ABSTRACT

Natural Language Processing is a field of computer science concerned with the interactions between computers and human languages. The syntax and semantics play a very important role in Natural Language Processing. Processing natural language such as English has always been one of the central research issues of artificial intelligence. The concept of parsing is very important. In this, the sentence gets parsed into Noun Phrase and Verb phrase modules. If there are further decompositions then these modules further get divided. In this way, it helps to learn the meaning of words.

Keywords: Parse tree, Parser, syntax, semantics etc

1. Introduction
The ultimate objective of natural language processing is to allow people to communicate with computers in much the same way they communicate with each other. Natural language processing removes one of the key obstacles that keep some people from using computers. More specifically, natural language processing facilitates access to a database or a knowledge base, provides a friendly user interface, facilitates language translation and conversion, and increases user productivity by supporting English-like input.

To parse a sentence, first upon the syntax and semantics of the sentence comes into consideration. One can parse the sentence using shallow parsing, full parsing etc.

2. Related Work
2.1 Issues in Syntax
“the dog ate my homework” – Who did what?

2.2.1 Identify the part of speech (POS)
Dog = noun ; ate = verb ; homework = noun
English POS tagging: 95%

2.2.2 Identify collocations
Example: mother in law, hot dog
Weather this sentence is grammatically well formed. The meaning may be useless here but syntax allow us if it is correct syntax. Then it allows us to answer question like who did what? who ate? The dog ate what? Whose homework? All these are components of this sentence and have relationship among each other and so relationship is preserved so strictly it is syntactically correct. So when one perform syntactic analogy of sentences noun,verb,pronoun,adjective etc plays role Which identify the role of word in that particular sentence

Organized By: GNI Nagpur, India
Example: Dog – Noun; ate-Verb
If one could write a program that would accept a sentence and will able to label each component of sentence with proper part of speech then that program is known as part of speech (POS).
In English 95% work is going on. Also word may not look always in isolation (separation)
Mother-in-law one cannot just look at word mother and try mother noun, in proposition and law-noun but in order to understand mother in law it is actually mapping to another concept. So inorder to understand sentence one have to understand part of speech POS

3. ResearchMethodology

3.1 Named entity recognition (NER):
Given a stream of text, determining which items in the text map to proper names, such as people or places. Although in English, named entities are marked with capitalized words, many other languages do not use capitalization to distinguish named entities.
Natural language generation
Natural language search
Natural language understanding
Optical character recognition
Anaphora resolution
Query expansion

3.2 Speech recognition
Speech recognition is an extension of natural language processing. The idea is to use a speech recognition routine to break continuous speech into a string of words, input the string into a natural language processing routine, and then pass the resulting commands to an application program.
One problem with speech recognition is that human language is imprecise and many words have multiple meanings that depend on context. Add multiple languages, dialects, and accents, and the problem becomes very complex. Additionally, few people are skilled at issuing orders or using language with precision.
Given a sound clip of a person or people speaking, the task of producing a text dictation of the speaker(s). (The opposite of text to speech.)
Spoken dialogue system
Stemming
Text simplification
Text-to-speech
Text-proofing

3.3 Concrete problems
Some concrete problems existing in the field include part-of-speech tag disambiguation (or tagging), word sense disambiguation, parse tree disambiguation, and Anaphora Resolution. While there are typically attempts to treat such problems individually, the problems can be shown to be highly intertwined. This section attempts to illustrate the complexities involved in some of these problems.
3.4 Parsing

A natural language parser is a program that works out the grammatical structure of sentences, for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb. Probabilistic parsers use knowledge of language gained from hand-parsed sentences to try to produce the most likely analysis of new sentences.

The conversion of a flat input sentence into a hierarchical structure that corresponds to the units of meaning in the sentence.

• There are different parsing formalisms and algorithms.
• Most formalism has two main components:
  grammar - a declarative representation describing the syntactic structure of sentences in the language.
  parser - an algorithm that analyzes the input and outputs its structural representation (its parse) consistent with the grammar specification

The aim of parsing is to check whether a particular sequence is a sentence, and, if so, to determine the grammatical structure of the sentence.

A grammar is a set of rules that specifies which sequences of words constitute proper sentences in a language.

A simple grammar:

Sentence → noun phrase, verb phrase
noun phrase → determiner, noun
noun phrase → proper name
verb phrase → transitive verb, noun phrase
verb phrase → intransitive verb
determiner → every, a
noun → man
noun → woman
proper name → john
transitive verb → loves
intransitive verb → lives

Grammar rules as Horn clauses
By thinking of a phrase as a list of words, we can treat the grammar rules as Prolog clauses:

sentence(S) :-
  append (NP, VP, S),

Organized By: GNI Nagpur, India
noun_phrase(NP),
verb_phrase(VP).
noun_phrase(NP) :-
append(D, N, NP),
determiner(D),
noun(N).
noun_phrase(NP) :-
proper_name(NP).
verb_phrase(VP) :-
append(TV, NP, VP),
transitive_verb(TV),
noun_phrase(NP).
verb_phrase(VP) :-
intransitive_verb(VP).
determiner([every]).
determiner([a]).
noun([man]).
noun([woman]).
proper_name([john]).
transitive_verb([loves]).
intransitive_verb([lives]).

It is not possible to generate all sentences of the language in a straightforward way, using:
?- sentence(S).
Prolog goes into a loop after finding four sentences.
however, use these rules to check whether a particular list of words constitutes a sentence, e.g.:
?- sentence ([john, lives]).

3.5 Shallow parsing:
Shallow parsing (also chunking, "light parsing") is an analysis of a sentence which identifies the constituents noun groups, verbs, verb groups, etc., but does not specify their internal structure, nor their role in the main sentence. Shallow Parsing

Shallow Parsing is a natural language processing technique that attempts to provide some understanding of the structure of a sentence without parsing it fully (i.e. without generating a complete parse tree). Shallow parsing is also called partial parsing, and involves two important tasks:-
1. Part of Speech tagging
2. Chunking

Part of Speech tagging

Part of Speech tagging is the process of identifying the part of speech corresponding to each word in the text, based on both its definition, as well as its context (i.e. relationship with adjacent and related words in a phrase or sentence.)
E.g. if we consider the sentence 'The white dog ate the biscuits' we have the following tags
There are two main approaches to automated part of speech tagging. Let us discuss them briefly.
Rule Based Part of Speech Taggers

Rule based taggers use contextual and morphological information to assigns tags to unknown or ambiguous words. They might also include rules pertaining to such factors as capitalization and punctuation.
E.g. 1. If an ambiguous/unknown word X is preceded by a determiner and followed by a noun, tag it as an adjective (contextual rule).
2. if an ambiguous/unknown word ends in an-ous, label it as an adjective (morphological rule).
Advantages of Rule Based Taggers:-

Organized By: GNI Nagpur, India
a. Small set of simple rules.
b. Less stored information.

Drawbacks of Rule Based Taggers:-
   a. Generally less accurate as compared to stochastic taggers.

Stochastic Part of Speech Taggers
Stochastic taggers use probabilistic and statistical information to assign tags to words. These taggers might use ‘tag sequence probabilities’, ‘word frequency measurements’ or a combination of both.

Example:
   1. The tag encountered most frequently in the training set is the one assigned to an ambiguous instance of that word (word frequency measurements).
   2. The best tag for a given word is determined by the probability that it occurs with the n previous tags (tag sequence probabilities)

Advantages of Stochastic Part of Speech Taggers:-
   a. Generally more accurate as compared to rule based taggers.

Drawbacks of Stochastic Part of Speech Taggers:-
   a. Relatively complex.
   b. Require vast amounts of stored information.

Stochastic taggers are more popular as compared to rule based taggers because of their higher degree of accuracy.

However, this high degree of accuracy is achieved using some sophisticated and relatively complex procedures and data structures.

**Chunking**

Chunking is the process of dividing sentences into series of words that together constitute a grammatical unit (mostly either noun or verb, or preposition phrase). The output is different from that of a fully parsed tree because it consists of series of words that do not overlap and that do not contain each other. This makes chunking an easier Natural Language Processing task than parsing.

E.g.

The output of a chunker for the sentence
'The white dog ate the biscuits'

Would be,

[NP The white dog] [VP ate the biscuits]

On the other hand, a full parser would produce a tree of the following form:-

Thus, chunking is a middle step between identifying the part of speech of individual words in a sentence, and providing a full parsed tree of it.

**Organized By:** GNI Nagpur, India
Chunking can be useful for information retrieval, information extraction, and question answering since a complete chunk (Noun, Verb or Preposition Phrase) is likely to be semantically relevant for the requested information. In the above example, “the white dog” might be an answer or part of a question that involves the document, and it has the potential to be more relevant than each of the words in it.

4. Conclusion

The sentence can be parsed into various components such as Noun Phrase and Verb Phrase using parsing techniques. If there is further possibility to decompose the sentence then it gets divided into sub components. In this way, the parse tree enables to clear meaning of the sentence. Yet, work is going in Part-of-Speech

5. References


Author:

Organized By: GNI Nagpur, India
Manoj Vairalkar received the Master of Technology in Computer Science & Engineering degree in Computer Science and Engineering from RTMN University, Nagpur, India. Attended the 2nd International Conference On Emerging Trends in Engineering and Technology-09 at Nagpur. He has a Life Membership in ISTE.