

# **CSE508 : Information Retrieval**

## **Assignment 1**

### **Instructions**

- The assignment is to be attempted individually
- Language allowed: Python
- For Plagiarism, institute policy will be followed
- You need to submit ReadMe, code files and analysis.pdf

### **Question**

1. Use the [20newsgroups dataset](#) to build a unigram inverted index.
2. Further, provide support for the following commands:
  - x OR y
  - x AND y
  - x AND NOT y
  - x OR NOT yWhere x and y would be taken as input from the user.
3. Provide support for searching for phrase queries using Positional Indexes. (For this question, build index only on comp.graphics and rec.motorcycles)