PSG COLLEGE OF TECHNOLOGY

**Department of Applied Mathematics and Computational Sciences**

**IX Semester M.Sc. SS – 2023-2024**

**Information Retrieval Lab**

**Exercise 1**

**Reading Assignment**

<http://lucene.apache.org/core/>

<http://tartarus.org/~martin/PorterStemmer/>

<http://www.crummy.com/software/BeautifulSoup/>

<http://nltk.org/>

**General resources**

[**http://www.acm.org/sigir/**](http://www.acm.org/sigir/)

[**http://trec.nist.gov/**](http://trec.nist.gov/)

**Data sets**

<http://ir.dcs.gla.ac.uk/resources/test_collections/>

<https://ai2-semanticscholar-cord-19.s3-us-west-2.amazonaws.com/historical_releases.html>

**Programming Assignment**

The aim of this programming assignment is to create a simple index and design a retrieval system.

* Create an Inverted Index
* Implement a program that reads through the list of sorted terms and creates an in-memory inverted index.
* Treat the 10 most prominent words as stop words and let your program compute the index size after that change.
* To process simple Boolean queries using the index file create above.
* Provide support for conjunctive queries and mixed operator queries.