International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

The Effectiveness of Natural Language Processing (Nlp) as a Processing Solution and Semantic Improvement

Afif Badawi

Email: afifbadawi@dosen.pancabudi.ac.id
Fakultas Sains Dan Teknologi Universitas Pembangunan Pancabudi

ABSTRACT
The purpose of writing this article is to discuss the function of Natural Language Processing (NLP) in semantic improvement. The writing method uses literature related to NLP on semantic processing and refinement. The discussion in this paper shows that natural language processing helps computers communicate with humans in their own language and makes scaling other language-related tasks easier and more systematic. Because NLP includes lexical/scanner analysis, syntactic/parser analysis, semantic/translator analysis and pragmatic/evaluator analysis. Each component is a sequence of interrelated processes and requires a knowledge base to process a language. Lexical analysis requires knowledge of vocabulary (lexicon) to understand word formation. Syntax analysis requires knowledge of grammar rules (grammar) to understand the structure of a sentence. Semantic analysis requires knowledge of the meaning and meaning of words to understand the relationship between words and the meaning of a sentence. Pragmatic analysis requires knowledge of a concept to understand the relationship between language and the context in which it is used.

Keywords : Natural Language Processing; Semantic; Effectiveness.

INTRODUCTION

Natural language processing (NLP) is a branch of artificial intelligence that helps computers understand, interpret, and manipulate human language. Computer science is closely related and has become the main attraction for applying NLP, including computational linguistics, in its efforts to fill the gap between human communication and computer understanding (Wangsanegara & Subaeki, 2015). Natural Language Processing is not a new thing or a new science, because this technology is developing very rapidly and continues to increase in line with the interest in human communication to machine tools, then it is also supported in terms of the availability of big data, advanced computing, and algorithms that are constantly being refined. In the process of processing, communication occurs from humans, then they can speak and write in English, Spanish or Chinese. However, the natural language of computers known as machine code or machine language is largely incomprehensible to most people (Admojo, 2018).

As in history, programmers used punch cards to communicate with the first computers 70 years ago. This difficult guide and process is understood by only a small number of people. Now you can say, “Alexa, I love this song,” and the device that plays music in your home will turn down the volume and reply, “Okay. Ranks saved,” in a human-like voice. Then the device adjusts its algorithm to play that song and other people like it the next time you listen to that music station (July, 2020). Natural language processing helps computers communicate with humans in their own language and scale other language-related tasks. For example, NLP allows computers to read text, hear speech, interpret it, measure sentiment, and determine which parts are important. Today's machines can analyze more language-based data than humans, tirelessly

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

and in a consistent and unbiased way. Given the amount of unstructured data that is generated every day, from medical records to social media, automation will be essential to fully analyze text and speech data efficiently.

Human language is very, very complex and diverse. We express ourselves in limitless ways, both verbally and in writing. Not only are there hundreds of languages and dialects, but within each language there is also a set of grammatical and syntactic rules, terms, and slang words. When we write, we often misspelt or abbreviate words, or omit punctuation. When we speak, we have regional dialects, and we sometimes mumble, stutter, and borrow terms from other languages (Admojo, 2018). While supervised and unsupervised learning, and in particular deep learning, are now widely used for human language modeling, there is also a need for syntactic and semantic understanding and domain expertise that is not always present in these machine learning approaches. NLP is very important because it helps resolve ambiguities in language and adds useful numerical structures to data for many downstream applications, such as speech recognition or text analysis (Industry, 2001).

Natural language processing includes a variety of language techniques for interpreting human language, ranging from statistical methods and machine learning to rule-based and algorithmic approaches. We need a broad range of approaches because text and voice-based data vary widely, as do practical applications. Basic NLP tasks include tokenization and parsing, lemmatization/stemming, part-of-speech tagging, language detection, and identification of semantic relationships. In general, the task of NLP is to break down language into shorter chunks of elements, try to understand the relationships between the pieces, and explore how the pieces work together to create meaning (Xxxix, 2017). Because the capabilities and advantages of using NLP make the use of NLP a solution in processing and repairing semantics because of errors that may occur a lot, so the role of NLP is needed as a process to correct typing errors. On this occasion, we will discuss the Effectiveness of NLP as a Processing Solution and Semantic Improvement.

LITERATURE REVIEW

The history of Natural Language Processing (NLP) dates back to the 1950s, although there has been research on NLP in earlier years. In 1950, Alan Turing (father of computer science) published his famous article entitled "Computing Machinery and Intelligence" in which Alan Turing proposed a test which is now known as the Turing Test. Turing test is] a test that measures the ability of a machine (in this case a computer program) to exhibit intelligent behavior (July, 2020).The existence of NLP helps computers to communicate with humans in their own language and scale other language-related tasks. For example, NLP allows computers to read text, hear speech, interpret it, measure sentiment, and determine which parts are important. Furthermore, currently machines can analyze language-based data more than humans, without fatigue and in a consistent way (Putri & Hendrowati, 2018).

There are various techniques in natural language processing to be interpreted into human language. Ranging from statistical methods to machine learning to rule-based and algorithmic approaches. This also requires a broad range of approaches because the data available is text-based as well as voice-based, as well as practical applications. Natural Language Processing (NLP) is a computer science that discusses the interaction between computers and human

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

language or natural language. Natural language is generally a form of a message to be conveyed through communication between humans. The main form of natural language is in the form of sound or speech (spoken language), and can also be in the form of writing. Natural language processing is also a branch of artificial intelligence (Artificial Intelligence) and its field of study intersects with computational linguistics (language). Studies of natural language processing include speech segmentation, text segmentation, part-of-speech tagging, and word sense disambiguation (Andayu, 2013).

The purpose of humans learning about language processing is to be able to carry out the process of making computational models from language, which can make a machine able to understand and understand the meaning of human language (natural language) then the machine can provide appropriate and appropriate responses, so that it can happen. The interaction between computers and humans and their intermediaries is natural language.

The following are areas related to natural language processing, including the following:

a. Phonetics and Phonology
   Phonetics and phonology are fields that deal with sound processing that can produce recognizable words. Its use is in applications that use the speech-based system method.
b. Morphology
   Morphology is a field of knowledge about words and their forms that are used to distinguish one word from another, and there is a separation between words and other elements (punctuation marks).
c. Syntax
   Syntax is a field that includes an understanding of word order and the formation of words into a sentence, and includes the relationship between words in the process of changing the form of a sentence into a systematic thing.
d. Semantics
   Semantics is a field that maps the form of syntactic structures by utilizing each word into a more basic form without being affected by the structure of a sentence.
e. Pragmatics
   Pragmatics is a field that deals with the level of knowledge from each different context and is conditioned on the situation and purpose of making the system.
f. Discourse knowledge
   Discourse knowledge is a field that introduces whether a sentence that has been read and recognized before affects the meaning of the next sentence. Information is known to perform meaning processing of personal pronouns and to interpret temporary aspects of information.
g. world knowledge
   World knowledge is a field that includes the meaning of a word in general and specifically in conversation (Andayu, 2013).

METHODS

The research methodology of this article uses literature study data and references that are in accordance with the discussion of NLP in semantic processing. This article provides an

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

illustration that NLP provides effectiveness in semantics that facilitates human programming languages for computers as media.

RESULTS AND DISCUSSION

Results

Natural Language Processing (NLP) technology is a technology that allows performing various natural language processing commonly used by humans. This system usually has input and output in the form of text. Natural language processing has many applications and below are some of the categories of applications as follows (Maghfira et al., 2017):

1. Natural Language Translator. Translator that translates one natural language into another natural language, for example Indonesian to English translator, Indonesian to Mandarin and other examples. A natural language translator is not only a dictionary that translates word for word but must be able to translate the syntax from the origin language to the destination language.
2. Natural language translators to artificial languages, namely translators where a machine or computer is able to convert commands in natural language into artificial language. Example: a translator that allows us to give natural language commands to the computer. With a system like this, system users can give commands in everyday language.
3. Text Summarization is a system that can "make a summary" of important things from a given sentence.

   The types of applications found in the field of natural language processing include:

   a) Text-based application

      Text-based applications are various kinds of applications that can process written text such as to find the main topic of a book, search for the contents of e-mails, translate a document and one language into another.

   b) Dialogue-base application

      Dialogue-based application is an approach that includes spoken language such as speech recognition, as well as interactions carried out by inputting or entering text via the keyboard. Applications applied to the dialogue-base: (i) a question and answer system, where natural language processing is used to obtain information from the database; (ii) an automated telephone service system; (iii) voice control on electronic devices; (iv) a problem solving system helps to solve problems that are commonly encountered in a job (Akhir, 2017).

NLP has a role in reducing ambiguity problems that exist in Indonesian texts. The stage in NLP that plays a role in dealing with this ambiguity is Part-of-Speech (POS) Tagging. POS tagging is one of the stages of the NLP study to determine word class. The results of POS tagging on documents can be used as a basis for research in other Natural Language Processing, such as: Language Generator, Information Retrieval, Text Summarization, Question and Answering, and Machine Translation (Yazid & Fatwanto, 2018).

In performing natural language processing, the system requires several components to analyze the meaning of sentences, including the following:

1) Parsers. It is a system that takes words that have been inputted by the user, then group them based on grammatical;

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

2) Knowledge Representation System. The system analyzes these word sets to determine their meaning;
3) Translator Output. After doing the analysis, the system gives a reply to the user. Output can be natural language or other output;
4) Pattern Matching. Pattern Matching is a pattern matching method used in artificial intelligence. In computer science, pattern matching is used to check the order of tokens in a sentence. One of the uses of pattern matching is to find matching pattern components, then replace them with other token sequences. An example of pattern matching is a sequence pattern. In sequence pattern, pattern search occurs using regular expressions (regex) and backtracking algorithms. In addition to regular expressions, the search for sentence patterns can be done with tree patterns. Tree patterns are used in several programming languages as a common tool for processing data based on their structure, for example, Haskell, ML, Scala and Mathematica. These programming languages have a special syntax for expressing tree patterns and language constructs for conditional execution and retrieval of values based on them. In terms of simplicity and efficiency, the use of regular expressions (regex) is still better than tree patterns (Aditama, 2020).

Discussion

The process components of NLP include lexical analysis/scanner, syntactic analysis/parser, semantic analysis/translator and pragmatic analysis/evaluator. Each component is a sequence of interrelated processes and requires a knowledge base to process a language. Lexical analysis requires knowledge of vocabulary (lexicon) to understand word formation. Syntax analysis requires knowledge of grammar rules (grammar) to understand the structure of a sentence. Semantic analysis requires knowledge of the meaning and meaning of words to understand the relationship between words and the meaning of a sentence. Pragmatic analysis requires knowledge of a concept to understand the relationship between language and its context of use (Restina et al., 2019).

Limitations in representing knowledge using relational databases can be overcome by using ontology data modeling. Ontology can describe the form of a hierarchical structure containing class definitions, relationships between entities, characteristics or properties and rules that apply in a field of knowledge, meaning that by using an ontology model a stored data can be viewed as a concrete form of knowledge because it has meaning, mutual related and have a relationship with other data. Research on natural language processing using semantic technology develops a web-based book catalog search application. The representation of library knowledge follows the standard RDF Simple Knowledge Organization System (SKOS) structure (Restina et al., 2019).

Simple command sentence processing based on lexical analysis and syntactic analysis using predefined sentence patterns without any semantic analysis process. Then those who developed a question answering (QA) system used two ontologies as a source of knowledge, namely ontolingua which was a representation of language knowledge based on standard Indonesian grammar rules, while ontologies which stored library knowledge were developed using the SKOS concept. Natural Language Processing (NLP) ability to understand complex

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

interrogative sentences both syntactically and semantically and able to detect sentences using Indonesian grammar rules (Admojo, 2018).

Two ontologies were also used which developed a semantic-based information retrieval system with natural language input. Two ontologies used are language ontology as knowledge for natural language processing and mountain ontology as knowledge for presenting information related to mountains and climbing. Language processing using ontology is able to understand natural language using syntactic analysis and semantic analysis before taking information from the mountain ontology. Information search can be carried out more dynamically based on user perceptions using keywords, phrases, clauses, command sentences or complex interrogative sentences so that users can search for information content according to the desired context (Admojo, 2018).

Semantic interpretation aims to translate sentences into a form of representation called logical form, while contextual interpretation will translate logical form into another form of representation called knowledge representation or final meaning representation. The input for the semantic interpretation process is the parse tree, while the output is an appropriate representation in the sense that it can be processed and understood by the computer. Commonly used representations are logical forms. Actually, even though the text is a form of representation, it is still in a form that cannot be understood by computers (Industry, 2001).

Underlying tasks are often used in higher-level NLP capabilities, such as the following:

1) Categorization of content. Linguistically based document summaries, including search and indexing, content notification and duplication detection.
2) Topic discovery and modeling. Accurately capture meaning and themes in text sets, and apply advanced analytics to text, such as optimization and forecasting.
3) Contextual extraction. Automatically pull structured information from text-based sources.
4) Sentiment analysis. Identifying subjective moods or opinions in a large number of texts, including average sentiment and opinion mining.
5) Speech-to-text and text-to-speech conversion. Transforms voice commands into written text, and vice versa.
6) Document summary. Automatically generates synopsis of large sets of text.
7) Machine translation. Automatic translation of text or speech from one language to another.

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

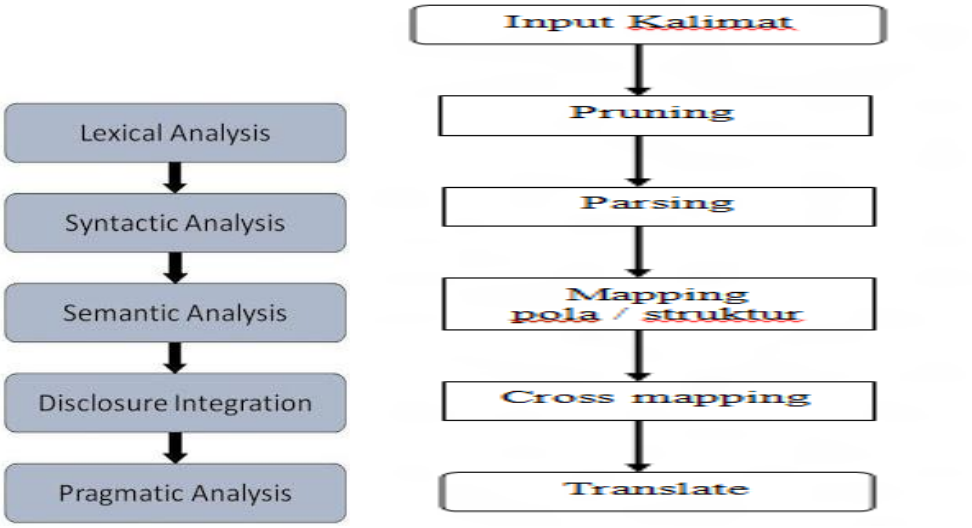

Figure 1. NLP Processing

Today, companies engaged in every field must be able to take advantage of what potential to be a source of important assets. However, it is challenging to process Natural Language Processing (NLP) the prospects are very high but it also depends on creativity. Regardless of the field at hand, every business today relies on large amounts of information in text. For example, a law firm works with a large number of research results, past and ongoing legal transaction documents. Notes, e-mail correspondence, as well as government information and special references in large sizes too. Because these types of information are mostly composed of language, NLP is needed in it.

Natural Language Processing (NLP) is the most important component in text mining and a subfield of artificial intelligence (AI) and computational linguistics. This science studies about how to understand natural human language. By changing the description of human language (such as text documents) into a more formal presentation (in the form of numerical and symbolic data). It is hoped that later it will be easier to be manipulated by computer programs. The goal of NLP itself is to go beyond syntax-based text manipulation (which is often referred to as 'word counting'). Becoming a true understanding and processing natural language that takes into account various semantic and grammatical constraints as well as context. NLP has also been successfully applied to various jobs through computer programs to automatically process human language which previously could only be done by humans. The success and popularity of text mining is very dependent on the development of NLP both in the process of generating and understanding human language. NLP allows the extraction of various features from unstructured text.

Text mining is a technology that seeks to extract useful information from unstructured textual data. Text mining is an extension of data mining (data mining) to textual data. This information is obtained by formulating and developing patterns and trends through statistical pattern learning. Text mining includes the process of structuring the input text (parsing, with the addition of some features that arise from the linguistic side and the reduction of some other

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

features), obtaining patterns from structured data, and evaluating and interpreting the output. Information can be said to be of high quality if it contains a combination of relevance and novelty. Common text mining steps include text categorization, text grouping (clustering), concept or entity extraction, sentiment analysis, document summarization and entity and relationship modeling. The bag-of-words (BoW) model is a simple representation used in natural language processing (NLP) and information retrieval (IR), also known as a vector space model. In this model, a text in the form of a sentence or a document is represented as a multiset bag of words contained in it, regardless of word order and grammar but still maintains its diversity.

CONCLUSION

Natural Language Processing (NLP) in semantic processing is proven to be used to determine the focus of words in a phrase or sentence by considering the location of the emphasis of the word. The process of making speech computational models proves that NLP can be used to research language. The utterances produced by the speakers in this study do not have a fixed pattern when compared to the spontaneous utterances that are the research reference. This may be influenced by linguistic knowledge that the stress in Indonesian is flat and there is no particular protrusion of speech. Therefore, the language computational model can be used to make a rule regarding how to determine the contour and stress pattern in speech by considering the segmental aspect as well as the suprasegmental aspect.

REFERENCES

Aditama, K. 2020. Pemanfaatan Natural Language Processing Dan Pattern Matching Dalam Pembelajaran Melalui Guru Virtual. *Elkom: Jurnal Elektronika Dan Komputer*, *13*(1), 121–133. https://journal.stekom.ac.id/index.php/elkom/article/view/187

Admojo, F. T. 2018. *Ontologi Bahasa Sebagai Basis Pengetahuan untuk Pemrosesan Bahasa Alami*. *08*(02).

Akhir, T. 2017. *Implementasi Natural Language Processing Untuk Mengurangi Risiko Terabaikannya Calon Pelanggan Pada Pt . Shafira Tour & Travel*.

Andayu, N. P. 2013. Perancangan Text To Speech Converter Engine Dalam Pengucapan Kata Berbahasa Arab Sehari-Hari. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, *1*(3), 144–149.

Industri, F. T. 2001. Studi Analisis Metode-Metode Parsing Dan Interpretasi Semantik Pada Natural Language Processing. *Jurnal Informatika*, *2*(1), 13-22–22. https://doi.org/10.9744/informatika.2.1.pp.13-22

Juli, E. (2020). *Natural languange processing*.

Maghfira, T., Cholissodin, I., & Widodo, A. 2017. Deteksi Kesalahan Ejaan dan Penentuan Rekomendasi Koreksi Kata yang Tepat Pada Dokumen Jurnal JTIIK Menggunakan Dictionary Lookup dan Damerau-Levenshtein Distance. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(6), 498–506.

Putri, W. T. H., & Hendrowati, R. 2018. Penggalian Teks Dengan Model Bag of Words Terhadap Data Twitter. *Jurnal Muara Sains, Teknologi, Kedokteran, Dan Ilmu Kesehatan*, *2*(1), 129–138.

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 1 page 36 - 44

Restina, D. J., Bijaksana, M. A., & Lhaksmana, K. M. 2019. Pembangunan Synonym Set untuk WordNet Bahasa Indonesia dengan Menggunakan Metode Komutatif. *Ind. Journal on Computing*, *4*(2), 147–156. https://doi.org/10.21108/indojc.2019.4.2.334

Wangsanegara, N. K., & Subaeki, B. 2015. *IMPLEMENTASI NATURAL LANGUAGE PROCESSING DALAM PENGUKURAN KETEPATAN EJAAN YANG DISEMPURNAKAN (EYD) PADA ABSTRAK SKRIPSI MENGGUNAKAN ALGORITMA FUZZY LOGIC Jurusan Teknik Informatika , Fakultas Sains dan Teknologi UIN Sunan Gunung Djati Bandung Ejaan yang*. *8*(2).

Xxxix, P. (2017). *PIBSI XXXIX, Semarang 7-8 November 2017*. *November*, 286–302.

Yazid, A. S., & Fatwanto, A. 2018. Penentuan Kelas Kata Pada Part of Speech Tagging Kata Ambigu Bahasa Indonesia. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *2*(3), 157. https://doi.org/10.14421/jiska.2018.23-05