**Parsing in NLP**:

is the process of determining the syntactic structure of a text by analyzing its constituent words based on an underlying grammar (of the language).

**Types of Parsing** :

1. Shallow Parsing (or Chunking)

2. Constituency Parsing (or Deep Parsing)

3. Dependency Parsing

Shallow Parsing (or Chunking):

=> It adds a bit more structure to a POS tagged sentence. The most common operation is grouping words into Noun Phrases (NP).

=> You can also group stuff into VP (Verb Phrases) and PP (Prepositional Phrases).

Constituency Parsing (or Deep Parsing):

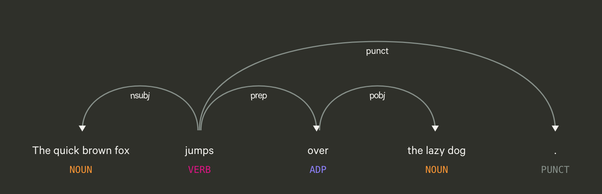
=> Adds even more structure to the POS tagged sentence. Such a parse is actually a tree, with words as leaves.

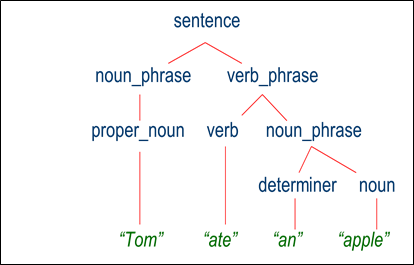
=> The other nodes are part of sentence tags: NP, VP, PP etc

Dependency Parsing:

=> Probably the most popular type of parse. It implies finding the dependencies between the words and also their type.

=> Here’s a quick example (used spaCy + displaCy for parsing and rendering)





**Deep Vs Shallow Parsing**

|  |  |
| --- | --- |
| **Deep Parsing** | **Shallow Parsing** |
| In deep parsing, the search strategy will give a complete syntactic structure to a sentence. | It is the task of parsing a limited part of the syntactic information from the given task. |
| It is suitable for complex NLP applications. | It can be used for less complex NLP applications. |
| Dialogue systems and summarization are the examples of NLP applications where deep parsing is used. | Information extraction and text mining are the examples of NLP applications where deep parsing is used. |
| It is also called full parsing. | It is also called chunking. |

## What is Chunking?

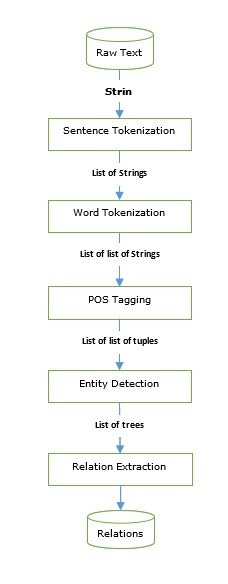
Chunking, one of the important processes in natural language processing, is used to identify parts of speech (POS) and short phrases. In other simple words, with chunking, we can get the structure of the sentence. It is also called ***partial parsing***.

### Chunk patterns and chinks

**Chunk patterns** are the patterns of part-of-speech (POS) tags that define what kind of words made up a chunk. We can define chunk patterns with the help of modified regular expressions.

Moreover, we can also define patterns for what kind of words should not be in a chunk and these unchunked words are known as **chinks**.

## Information Extraction



## Relation extraction

Relation extraction, another commonly used information extraction operation, is the process of extracting the different relationships between various entities. There can be different relationships like inheritance, synonyms, analogous, etc., whose definition depends on the information need.

## Named-entity recognition (NER)

Named-entity recognition (NER) is actually a way of extracting some of most common entities like names, organizations, location, etc. Let us see an example that took all the preprocessing steps such as sentence tokenization, POS tagging, chunking, NER

### Bag of Words (BoW) model

BoW, one of the simplest models in NLP, is used to extract the features from piece of text or document so that it can be used in modeling such that in ML algorithms. It basically constructs a word presence feature set from all the words of an instance. The concept behind this method is that it doesn’t care about how many times a word occurs or about the order of the words, it only cares weather the word is present in a list of words or not.

https://www.quora.com/What-is-parsing-in-NLP