IJASRM International Journal of Advanced Scientific Research and Management, Volume 3 Issue 8, Aug 2018
www.ijasrm.com
ISSN 2455-6378

Natural Language Processing: Evolution, Methodology, Problems and Applications in Real World

Garima Tyagi[1]

[1] Career Point University,
Kota (Raj)

Abstract

Natural language processing is a stream of artificial intelligence that has attracted many researchers and scientists from different fields. These fields are related with the computational linguistics, statistics and computer science engineering. The main goal of Natural Language Processing is to develop computational models for language analysis and generation. Certain methodologies are used in natural language processing for sentence analysis and some of them are widely accepted. This paper is focusing on the best methodology that can be used for sentence analysis in NLP; knowledge based system for multilingual text classification; problems in natural language processing and the applications of natural language processing in real world.

Keywords: *information extraction, linguistic, multilingual text, Natural language processing, NLP, Machine Translation*

1. Introduction

Natural languages are defined as the languages those are naturally evolved and used for communication purposes e.g. Hindi, English, French, Spanish, German etc. The tern natural languages is often used to define languages those are used by the human for communication purpose and used to distinguish them with computer languages. Natural language processing is a field of study that is related with the scientific study of natural languages and concerned with the interaction of computer with languages. Sometimes NLP can also be explained as the extension of computer linguistics.

In NLP different techniques are used so that the computer performs in such a way that it can understand the commands issued in natural languages and perform the operations accordingly. Natural Language Processing is the area that is basically used to specify that how computers can be used to understand natural languages for different purposes. NLP researchers are using the techniques to find that how a human can understand and using the language, so that different tools and techniques can be developed (Chawdhary 2003). For the development of such type of systems a series of certain activities to be done. These activities may include –

1. Development of Natural Language Text Processing System
2. Development of Natural Language Interface
3. Evaluation of the System

The main aim of Natural language processing is to develop tools for processing natural languages. A large amount of Natural language processing tools and services are available. These tools and services are available in the current world and there is also the emergence of new one, many web services are also available for facilitates the working. These tools and services are available for language detection, text classification, morphological analysis etc. (Hellmann, et al. 2013) These tools have to be judged using the programming interface.

2. Evolution

As we all know NLP comes in comparatively new era and it involves various development stages –

It is basically started with the with the development of automation in 1950s with the advent of Turing's model for computation (Kumar 2012). With the development of Turing's model, different other models have been developed, these models define the field of formal language theory including context-free grammar.

After this in early 1960's some natural language system were discussed, these were the simple systems

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 8, Aug 2018
www.ijasrm.com
ISSN 2455-6378

working on the combination of pattern matching and keyword search using the heuristic search method.
In late 1960's, the models for human language processing was developed, they were based on the ' transformational grammar ' and also on 'online corpus' Predicate logics were also used to develop and then by using the probabilistic and data driven models, natural processing systems were enhanced. Algorithms for different types of processing like parsing, tagging and disclosure processing were used and certain strategies were incorporated in NLP from speed recognition and information retrieval.

3. Early Systems

3. 1 ELIZA (Weizenbaum 1966) - It is one of the first natural language processing systems. It was developed in 1966. It was one of the most popular AI systems of that time. It carries a conversation with user via a terminal.

3.2 SysTrans – System Translation - This was the first machine translation system developed in 1968-69 by Dr. Peter Toma. It was originally used for Russian to English translation and was provided the first on-line machine translation service called Babel Fish for Yahoo. It was used by Google's language tools for a long time.

3.3 Lunar (Woods 1978) - It was the question answering system worked on artificial Intelligence technique. It was having a separate syntax analyzer and a semantic interpreter.

3.4 SHRDLU (Winograd, 1972) - It was developed in 1970's and it was a dialogue system that can converse with a human user. It uses syntactic parsing and semantic reasoning for understanding the instructions. The user can ask for the manipulation of blocks.

3.5 Taum Metero - It was a natural language processing system used in Canada for generating weather reports.

4. Problems and Challenges with NLP

There are number of factors those are responsible for making working with NLP so difficult. These are basically related with the representation and Other tasks that can be performed in NLP are –

Finding Parts of Text, To find the sentences, To find the different entities like person , place or thing, To detect the part of the speech, To classify the documents

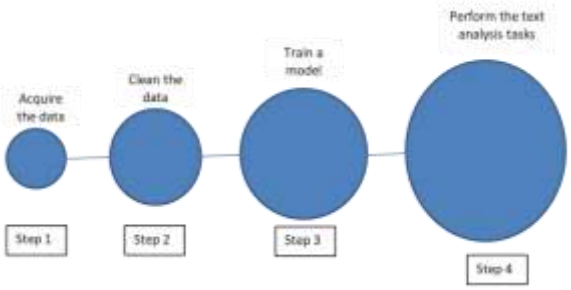
interpretation. In the field of computational linguistics the representation of the content should be précised. If the natural languages are used in the working with system, it may be ambiguous. It is very difficult to capture knowledge from one source and transferred to another.
The major issue with the natural language processing is dealing with semantics. As we know that the word, its composition in the sentence and the syntax and semantically composition of the sentence, gives it the complete meaning. Without it the sentence may not have a perfect meaning. The another problem with NLP is ambiguity and is firstly arises at word level as any word may have different meaning at different places. In certain cases the words used in the sentence are not ambiguous but the the sentence as whole may be ambiguous.

5. Methodology

Natural Language Processing is concerned with the analysis of text to enhance applications or to create applications. The NLP process typically consists of various steps including data acquisition, data preparation, creation of analysis model and then to perform the actual analysis.
NLP has been used for a variety of tasks including –

- Identify names and places with in the documents
- It is used in machine translation
- Summarizing the contents of text documents
- Determining the sentiments of the documents to check the review.
- Speech reorganization and synthesis

NLP steps can be defined as follows –



5.1 Out of the above tasks that can be performed by NLP, here it is explained that how we can deal with **finding the parts of the text**

It involves different steps to be performed –

1. Firstly the text is divided into sub parts those are called as tokens and for this process the tokenizes are involved.
2. Tokens are basically corresponds to certain words those are the parts of the text.
3. With the above process other tasks involved are –
    a. Stemming
    b. Lemmatization

5.2 Stemming

It is an important feature supported by search systems (Jivani 2011) and it is the process of finding the base of a word eg. the word walk and walking have the word walk as the base word . This process uses only the current word for analysis. The actual function of stemming is to reduce the word to its root.

5.3 Lemmatization

This is the process to find the normalized form of a word (Mladenic, Lavrac and Plisson n.d.). It determines the lemma, or the dictionary form of the word. It means lemmatizing deals with obtaining the 'lemma' of a word which involves reducing the word forms to its root form after understanding the POS and the context of the word in the given sentence. It also used to find that what is the actual meaning of the word because a word can be used for different ways in the text. It is basically uses the context of the word to perform the analysis.

- The process of finding sentences is known as **Sentence Boundary Disambiguation (SBD)**
- Another task that needs to be performed in NLP is **to identify entities** like person, place or thing etc. but due to ambiguity present in different languages make it difficult.
- It is very important to determine the grammatical elements of a sentence, and it is also useful for subsequent tasks. In this process the elements of a sentence are assigned witah tags to describe it, but it is not so easy as a sentence may have more than one set of tags.

5.4.Classifying documents

With the proliferation of documents generated either manually or automatically, the automatic classification of text is becoming increasingly important. Classification is basically used to sort the documents by type and also to find the ownership of the document. The documents may be classified on the basis of predetermined categories or sometimes there is a procedure to determine the category of the document, if it is not already present.

We can use sentiment analysis for classifying the documents.

Extending Relationship Between Textual Elements

A parser is used for extending the relationship between textual elements and it provides the types or tags to the given elements eg it may be a personal relationship like who is the mother and who is the daughter. It may be a professional relationship like CEO, manager or their sub ordinates and in certain cases a physical relationship like a car may have different elements.

To define the relationship between the elements a graphical representation may be done or defined .

6 Applications of NLP

NLP providing different computational techniques that can be used in learning, understanding and then producing natural language content. Previously it includes the analysis of structural language and then developing certain technologies like machine translation, speech recognition and speech synthesis (Hirschberg and Manning 2015). The first area of NLP applications is Machine Translation. The complete analysis of a natural language sentence and also the generation of output sentence using the linguistic analysis are done in Machine Translation. It is one of the most challenging tasks in the area of Artificial Intelligence.

Now a days different application have been used in information retrieval, information extraction, text summarization etc.

So the applications those are using NLP are as follows –

6.1 Machine Translation

- It can be defined as the translation of text from one natural language to another, For such type of translation deep knowledge of grammar, words and phrases of both the languages is necessary. Different levels of NLP have been utilized in Machine Translation systems, ranging from the word-

based approach to applications require higher of analysis. (Preeti and Sidhu 2013)

6.2 Speech Recognition

- It is a technique that includes the conversion of speech signals in the set of words. The problem associated with this technique is that there is a wide variation in the pronunciation of words and it may lead to ambiguity.

6.3 Speech Synthesis

- This is actually an opposite technique to speech recognition, here the speech is produced from a set of word of sentences and convert them into speech.

6.4 Information Retrieval

- This is related to identifying the relevant documents and it includes indexing of the words using stemming and phrase extraction. Thesaurus is an example of information retrieval.

6.5 Information Extraction

- This is the process in which factual information available in document is captured and returned . The information will be stored in the form of data in the database .

6.6 Text Summarization

- It includes the creation of summary of document and involves the semantic and syntactic level processing of text.

6.7 Dialogue Systems

- These systems developed for end users and these systems are generally answering machines. It includes telephone without an operator, teaching systems, voice controlled machines and general problem solving systems

Now a days different types of information retrieval systems and search engines were developed to search the relevant information from a huge amount of data available eg data available on Web. The applications those are required at that time are text extraction, question answering and text summarization, there all the applications are some how related with the information retrieval but mostly deal with NLP instead of Information retrieval.

Dialogue systems are a popular topic in NLP. Different spoken dialogue systems have developed for communication between humans or between humans and machine. These systems require tools for the automatic speech recognition, dialogue management and text-to speech synthesis, thus these tools are developed using NLP.

7. Conclusion

Natural Language Processing is concerned with number of applications and dealing with the development of models related with the processing of human languages. Different systems have been developed for NLP but as the natural languages are highly ambiguous and vague in nature , so it is very difficult to create a system that can perfectly done the processing of natural languages without an error. As there is the continuous development of NLP systems, but it is very complicated to develop a perfect one due to the complexity of human language. In the real world different applications like spoken dialogue system and translation system are developed. Now a day with the advent of web based models those are more user friendly, NLP plays an important role in the development of such models. Different types of models are used for processing the information quickly, so the main focus should be on removing the ambiguity in developing the applications, so that the perfect models can be developed.

References

[1] Chawdhary, Gobinda G. *Natural Language Processing.* 2003.
[2] Hellmann, Sebastian, Jens Lehmann, Soren Auer, and Martino Brummer. "Integrating NLP Using Linked Data." *The Semantic Web – ISWC.* 2013. 98-113.
[3] Hirschberg, Julia, and Christopher D. Manning. "Advances in Natural Language Processing." *Science*, 2015: 261-266.
[4] Jivani, Anjali Ganesh. "A Comparative Study of Stemming Algorithms." *IJCTA*, 2011.
[5] Kumar, Ela. *Natural Language Processing.* New Delhi: I.K. International Publishing House Pvt. Ltd., 2012.
[6] Mladenic, Dunja, Nada Lavrac, and Joël Plisson. "A Rule based Approach to Word Lemmatization." n.d.
[7] Preeti, and Brahmaleenr Kaur Sidhu. "Natural Language Processing." *INternational Journal of Computer Technology & Applications*, 2013: 751-758.

[8] Weizenbaum, Joseph. "ELIZA: A computer program for the study of Natural Language Communication between Men and Machine." *Communications of the ACM*, 1966: 36-43.

[9] Woods, W.A. "Semantics and Quantification in Question and Answerin." *Advances in Computers*, 1978: 1-87.