Summary: Natural-Language Processing, Parsing and Generation

Author: Oisin Farrell and Tobias Budig.

Date: 7th December 2020.

Natural Language Processing:

Natural Language Processing, usually shortened as NLP, is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language. The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable. Most NLP techniques rely on machine learning to derive meaning from human languages [1].

Parsing:

Parsing usually applies to text - the act of reading text and converting it into a more useful in-memory format, "understanding" what it means to some extent [2]. Parsing, in the context of NLP refers to the ability to read some information and break it down into logical pieces so that the system can make use of it. In order to parse some data correctly we first must introduce POS tagging. POS tagging is the process of marking up a word in a corpus to a corresponding part of a speech tag, based on its context and definition. This task is not straightforward, as a particular word may have a different part of speech based on the context in which the word is used [3].

Table

Description automatically generated

Figure 1: POS tagging [4].

Above, In figure 1, you can see how the tags are allocated. Sentences are broken down into nouns, verbs, adjectives and adverbs. This is useful as it gives the parsed data context.

Generation:

Generation is the more difficult area of NLP. It refers to the creation of natural language by the system. An example of where we encounter Natural language generation (NLG) is when a chatbot deciphers what we have said and produces a valid response.

Questions:

1. What are some of the components of NLP?

* Entity extraction: It involves segmenting a sentence to identify and extract entities, such as a person (real or fictional), organization, geographies, events, etc.
* Syntactic analysis: It refers to the proper ordering of words.
* Pragmatic analysis: Pragmatic Analysis is part of the process of extracting information from text [7].

1. What is POS tagging?

POS tagging is the process of marking up a word in a corpus to a corresponding part of a speech tag, based on its context and definition. This task is not straightforward, as a particular word may have a different part of speech based on the context in which the word is used [3].

1. What is regular grammar?

Regular grammar is used to represent a regular language.

A regular grammar comprises rules in the form of A -> a, A -> aB, and many more. The rules help detect and analyze strings by automated computation.

Regular grammar consists of four tuples:

* ‘N’ is used to represent the non-terminal set.
* ‘∑’ represents the set of terminals.
* ‘P’ stands for the set of productions.
* ‘S € N’ denotes the start of non-terminal.

1. What is a regular expression?

A regular expression is used to match and tag words. It consists of a series of characters for matching strings.

Suppose, if A and B are regular expressions, then the following are true for them:

* If {ɛ} is a regular language, then ɛ is a regular expression for it.
* If A and B are regular expressions, then A + B is also a regular expression within the language {A, B}.
* If A and B are regular expressions, then the concatenation of A and B (A.B) is a regular expression.
* If A is a regular expression, then A\* (A occurring multiple times) is also a regular expression.

1. What is semantic analysis?

Semantic analysis helps make a machine understand the meaning of a text. It uses various algorithms for the interpretation of words in sentences. It also helps understand the structure of a sentence.

Techniques used for semantic analysis are as given below:

* **Named entity recognition:** This is the process of information retrieval that helps identify entities such as the name of a person, organization, place, time, emotion, etc.
* **Word sense disambiguation:** It helps identify the sense of a word used in different sentences.
* **Natural language generation:** It is a process used by the software to convert the structured data into human spoken languages. By using NLG, organizations can automate content for custom reports.

References:

[1] [https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32 (Links to an external site.)](https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32)

[2] [https://stackoverflow.com/questions/1788796/what-is-parsing (Links to an external site.)](https://stackoverflow.com/questions/1788796/what-is-parsing)

[3] [https://medium.com/analytics-vidhya/pos-tagging-using-conditional-random-fields-92077e5eaa31#:~:text=POS%20tagging%20is%20the%20process,which%20the%20word%20is%20used (Links to an external site.)](https://medium.com/analytics-vidhya/pos-tagging-using-conditional-random-fields-92077e5eaa31#:~:text=POS%20tagging%20is%20the%20process,which%20the%20word%20is%20used).

[4] [https://www.google.com/search?q=pos+tagging&source=lnms&tbm=isch&sa=X&ved=2ahUKEwj81OX5yLvtAhVF-yoKHSHyD1QQ\_AUoAXoECBEQAw&biw=1396&bih=686#imgrc=wWynGWkl6DPmYM (Links to an external site.)](https://www.google.com/search?q=pos+tagging&source=lnms&tbm=isch&sa=X&ved=2ahUKEwj81OX5yLvtAhVF-yoKHSHyD1QQ_AUoAXoECBEQAw&biw=1396&bih=686#imgrc=wWynGWkl6DPmYM)

[5] [https://www.google.com/search?q=pos+tagging&source=lnms&tbm=isch&sa=X&ved=2ahUKEwj81OX5yLvtAhVF-yoKHSHyD1QQ\_AUoAXoECBEQAw&biw=1396&bih=686#imgrc=6fN0yB7PKtaoIM (Links to an external site.)](https://www.google.com/search?q=pos+tagging&source=lnms&tbm=isch&sa=X&ved=2ahUKEwj81OX5yLvtAhVF-yoKHSHyD1QQ_AUoAXoECBEQAw&biw=1396&bih=686#imgrc=6fN0yB7PKtaoIM)

[6] [https://intellipaat.com/blog/interview-question/nlp-interview-questions/ (Links to an external site.)](https://intellipaat.com/blog/interview-question/nlp-interview-questions/)

[7] [https://www.onlineinterviewquestions.com/nlp-int (Links to an external site.)](https://www.onlineinterviewquestions.com/nlp-int)