

Semantic Analysis

What Is Semantic Analysis?

- ▶ The purpose of semantic analysis is to draw exact meaning, or you can say dictionary meaning from the text. The work of semantic analyzer is to check the text for meaningfulness.

How is semantic analysis different from lexical analysis?

- ▶ Lexical analysis also deals with the meaning of the words, then Lexical analysis is based on smaller token but on the other side semantic analysis focuses on larger chunks.

Parts of Semantic Analysis

Semantic analysis can be divided into the following two parts –

- ▶ **Studying meaning of individual word**

It is the first part of the semantic analysis in which the study of the meaning of individual words is performed. This part is called lexical semantics.

- ▶ **Studying the combination of individual words**

In this second part, the individual words will be combined to provide meaning in sentences that leads to compositional semantics.

Lexical Semantics

- ▶ The first part of semantic analysis, studying the meaning of individual words is called lexical semantics.
- ▶ It includes words, sub-words, affixes (sub-units), compound words and phrases also. All the words, sub-words, etc. are collectively called lexical items.
- ▶ In other words, we can say that lexical semantics is the relationship between lexical items, meaning of sentences and syntax of sentence.

In order to understand the meaning of a sentence, the following are the major processes involved in Semantic Analysis:

1.Word Sense Disambiguation

2.Relationship Extraction

WSD (word sense disambiguation).

- Lexical ambiguity, syntactic or semantic, is one of the very first problem that any NLP system faces. Part-of-speech (POS) taggers with high level of accuracy can solve Word's syntactic ambiguity.
- On the other hand, the problem of resolving semantic ambiguity is called **WSD (word sense disambiguation)**.

Word Sense Disambiguation involves interpreting the meaning of a word based upon the context of its occurrence in a text.

Eg: Orange, Date, Bark, rock....

Consider the two examples of the distinct sense that exist for the word "*bass*" –

I can hear bass sound.

He likes to eat grilled bass

The occurrence of the word **bass** clearly denotes the distinct meaning. In the first sentence, it means frequency; in the second, it means **fish**. Hence, if WSD would disambiguate it then the correct meaning of the above sentences can be assigned as follows –

I can hear bass/frequency sound.

He likes to eat grilled bass/fish.

- ▶ WSD needs inputs like Dictionary and Text Corpus.
- ▶ Approaches includes
 - Dictionary-based or Knowledge-based Methods
 - Lesk Method
 - Supervised Methods
 - SVM, Naïve bayes....
 - Unsupervised Methods
 - CBOW, skip gram.....

Relationship Extraction

- ▶ Another important task involved in Semantic Analysis is Relationship Extracting. It involves firstly identifying various entities present in the sentence and then extracting the relationships between those entities.

- ▶ Includes

Named Entity Recognition

Semantic Role Labeling.

The general theory in compositional semantics: The meaning of a phrase is determined by combining the meanings of its subphrases, using rules which are driven by the syntactic structure.

Consider the following sentence and plausible representations of meaning.

Sentence: "John ate a ripe apple."

Syntax tree :

```
S ----> NP ----> Name ----> John
  |
  |--> VP ----> Verb ----> ate
        |
        |--> NP ----> Det ----> a
              |
              |--> Adj ----> ripe
              |
              |--> Noun ----> apple
```

NER(Named Entity Recognition)

- ▶ The named entity recognition (NER) is one of the most popular data preprocessing task. It involves the identification of key information in the text and classification into a set of predefined categories. An entity is basically the thing that is consistently talked about or refer to in the text.
- ▶ At its core, NLP is just a two-step process, below are the two steps that are involved:
 - Detecting the entities from the text
 - Classifying them into different categories

Some of the categories that are the most important architecture in NER such that:

- Person
- Organization
- Place/ location

Other common tasks include classifying of the following:

- date/time.
- expression
- Numeral measurement (money, percent, weight, etc)
- E-mail address

TYPE	DESCRIPTION
PERSON	People, including fictional.
NORP	Nationalities or religious or political groups.
FAC	Buildings, airports, highways, bridge, etc.
ORG	Companies, agencies, institutions etc.
GPE	Countries, cities, states.
LOC	Non-GPE locations, mountain ranges, bodies of water.
PRODUCT	Objects, vehicles, foods, etc. (Not services)
EVENT	Named hurricanes, battles, wars, sports events, etc.
WORK_OF_ART	Titles of books, songs, etc.
LAW	Named documents made into laws.
LANGUAGE	Any named language.
DATE	Absolute or relative dates or periods.
TIME	Times smaller than a day.
PERCENT	Percentage, including "%".
MONEY	Monetary values, including unit.
QUANTITY	Measurements, as of weight or distance.
ORDINAL	"first" , "second" , etc.
CARDINAL	Nueralns that do not fall under another type.