# Natural Language Processing

File Operations Project – Part I

**Introduction**

This project will create a virtual assistant which will process the commands/sentences typed by the user and perform various file operations based on it. The commands can include different type of commands such as ‘create’, ‘delete’, ‘insert’, ‘find’, etc.

**Description**

We have assumed that every sentence can be reduced to the form of ‘action [object]+’.Where action is a command and object is one or more objects to perform the action upon

When the user inputs the command, it is broken up into tokens by the tokeniser. All the actions and the objects on which the action will take place are identified. To achieve this we have written list files for all possible file actions such as ‘create’, ‘delete’, etc and list files for objects such as files, folders and phrases.

The pipeline for running the Gate application is as follows

1. Document reset PR
2. ANNIE Sentence splitter
3. ANNIE English tokeniser
4. ANNIE Gazetteer with the def file ‘all.def’ in folder gazetteer
5. ANNIE POS Tagger
6. ANNIE NE Transducer loaded with the main.jape file

Gazetteer Lists

We have made a number of lists to recognize all possible actions and to find all references to files and folders .

Jape Rules

The main.jape runs the following jape rules in phases

1. Preprocessing1.jape – This jape rule collects all conjunctions in the corpus (code from <http://gate.ac.uk/wiki/quick-start/GATE-Tutorial-2010-Gate4.0.pdf>)
2. PreprocessingActions.jape – Rules to recognize all actions in the corpus
3. PreprocessingObjects.jape – Rules to find – Quotes , References to files , directories and phrases
4. PreprocessingObjects2.jape – Rules to find any file,directory or phrases in quotes
5. PreprocessingObjects3.jape – Rules to find actual files and directory references

Assumptions

1. All phrases and directories should be in double quotes this is a requirement for now , but we may be able to change this rigidity once we get to the java part
2. All file names must have extentions.









