International Journal of
Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: http://doi.org/10.31695/IJASRE.2018.32708

Volume 4, Issue 5
May - 2018

Natural Language Processing

Matthew N. O. Sadiku, Yu Zhou, and Sarhan M. Musa

Roy G. Perry College of Engineering

Prairie View A&M University

Prairie View, TX 77446

U.S.A

ABSTRACT

Natural language processing (NLP) refers to the field of study that focuses on the interactions between human language and computers. It is a computational approach to text analysis. It is the application of a wide range of computational techniques for the understanding, automatic analysis, and representation of human language. This paper provides a brief introduction to NLP.

Key Words: Natural language processing, Computational linguistics.

1. INTRODUCTION

Although globalization has opened borders for exchange, users may find sensitive data in foreign websites in foreign languages. Language is crucial around the world in communication, entertainment, media, culture, drama, movie, and economy. There is the need for the establishment of language processing systems.

Natural language processing (NLP) refers to the application of computational techniques to the understanding and generation of human language. It involves

the study of mathematical and computational modeling of various aspects of language. It implies the development of a wide range of systems such as spoken language systems that integrate speech and natural language, multilingual interfaces, and machine translation. NLP systems perform useful roles, such as correcting grammar, converting speech to text, and automatically translating between languages

NLP (also known as computational linguistics) is a interdisciplinary field involving computer science, linguistics, logic, and psychology. NLP as an area of research and application began in the 1950s as the intersection of artificial intelligence (AI) and linguistics. Since then, NLP research has been focusing on tasks such as machine translation, information retrieval, text summarization, question-answering, information extraction, and opinion mining [1]. The late 1980s witnessed a revolution in NLP with the introduction of machine learning algorithms and statistical techniques for language processing. Machine learning approaches extract statistical information from large amounts of documents and learn the rules of language without explicitly listing them.

Human language is inherently complex and diverse. We express ourselves in many ways, both verbally and in writing. There are hundreds of languages (English, Spanish, Chinese, etc.) and dialects, each with a unique set of grammar and syntax rules, accent and slang [2]. We share meaning in several ways: texts, gestures, sign languages, and face expressions.

1.1 NLP BASICS

Natural language processing is a technique where machine can become more human and thereby making human to communicate with the machine easily. NLP seeks to make software intelligent enough to process a natural language as humans. For example, imagine a machine that takes instructions by voice.

NLP analysis generally consists of the following three levels [3]:

- *Syntax*, the study of sentence structure. Syntax deals with the formation of a sentence from individual words. Syntax alone suggests the proper interpretation of "Jimmy loves Lucy."
- *Semantics*, the study of context-independent meaning. This derives the meaning of a sentence based on the meanings of the words/phrases. For example, semantics determines whether the word "bank" refers to a river bank or to a financial institution.
- *Pragmatics*, the study of context-dependent meaning. Pragmatics deals with how meaning changes in the presence of a specific context and how the contexts affect the meaning of the sentences. This level is concerned with the purposeful use of language in situations

2. APPLICATIONS

NLP is important because of the major role language such as English plays in human intelligence and because of the wealth of potential applications. NLP is commonly used for text mining, machine translation, and automated question-answering. Applications of NPL include interfaces to expert systems and database query systems, machine translation, text generation, story understanding, automatic speech recognition, and computer-aided instruction. It also has great potential in healthcare, mobile technology, cloud computing, virtual reality, election, social work, and social networking. We will consider just a few of these as examples.

- *Healthcare*: NLP is important to health. NLP clinical systems can be used to represent clinical knowledge and clinical decision support interventions in standardized formats. They have been developed to process unstructured text and transform it into a desired coded form to support several healthcare-related activities. Such systems require a higher degree of accuracy as results are incorporated into critical decisions related to patient care. There is also NLP-based approach to extract information from patient records [4]. We will consider a few of these as examples.
- *Social media*: The role of social media in people's lives is becoming significant. Natural language processing is one of the most promising avenues for social media data processing.
- *Summarizer:* This will automatically summarize a block of text, exacting topic sentences, and ignoring the rest. Opinion summarization focuses on the opinionated parts, while traditional summarization focuses on extracting informative parts and getting rid of redundancy.
- *Sentiment Analysis:* Sentiment analysis is a task in which NLP automatically identifies the sentiment of a text based on the opinions or feelings written in the text. This can be used to identify the feeling, opinion, or belief of a statement. The main issue in sentiment analysis is to identify how sentiments are expressed in texts and whether the expressions indicate positive or negative opinions toward the subject. For example, sentiment analysis of an online videogame can provide game designers with insights into what the users consider as favorable or unfavorable about the game [5,6].

Applications for processing large amounts of texts require NLP expertise. These include classifying text into categories, indexing, automatic translation, speech understanding, information extraction, automatic summarization, knowledge acquisition, games, opinion mining, spell-checking, context-based thesaurus, and genre-based word prediction, and text generations [7]. Some of NLP applications involve generation as well as analysis of the language. Over the years, a significant number of NLP software have been developed; some of these are available freely, while others are available commercially.

3. CHALLENGES

NLP systems usually require a huge amount of textual data which is often hindered by privacy and data protection issues. It is essential that personal/sensitive data is treated with confidentiality. Training speed and accuracy are two major challenges of large-scale natural language processing systems. It is nontrivial to improve both the training speed and the accuracy at the same time. A compromise may have to be made.

There is apparent tension that exists between determinability and uncertainty when computers process the natural language. While humans are often ambiguous and imprecise, computers require unambiguous and precise messages.

4. CONCLUSION

Natural language processing (NLP) is a popular technique of artificial intelligence for extracting the elements of concerns from raw plain text information. It deals with how to program computers to fruitfully process large amounts of natural language data. It is a multidisciplinary field that is related to linguistics, cognitive science, psychology, philosophy and logic. More information about NLP can be found in [8-11].

REFERENCES

[1] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," *IEEE Computational Intelligence Magazine*, May 2014, pp. 48-57.

[2] "Natural-language processing," *Wikipedia*, the free encyclopedia
https://en.wikipedia.org/wiki/Natural-language_processing

[3] J. Hirschberg, B. W. Ballard, and D. Hindle, "Natural language processing," *AT&T Technical Journal*, Jan./Feb. 1988, vol. 67, no. 1, 1988.

[4] C. Friedman, T. C. Rindflesch, and M. Corn, "Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine," *Journal of Biomedical Informatics*, vol. 46, 2013, pp. 765–773.

[5] T. Nasukawa and J. Yi, "Sentiment analysis: capturing favorability using natural language processing," *Proceedings of the 2nd International Conference on Knowledge capture*, October 2003, pp. 70-71.

[6] J. P. Zagal, N. Tomuro, and A. Shepitsen, "Natural language processing in game studies research: An overview," *Simulation & Gaming*, vol. 43, no. 3, 2012, pp. 356–373.

[7] S. M. Chandhana, "Natural language processing future," *Proceedings of International Conference on Optical Imaging Sensor and Security*, Tamil Nadu, India, July 2-3, 2013.

[8] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, 2009.

[9]J. Olive, C. Christianson, and J. McCary (eds.), *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*. Springer, 2011.

[10] C. D. Manning and H. Schutze, *Foundations of Statistical Natural Language Processing.* Cambridge, MA: The MIT Press, 1999.

[11] D. Jurafsky and J. H. Martin, *Speech and Language Processing.* Englewood Cliffs,NJ: Prentice Hall, 2000.

ABOUT THE AUTHORS

Matthew N.O. Sadiku , is a professor at Prairie View A&M University, Texas. He is the author of several books and papers. He is an IEEE fellow. His research interests include computational electromagnetics and computer networks.

Yu Zhou , is a doctoral student at Prairie View A&M University, Texas. He worked for Institute of Automation Shandong Academy of Sciences. His research interests are in natural language processing, image processing, machine learning, deep learning, and autonomous underwater vehicle design.

Sarhan M. Musa , is a professor in the Department of Engineering Technology at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow.

DOI: 10.31695/IJASRE.2018.32708