**INFORMATION RETRIEVAL \_ ASSIGNMENT 2**

**Content**: Vector Space Model

1. PREPARING:

Download from Course Site:

- Dataset

- File of stopwords

2. INDEXING

2.1. Remove stopwords

2.2. Build vocabulary: Choose randomly 500 words from the remaining words from the collection of documents.

2.2. Perform inverted index.

3. QUERY

3.1. Implement a function for computing tf\_idf vector for a document/query.

3.2. Implement functions for computing relevance between a query and a document in two fashions: Using Distances and Using Angles between tf\_idf vectors

3.3. Implement a function for searching for relevance documents given a query.