TAUGHT BY:

PROF. NADA NAJI

INFORMATION RETRIEVAL

SEMESTER I, FALL 2016

December 9, 2016

ROHIT RAJ SAMANJATE SOOD Sumit BHANWALA

INTRODUCTION

The goal of this project is to design and build an Information retrieval system. The performance of this system is then evaluated and compared in terms of retrieval effectiveness.

The first task of the project is to build four different retrieval modules namely Lucene, BM25, tf-idf, and cosine similarity. These programs are then run on the CACM test-collection. We then use BM25 as the base search engine and perform query expansion using Pseudo relevance feedback approach. For the next task we use BM25 as our search engine and perform Stopping based on the common words list given. These words are neither considered for document scoring nor as a query term. In the same task we index the stemmed corpus and run the stemmed queries on this index.