International Journal of Engineering & Technology, 7 (2.33) (2018) 452-455

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A literature survey on question answering system in natural language processing

A. Clementeena [1] *, Dr. P. Sripriya [2]

[1] Research Scholar, School of Computing Sciences, VISTAS
[2]Associate Professor, School of Computing Sciences, VISTAS
**Corresponding author E-mail: clementeena.phd@gmail.com*

Abstract

In order to respond to usual free form of questions that is contained in a large collection of informative texts or information. For this to hap-pen one must understand what type of questions are being imposed and thus to also know a few knowledge about the process. The con-straints will include the semantic and syntactic knowledge and framing that sort of question with possible correct answers. Thus this paper gives you a clear vision of how its being worked on question answering and thus it presents a machine learning approach. Also ways a important way of giving answers to closed domain in a different data set of know.

Keywords: *Natural Language Processing (NlP); Semantic Knowledge; Pragmatic Knowledge; Syntactic Knowledge.*

1. Introduction

Natural language processing conveys all about the computational linguistics or a theory or a work on NLP completely began more than 60 years ago. Many companies' approx. more than 500 companies they have just invested in voice that is speech systems. For example: AT &T has been performing in replacing 200 operators with voice recognition software. Verizon and BellSouth communications, they have already used this voice software that helps to control city regulations by the way is also used for sequencing out all the works during all the directory assistance calls.

2. Applications of NLP

2.1. Lexical semantics

Lexical meaning tells what is the computational meaning of format for the particular context.

2.2. Machine translation

Machine Translation mainly concentrates on the transformation of text from one language to another human language and it is a member of a group of problems which is termed as "AI-Complete". It is used widely to acquire more knowledge on the human like knowing out their semantics, facts and real happenings of the real world.

2.3. Named entity recognition

This terms out the particular match of the text to the proper names and tells what type of name it is and thus capitalization helps in finding out the named entities in languages like English. Some languages like German don't see the capitalization and also French and Spanish they do not capitalize their names as they serve as the adjectives.

2.4. Natural language processing

Converts the computer database texts into a human readable language.

2.5. Natural language understanding

Converts the blocks of data into understandable representation as First order logic structures which are very easy and fast for the computers to manipulate.

2.6. Optical character recognition

An image is represented in a text format from which to determine the corresponding formal text.

2.7. Question answering

Giving out a human-language question and giving a proper answer for it. Sometimes a specific question is asked and also sometime a open ended question can also be asked. Recent research works has given even more difficult questions.

2.8. Recognizing text entailment

There are two text fragments are given in which one text fragment is being true and the other entails false.

2.9. Relationship extraction

Given out blocks of texts and so identify the relationship among the named entities.

2.10. Sentiment analysis

It is widely in use of finding the future trends of the social media that helps in marketing field and also giving online reviews about people and other factors.

2.11. Topic recognition and segmentation

Given out a block of text in which it is separated into segments and then to find out the topic for each segment.

2.12. Word sense disambiguation

Many of the words in a block of paragraphs will have one meaning and thus we have to select the sensible context for the word. Nowadays many researchers are focusing more on topics like natural language processing, natural language understanding, speech recognition or voice recognition, translation of machine language words and checking of grammatical options.

3. Methodology used in NLP

The methodology that is used in processing NLP is that normal IPO method and that is elaborately explained as input process output and the flowchart of the process is given as follows.

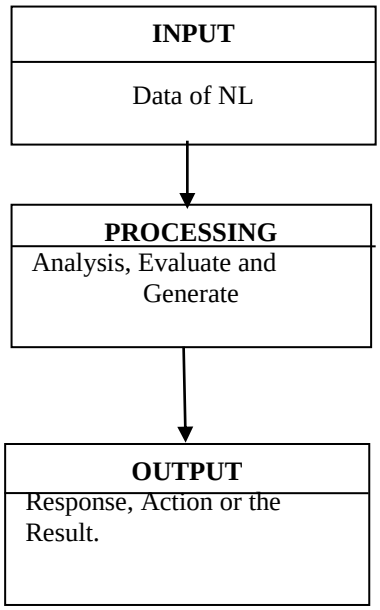

Fig. 1: Methodology in NLP.

4. Knowledge discovery in NLP

Knowledge understands of something such as information, facts, descriptions or skills. Knowledge can be defined as the practical understanding of a subject. The levels of knowledge can be categorized as many.

4.1. Phonetic and phonological knowledge

This type of knowledge tells how the words are realized in manual way as sounds

4.2. Morphological knowledge

This type of knowledge tells how the words are brought out from basic and meaningful units that are also known as phonemes.

4.3. Syntactic knowledge

This type of knowledge tells how the words and letters that can be clubbed together in a unique sentences.

4.4. Semantic knowledge

This tells us that some of the related words and also tell us how and in what way these meanings extraction combine the sentences to form meanings for those sentences.

4.5. Pragmatic and discourse knowledge

This concern more on how those sentences are incorporated in different context which mainly affects the sentence interpretation.

4.6. World knowledge

Knowledge includes the current affairs about the feature and outline of the world.

5. Issue of ambiguity

- Issues mainly rely on syntactical and morphological ambiguity.
- Issues also happens on semantic knowledge by giving an example as [Make a verb or verb that can be either “create” or “cook”].
- It can be solved by parts-of-speech (POS).
- It can also be solved by words

6. Semantic knowledge

Semantic knowledge is the set of practises that can classify content as it can be accessed at any time and delivered in a stipulated time. Semantic knowledge is all about representing the knowledge or data about organisations. Some of the specific practises are maintaining the cluster documents into a single document, converting the data’s or contents into various file formats, any keyword search for a particular word or does it very effectively, involving more context based contents in creation of that content, reducing the rates of the production cost, reduce the production costs for the file formats, reduce the manual representation of forming right data knowledge at right time. This knowledge gives you the study of meaning of linguistic utterances. This can also beused as to know well versed about how the meaning of the happening is closely related to meaning of words, phrases and then morphemes. Their lies many issues in it as with knowledge based representation of semantic knowledge as Verifiability, Unambiguous representation and vagueness. Notable NLP systems /prototype are ELIZA and LUNAR.

6.1. ELIZA

ELIZA is the simple computer which is induced to be the study in the natural language. Main feature of it is that it can repeat everything when the user asks in the form of a question.

6.2. LUNAR

LUNAR is formed by NASA that allows the geologists to make more and more questions about the chemical composition of lunar rock and soil samples

7. Syntactic knowledge

Syntactic knowledge is about describing how the individual words can be clubbed into a proper meaningful sentences, phrases and utterances. Computers are actually very much speed in their work and it’s also the most powerful machines which can sum up and execute the text that is written by human minds in a different way

and treating them in a different way. The main goal of syntactic analysis is to find whether the given data in an input form is a perfect sentence in a given natural language. Some description of syntactic structure of a sentence is given as an illustrations in the form of derivation tree and such formalizations are mainly aiming at more focus on the understanding of the computers. Also finding and understanding the relationship between the words and also between the corresponding people, action and things to be done.
Syntactic analysis can be mainly focussed on developing on punctuation sentences and also dialogue series with having natural language interface or as also a building block in a machine translation system. Czech is the language exhibiting rich inflection which requires much knowledge in grammar rules than any other languages. According to the research this syntactic kind of understanding is very tough and hard to analyze. The NLP laboratory is currently on developing site of synt, syntactic analyzer. According to the test performance done the synt reaches the recall of 92% and precision of about 84% and mainly visualizing all types of derivation tree. This knowledge gives you the clear cut vision of formal relationships between the words. In 1956, famous linguist, Noam Chomsky who first created the context-free grammar parses trees.
The most common methods or ways in which parsing is done they are:

- Top-Down parsing→Parsing is done from the root node to the base of the leaves
- Bottom-Up parsing→This kind of parsing starts with input as words and tries to build the words from those tree forms and is done by applying the rules and norms of the grammar which can be done one at a time.
- Depth-First Parsing→This kind of parsing expands the search space incrementally by one state at a time.
- Repeated Parsing of Sub trees→this kind of parsing deals with the findings of inefficiency of other parsing algorithms.
- Dynamic Programming Parsing Algorithms→this kind of parsing is used to rely on resolving ambiguity.

8. Pragmatic knowledge

Pragmatic knowledge is the concept of how the individuals can communicate meanings for the words and how can them produce the sentences or utterances in a proper way. Pragmatic knowledge includes social way of delivery and also functional knowledge. Pragmatics is the Sub field containing of two major parts as linguistics and semiotics. This pragmatics contributes the context content to the defined meaning. Pragmatics encompasses the speech act theory, talk in interaction and also other approaches like sociology, philosophy, anthropology and linguistics. For example in the case of semantics which examines meaning that is conventional or that is coded in a given language, and pragmatics study of how the transmission of meaning depends not only on structural and linguistic knowledge. Pragmatics explains how language users are able to overcome all the issues and since it relies only on the utterances, manner, time and place.

9. Question answering system

9.1. Introduction

Giving out a human-language question and giving a proper answer for it. Sometimes a specific question is asked and also sometime a open ended question can also be asked. Recent research works has given even more difficult questions.

9.2. Literature survey on question answering system

Question answering application is the computer science field in which collective to natural language processing and information retrieval. Question Answering goes with building defined systems in which it answers the question that is posted by the humans in a natural language format. In implementation process, a computer program usually generates its own answers by asking or querying a structure database of information or knowledge which is usually termed as knowledge base. Sometimes even it can take out some answers from some unstructured dataset collection of information or knowledge. Some of samples collected for natural language document as:

- A bunch of local collection of reference texts that is so important
- Internal documents and regarding WebPages that is being collected.
- Compiled newswire reports.
- A set of Wikipedia pages.
- A subset of World Wide Web pages.

In early stages there were two QA systems namely BASEBALL and LUNAR. This BASEBALL answered more questions for a duration of one year. Thus, LUNAR answered the geological analysis of rocks that was returned by Apollo moon missions. Both the QA systems were very useful and effective. In 1971, LUNAR was demonstrated by the lunar science and it was able to answer 90% of the question that has been asked.Then the language abilities of BASEBALL and LUNAR used similar ways like ELIZA and DOCTOR. These were the first chatterbox programs. SHRDLU was the highly effective and very familiar question answering which was formed by terry wino guard in the late 60's and early 70s and it is mainly concerned with the operation of robots in the toy world.
In 1970s knowledge base were developed to stream the domains of knowledge. Question answering system develops a interface with these expert systems. These expert systems resemble the modern QA system but except the internal architecture. In 1970s and 1980s formed the computational linguistics which led to more useful and effective projects in the field of question answering and comprehension of text. Recent QA system EAGLI has been developed for health and life benefits.

9.3. Question answering methods

QA is very much dependent on the good search called corpus and thus the larger the collection data size will be giving out a better performance. The focus of the data redundancy in a vast collection which means the small bits of information or data are to phrased in different manner in different ways as their By having all the right information that appears in many forms and thus it forms the too much work on the QA system to do the bigger or complex works that has to be understood and also Having the correct values nearby then the incorrect values can be corrected and can be rectified. The QA system completely rely on the reasoning power and also there are many number of question answering systems which is designed and developed in prologue which is a logic programming language that is linked with artificial intelligence.

9.4. Types of QA system

9.4.1. Closed domain question answering

It deals with particular domain or topic. It can be an easier task of asking questions and getting answers because NLP systems are very good at finding specific topic questions and retrieving answers will be very easy. Closed domain can work out very fast where the questions are on to a specific node so that it can retrieve the answers for the question that is being asked. Question asking for descriptive will be more effective than procedural ones. Even machine reading applications are also created in the medical field.eg: Alzheimer's disease.

9.4.2. Open domain question answering

In Information retrieval, it tells that in an open domain QA system goals at giving out the proper answers for the intended questions

asked by the user. Thus the returned text is given out in the form of short fragmented text and not in the form of documents. Thus the QA system finds the answers out of the help with the information retrieval, computational linguistics and knowledge representation of the data for finding those answers for the questions given. The QA system takes the natural language question as a input than the group of keywords.

Thus having the main input as the natural language question it is more easy and user-friendly but it's tougher to implement as this issue there are many number of question types and it is difficult for the system to find the correct one out of it in regards to give the appropriate correct answer. Initiating each question type to a question is a critical and a hard job to do. The whole retrieving process lies on finding the correct question type and which gives the correct answer type that helps in giving the correct appropriate value.

- As the name relies it tells that question and answering can be an open ended and can be nearly any kind of domain or topics.
- And so these domain questioning will have a large set of data base and thus the answers can be retrieved at any time.

10. Conclusion

This paper jots out the machine learning approach to question answering system. This paper shows the question answering in a closed domain that makes the listener to easily answer those respected questions. Thus closed domain question answering approach is highly intact with the users. In future we work to inhibit abundant knowledge of semantic analysis and its importance and to resolve some other issues and difficulties that is happened in those days and will definitely overcome all those issues by upcoming renders.

References

[1] Chowdhury, Gobinda G. "Natural language processing." Annual review of information science and technology 37.1 (2003): 51-89.

[2] Liddy, Elizabeth D. "Natural language processing." (2001).

[3] Spyns, P. "Natural language processing." Methods of information in medicine35.4 (1996): 285-301.

[4] Grishman, Ralph. "Natural language processing." Journal of the Association for Information Science and Technology 35.5 (1984): 291-296.

[5] Hirschman, Lynette, and Robert Gaizauskas. "Natural language question answering: the view from here." natural language engineering 7.4 (2001): 275-300.

[6] Burke, Robin D., et al. "Question answering from frequently asked question files: Experiences with the faq finder system." AI magazine 18.2 (1997): 57.

[7] Lehnert, Wendy G. Strategies for natural language processing. Psychology Press, 2014.

[8] Covington, Michael A. Natural language processing for Prolog programmers. Englewood Cliffs (NJ): Prentice hall, 1994.

[9] Lehnert, Wendy G. The Process of Question Answering. No. RR-88. YALE UNIV NEW HAVEN CONN DEPT OF COMPUTER SCIENCE, 1977.

[10] Moldovan, Dan, et al. "The structure and performance of an open-domain question answering system." Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2000.

[11] Sowa, John F. Knowledge representation: logical, philosophical, and computational foundations. Vol. 13. Pacific Grove: Brooks/Cole, 2000.