**National University of Computer & Emerging Sciences**

**Karachi Campus**



**Text Classification for Context-Free Grammar**

**Project Proposal**

**DATA STRUCTURES**

**Section: H**

**Group Members:**

**19k-0305Ashmal Vayani**

**19k-0204 Hasnain Somani**

**19k-1396 Abdul Samad**

**Description:**

This project implements the CFG (Context-Free Grammar) which is a formal grammar in which every production rule is formed. CFG’s are used to describe the structure and formations of sentences, words, and phrases in natural language. They not only just support the natural language of linguists, but also forms the structure of programming language, but our project will be restricted to the natural language of linguistics. In our project, English language statements will be parsed and words will be matched from the corpus (collection of already defined words of that norm) from open source dictionaries or URLs to check if the words and sentences fall on proper grammar such as “I am out tonight” is a valid structure whereas “Boys cricket play” is invalid. Our project follows the supervised learning approach of the program as a general concept because the project or code will judge sentences on sample inputs and some pre-defined or consistent rules fed in texts to compare the entered text.

The data structures that we will be using in our projects will be mainly queues to determine the grammar of the word whether it is a noun, a preposition, a verb, and so on. Moreover, it also enqueues the type of sentence and implies respective operations on the data-set. The types of words and sentences are manipulated and reformed using functions in the program (which is why Classes have been included). A queue has been used instead of other DSA because uses the FIFO method to access data, and we need to access the sentence word to word, without affecting its actual order. Henceforth, the project uses the basics of OOP (mainly classes) and Data Structure Algorithms / Techniques (such as Parsing algorithm) to effectively work with the user's inputs, and to provide the required output. The IDE to be used will be Visual Studio, and the programming language will be C++.