Named Entity Recognition

Apply on Knowledg Graph and Sentiment Analysis

Phiphat Chomchit

CMU

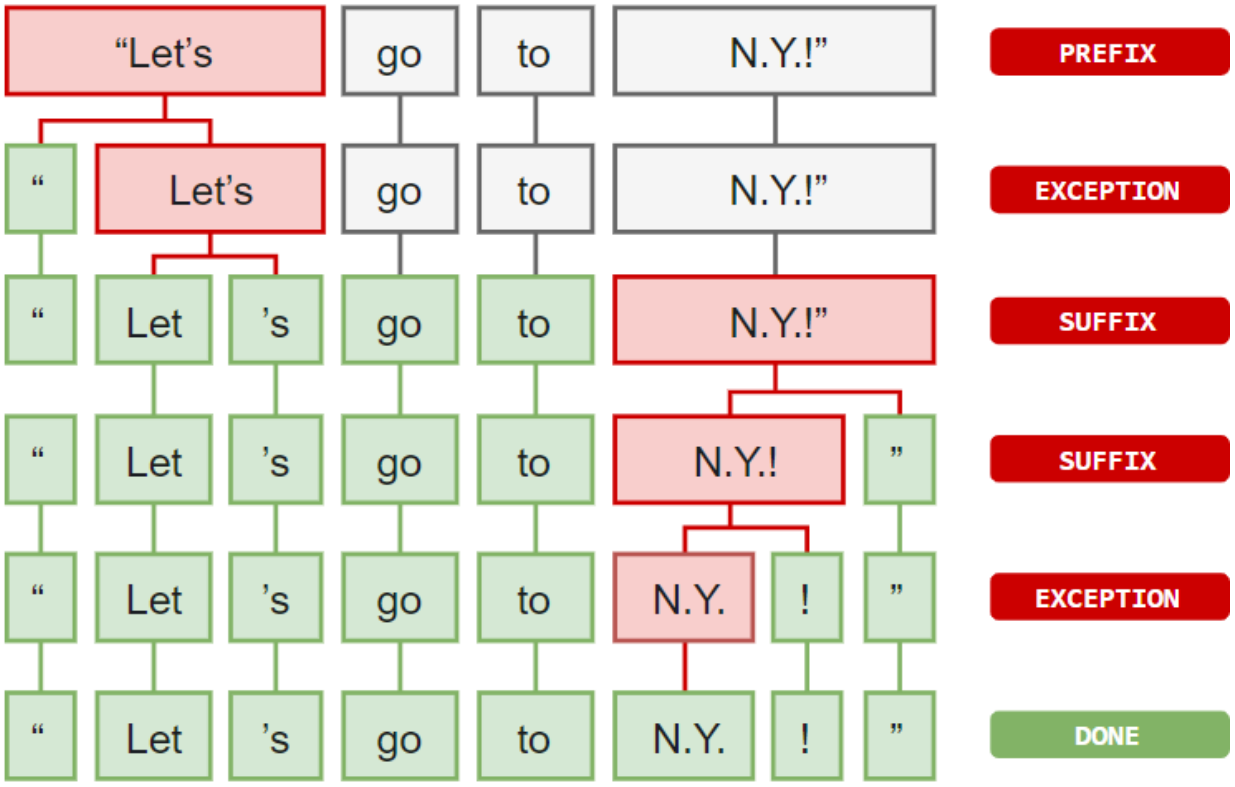
Linguistic annotations

Linguistic annotations give you insights into a text's grammatical structure.

This includes the word types, like the parts of speech, and how the words are related to each other.

Tokenization

Segmenting text into words, punctuations marks etc.



Part-Of-Speech (POS) Tagging

Part of speech or POS is a grammatical role that explains how a particular word is used in a sentence. There are eight parts of speech.

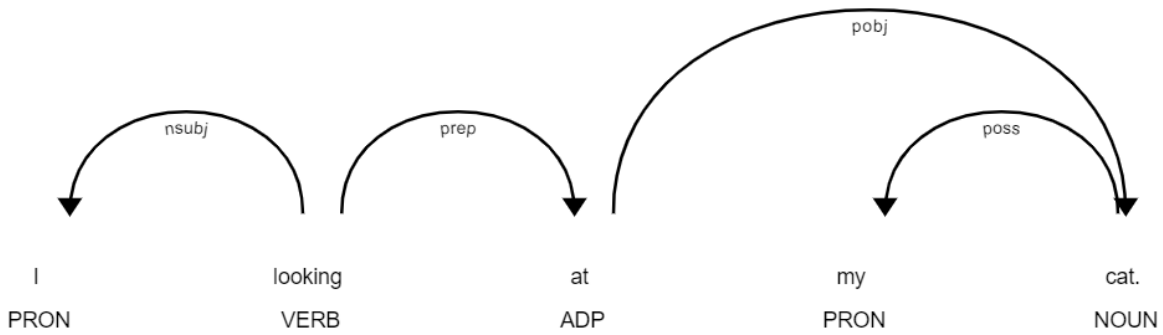
1. Noun
2. Pronoun
3. Adjective
4. Verb
5. Adverb
6. Preposition
7. Conjunction
8. Interjection

Dependency Parsing

Dependency parsing is the process of extracting the dependency parse of a sentence to represent its grammatical structure.

The dependencies can be mapped in a directed graph representation:

1. Words are the nodes.
2. The grammatical relationships are the edges.



Lemmatization

Lemmatization is the process of reducing inflected forms of a word while still ensuring that the reduced form belongs to the language. This reduced form or root word is called a lemma.
example:

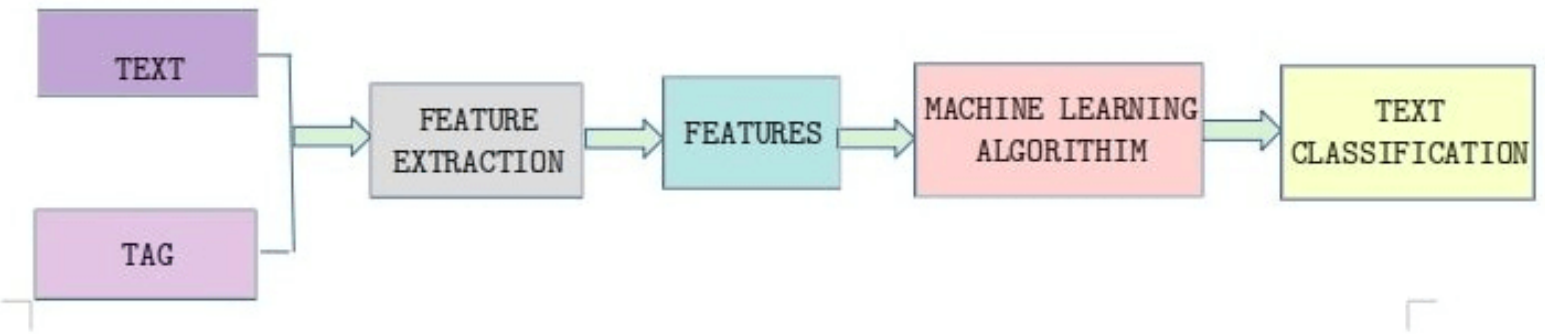
is –> be
looking –> look
analytics –> analytic

Named Entity Recognition (NER)

A named entity is a "real-world object" that's assigned a name – for example, a person, a country, a product or a book title.

Text Classification

Assigning categories or labels to a whole document, or parts of a document.



What is Knowledge Graph?

A knowledge graph is a way of storing data that resulted from an information extraction task. Many basic implementations of knowledge graphs make use of a concept we call triple, that is a set of three items(a subject, a predicate and an object) that we can use to store information about something.

Code