaningful representation of the text.

Pragmatic: Pragmatic is concerned with the use of language in situations for understanding the goal and explains how extra meaning is read into texts without literally being encoded in them. This require much world knowledge, including the understanding of intentions, plans, and goals. For example, the following two sentences need aspiration of the anaphoric term 'they', but this aspiration requires pragmatic or world knowledge [7][4].

B. Employing Natural Language Processing to Analyse Grammatical Error in a Simple Japanese Sentence [16]

This paper presents the research in building an error analysis system to help the language learning process, especially in writing a Japanese sentence with correct grammar. Writing error, particle error, sentence formation error and predicate usage error can be analyzed by the system. The NLP tools used are morpheme analyzer, lexical analyzer and syntax analyzer. For the morpheme analyzer, JUMAN is used; and for the lexical analyzer and syntax analyzer, KNP is used. The system evaluation had been divided into sentence's completeness evaluation, tree structure evaluation, and grammatical function evaluation. Given a Japanese sentence input, the system detects the errors, analyses the cause of the errors, and finally gives instructions to the user about how to correct the sentence. This research indicates that the accuracy is 80.30%. Figure. 4 shows the input and system response of this research. After 公園 (park) で should come instead of に (this particle is used specify timings etc.) and after ご飯 (lunch) を should come instead of の (this particle is used when indicating belongs to) and the system has identified this accurately as shown below.

Input Sentence
私は公園にご飯の食べます (watashi ha kouen ni gohan no tabemasu)
System Response
The usage of particle "に" (ni) after "公園" (kouen) is slightly inappropriate. To mark a place where an activity is done, please use particle "で" (de) instead. The usage of particle "の" (no) after "ご飯" (gohan) is slightly inappropriate. To mark the object which is affected by an action, please use pariticle "を" (wo) instead.

Figure 4: Input and System Response from Reference [16]

C. Natural Language Processing in a Japanese Text-To-Speech System for Written-style Texts [17]

This paper demonstrates newly developed Japanese text-to-speech (TTS) system, AUDIOTEX, that modifies written-style expressions for listening and generates prosodic information (Kanji reading, accent, and pause). It states that the conventional TTS systems have several problems: (1) most texts are written to be read with the human eye, thus they are inadequate for listening, and (2) kanji reading errors. The steps involved in this research are (1) A rewriting process modifies the sentence for listening. (2) A morphological

analysis separates a sentence into words using dictionary. (3) A prosodic process determines accentual phrases and pause locations and generates Katakana characters which corresponds to the syllables. (4) A speech synthesis process covers katakana characters to phoneme strings [17]. Figure. 5 represents the System configuration of this research.

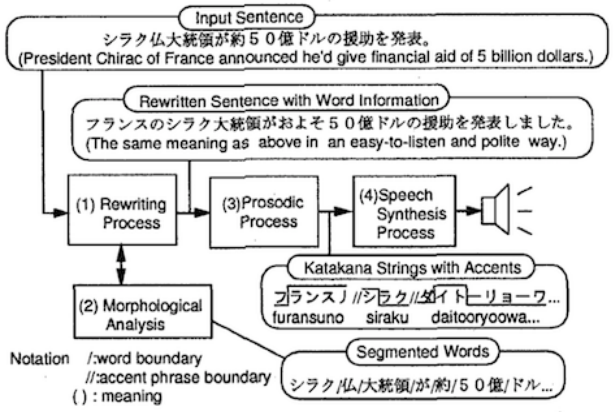

Figure 5: System configuration of Reference [17]

D. *Natural Language Processing for SemanticWeb Services[18]*

This research has implemented a prototype of Semantic Web Service, that uses natural language processing to interpret the conventional, natural language content of traditional Web page and to retrieve the information asked for in natural language as well. Semantic Web is an emerging concept that advocates encoding of information and services in a structured machine interpretable manner [18]. This research uses Hierarchical Semantic Form (HSF) which resembles Hierarchical Organized Neural Networks. This research uses SOUL (Space of Universal Links) algorithm. This research states that HSF with SOUL algorithm is scalable and facilities easy defining of new semantic categories, thus enhancing cognitive capabilities of the system. Figure. 6 represents the Complex Semantic category for the following sentence "Could you please give me the first five morning Lufthansa flights departing from Berlin and arriving at Rome on next Sunday?".

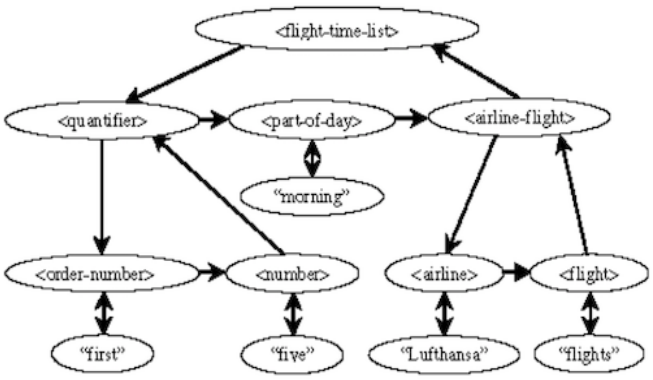

Figure 6: Complex Semantic Category of Reference [17]

E. *Development of GUI for Text-to-Speech Recognition using Natural Language Processing[19]*

Reference [19] is a Text-to-speech synthesizer that converts text into spoken word, by analyzing and processing it using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into speech representation of the text. This research has developed a useful text-to-speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can also be saved as an audio file. TTS is the automatic conversion which configures the concept of speech recognition, speech analysis, speech synthesis, speech tuning, speech alteration. Figure. 7 describes about the TTS synthesis model used in this research. The first interface converts raw text containing alpha-numeric symbols like numbers and abbreviations in terms of speech into the equivalent of out words [18]. Then symbolic linguistic representation is generated which is the output of TTS.

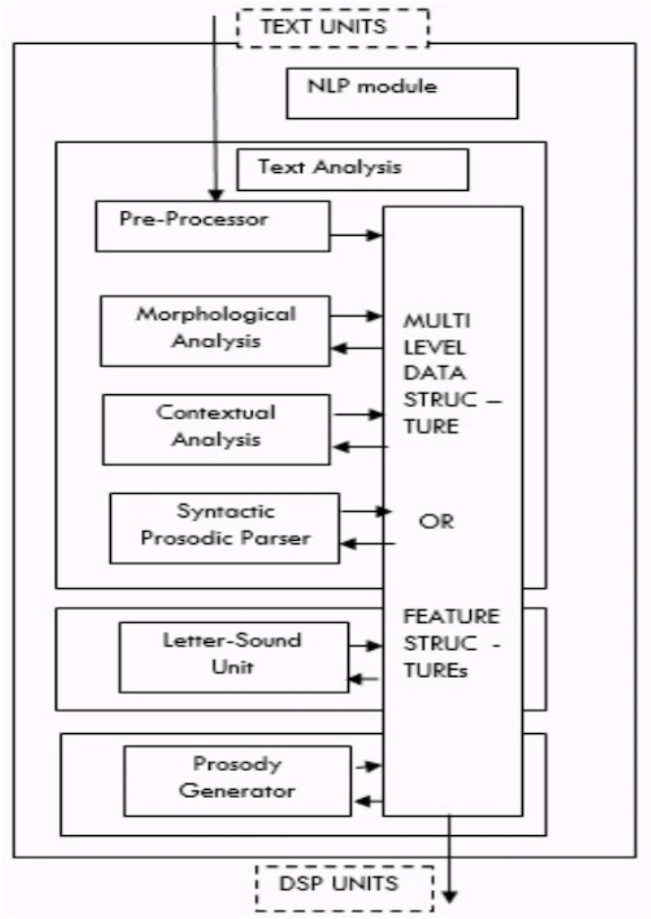

Figure 7: TTS Synthesis model of Reference [19]

F. *Class diagram extraction from textual requirements using Natural language processing (NLP) techniques[20]*

Reference [20] proposes a method and a tool to facilitate requirements analysis process and class diagram extraction from textual requirements supporting natural language processing NLP and Domain Ontology techniques. A desktop application called RACE (Requirements Analysis and Class Diagram Extraction) is created in this research which generates the class diagram from the textual requirements. OpenNLP parser has been used in this research which provides the lexical and syntactic parser. WordNet is used to validate the semantic correctness of the sentence generated at the syntactic analysis. Figure. 8 demonstrates the architecture of the RACE algorithm. The concept extraction engine uses OpenNLP parser. This research was able to find four types of relationship - Generalization, Association, Composition, aggregation and dependency.

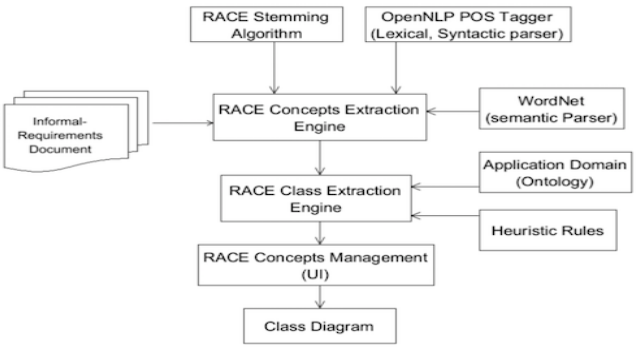

Figure 8: RACE System Architecture of Reference [20]

G. *Natural Language Requirements Processing: from Research to Practice[21]*

Reference [21] basically talks about Natural Language Processing from Research to Practice. It states that Natural language processing (NLP) techniques have been largely applied to automate several requirements engineering (RE) tasks, including model synthesis, requirements categorization, trace ability, detection of equivalent requirements, information extraction, defect detection, and, more recently, classification of online product reviews [21]. It talks about Besides the Python NLTK library for text processing, one relevant example is GATE (General Architecture for Text Engineering), a user-friendly tool for information extraction and pattern identification that was applied in RE industrial case studies. It implies that NLP has not developed much in the Requirements analysis. The requirement analysis is not constant for one company to another and this article provides us a possible solution to these problems.

H. *Information Processing and Retrieval from CSV File by Natural Language[22]*

Reference [22] presents a new model that will allow users to easily retrieve information from CSV files by natural language. Figure. 9 describes about the architecture used for this research. First the csv file is stored as a table in the database. Once the user gives the condition in the natural language, lexical, Syntactic and Semantic analysis are being done in the text the text is converted to a SQL query. This SQL query is used in the database to retrieve the values. Thus, in turn retrieving the values from the csv file based on the condition.

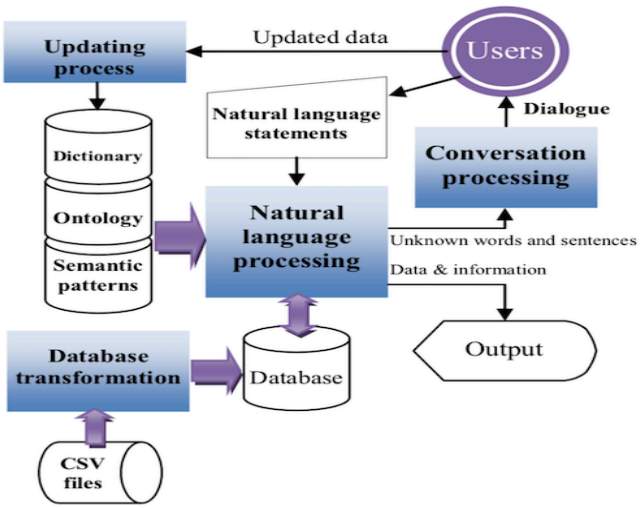

Figure 9: NLP models of Reference [22]

I. *An Automated Tool for Generating UML Models from Natural Language Requirements[23]*

Reference [23] describes a domain independent tool, named, UML Model Generator from Analysis of Requirements (UMGAR), which generates UML models like the Use-case Diagram, Analysis class model, Collaboration diagram and Design class model from natural language requirements using efficient Natural Language Processing (NLP) tools. NLP technologies like Stanford Parser, JavaRAP, and WordNet 2.1 are used. Standford Parser does tokenization, sentence splitting and part-of-speech (POS) tagging. JavaRAP groups nouns, verbs and adjectives into cognitive synonyms (Synsets). WordNet2.1 helps in performing morphological analysis for converting plurals into singulars. It helps to resolve pronouns up to third person pronouns. Three kinds of model developer are present in the architecture. They are Use case model developer, Conceptual Model Developer and Design Class Model developer.

J. *Stock Prices Prediction using the Title of Newspaper Articles with Korean Natural Language Processing[24]*

Reference [24] uses CNN model of stock price prediction using Korean natural language processing. This research converted Korean sentences into nouns and vectorized them using skip grams to extract the characteristics of the words. Then, the vectorized word sentence was used as input data of the CNN model to predict the stock price after 5 days of trading day.

K. *Voice Controlled Home Automation System Using Natural Language Processing (NLP) And Internet Of Things (IOT)[25]*

Reference [25] proposes voice-controlled home automation system using Natural Language Processing. Arduino MK1000 is used as practical and economic solution for innovators seeking to create Wi-Fi enabled projects with minimal pre-requisite knowledge in the field. Using mobile the voice is delivered. Using NLP, the system understands the meaning of the sentence and it interacts with the hardware which in-turn interacts with the hardware for the changes.

L. *Research on Web Monitoring System Based on Natural Language Processing[26]*

Reference [26] studies and designs a Web monitoring system for controlling sensitive information spreading on the web. The methodology of this research is like once the web page is got then Structure mining, Extraction of main body text, Partitioning, Tagging, Lexical and Grammar analysis are performed in the page and if the text contain any sensitive information the web page will not be opened and will be screened. Data cleaning, Format conversion and detection of sensitive terms are some of the methods used in this research. Based on a threshold the system determined whether the web page is screened or not. Figure. 10 displays the web data mining classification from this reference.

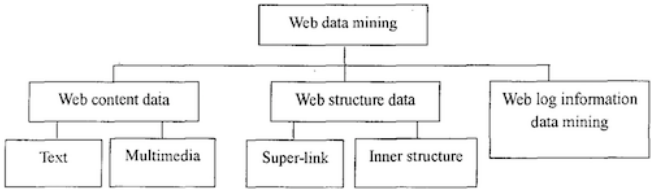

Figure 10: Web Data Mining Classification of Reference [14]

V. Conclusion

This paper demonstrates about Natural Language Processing its terminology, techniques, methodology, applications and its state of research conducted. Also 12 research papers were randomly selected and are explained in brief in the Literature Review Section. Natural language processing (NLP) has recently gained much attention for representing and analyzing human language computationally [4]. It is a way of computer processing of human language, and the computer returns a correct response that people expect. The study of natural language processing is divided into two kinds, one is written language processing, the other is spoken language processing [26]. Relatively: written language is more standard than spoken language, so it is easier to process written language by computer than spoken language [12]. It is said that language is the direct implementation of man's thinking. The society development depends on language (characters or others), so every subject tie up with language. When we study natural language processing, we should not only include words: grammar, sentence, word meaning in different situation but also include knowledge of real world and knowledge of relative subject [11].

References

[1] SAS. (2020) – Machine Learning : What is it and Why it matters https://www.sas.com/en_us/insights/analytics/machine-learning.html

[2] Dr. Michael J. Garbade (2018) : A Simple Introduction to Natural Language Processing – https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32

[3] Edureka(2018) - NLP Tutorial For Beginners | NLP Training | Edureka - https://www.youtube.com/watch?v=5ctbvkAMQO4

[4] Khurana, Diksha & Koli, Aditya & Khatter, Kiran & Singh, Sukhdev. (2017) - Natural Language Processing: State of The Art, Current Trends and Challenges.

[5] Tutorials Point Learn Python Text Processing (2020) : Python – Tokenization https://www.tutorialspoint.com/ python_text_processing / python_tokenization.htm

[6] Python | Stemming words in NLTK (2020) GeeksForGeeks : https://www.geeksforgeeks.org/python-stemming-words-with-nltk/

[7] Categorization and Tagging Words (2019) - http://www.nltk.org/book/ch05.html

[8] Susan Li (2018) : Named Entity Recognition with NLTK and Spacy - https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da

[9] Tutorials Point Learn Python Text Processing (2020) : Python Chunks and Chinks - https://www.tutorialspoint.com/ python_text_processing/ python_chunks_and_chinks.htm

[10] Alice Zhao. (2018) Comparing stand up comedians using natural language processing – https://github.com/adashofdata/nlp-in-python-tutorial.

[11] Alice Zhao. (2018) Natural Language Processing in Python – https://www.youtube.com/watch?v=xvqsFTUsOmc

[12] Susan Li (2018) https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24

[13] Nation, K., Snowling, M. J., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: Semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. Advances in Speech Language Pathology, 9(2), 131-139.

[14] Feldman, S. (1999). NLP meets the Jabberwocky: Natural Language Processing in Information Retrieval.ONLINE-WESTON THEN WILTON-, 23, 62-73

[15] Liddy, E. D. (2001). Natural language processing.

[16] A. N. S. Kasmaji and A. Purwarianti, "Employing natural language processing to analyse grammatical error in a simple Japanese sentence," 2015 International Conference on Electrical Engineering and Informatics (ICEEI), Denpasar, 2015, pp. 82-86.

[17] K. Matsuoka, E. Takeishi, H. Asano, R. Ichii and Y. Ooyama, "Natural language processing in a Japanese text-to-speech system for written-style texts," Proceedings of IVTTA '96. Workshop on Interactive Voice Technology for Telecommunications Applications, Basking Ridge, NJ, USA, 1996, pp. 33-36.

[18] M. Stanojevic and S. Vranes, "A Natural Language Processing for Semantic Web Services," EUROCON 2005 - The International Conference on "Computer as a Tool", Belgrade, 2005, pp. 229-232.

[19] P. M. ee, S. Santra, S. Bhowmick, A. Paul, P. Chatterjee and A. Deyasi, "Development of GUI for Text-to-Speech Recognition using Natural Language Processing," 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, 2018, pp. 1-4

[20] M. Ibrahim and R. Ahmad, "Class Diagram Extraction from Textual Requirements Using Natural Language Processing (NLP) Techniques," 2010 Second International Conference on Computer Research and Development, Kuala Lumpur, 2010, pp. 200-204.

[21] A. Ferrari, "Natural Language Requirements Processing: From Research to Practice," 2018 IEEE/ACM 40th International Conference on Software Engineering: Companion (ICSE-Companion), Gothenburg, 2018, pp. 536-537.

[22] C. Tapsai, "Information Processing and Retrieval from CSV File by Natural Language," 2018 IEEE 3rd International Conference on Communication and Information Systems (ICCIS), Singapore, Singapore, 2018, pp. 212-216.

[23] D. K. Deeptimahanti and M. A. Babar, "An Automated Tool for Generating UML Models from Natural Language Requirements," 2009 IEEE/ACM International Conference on Automated Software Engineering, Auckland, 2009, pp. 680-682

[24] H. Yun, G. Sim and J. Seok, "Stock Prices Prediction using the Title of Newspaper Articles with Korean Natural Language Processing," 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, 2019, pp. 019-021.

[25] P. J. Rani, J. Bakthakumar, B. P. Kumaar, U. P. Kumaar and S. Kumar, "Voice controlled home automation system using Natural Language Processing (NLP) and Internet of Things (IoT)," 2017 Third International Conference on Science Technology Engineering \& Management (ICONSTEM), Chennai, 2017, pp. 368-373.

[26] Liu Lin, Fan Xiaozhong and Zhao Xunping, "Research on Web monitoring system based on natural language processing," International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003, Beijing, China, 2003, pp. 746-751.