Assignment AIR-3

Title: Syntax analysis

Problem statement
Implement syntax analysis for the assertive English statements.
The stages to be executed are
1) Sentence segmentation
2) Word Tokenization
3) Part-of-speech/morpho syntactic tagging
4) Syntactic parsing (use parses like Stanford)

Objective
Understand sentence segmentation and tokenization concepts
Implement syntax analysis on the text

Outcomes
To be able implement different text analysis and parser techniques

Theory related concepts
Sentence segmentation

1. Breaking text apart into seperate sentences word tokenization
2. Breaking the sentences into words i.e. token tokenization
3. These words are then captured for further analysis
4. Sentences are separated on space
5. Punctuation marks are treated as separated tokens since punctuations also have meaning.

Part of speech tagging

In corpus linguistics, a part of speech tagging is also called grammatical tagging. They are word category disambiguation. The process of marking up a word in a text (corpus) as corresponding to a particular part of speech based on both its definition and its context, its relation with adjacent and related words in a phrase sentence or paragraph

Sentence parsing

Parsing, syntax analysis is the process of analysis a string of symbols either in natural language, computer language or data structures conforming rules of a formal grammar

Stanford parser

A statistical parser is a program that works and the grammatical structure of sentences, for instance which groups go together and which words are the subject or object of a verb. Probabilistic parsers are ~~trained~~ knowledge of language.

Test Cases

A lion was once sleeping in the jungle, when a mouse started running up and down his body just for fun.

A - DT
lion - NN
was - VBD
once - RB
sleeping - VBA
in - IN
the - DT
jungle - NN
when - WRB
a - DT
mouse - NN
started - VBD
running - VBG
up - RB
and - CC
down - IN
his - PRP
body - NN
just - RB
for - IN
fun - NN

NN - noun sig singular
DT - Determiner
RB - Adverb
IN - Preposition
CC - Coordinating conjunction
PRN - Personal pronoun

Conclusion
Thus we have successfully understood and implemented the syntax analyser