# Web Information Retrieval (67782) Ex1: Index Structure Analysis

**Submitted by:**

## 1 General Explanation and Diagram

This section should contain the precise details behind the index structure that you have implemented. Your discussion should be very specific and should allow the reader to precisely understand the format of your index files, stored on disk. Provide a diagram that depicts the structure of the index.

## 2 Main Memory Versus Disk

Put an explanation of which portions of the index are read into memory when an IndexReader object is created, and which portions will be read as needed.

## 3 Theoretical Analysis of Size

Theoretically analyze the expected size (in bytes) of all of your index structures. In your analysis, the size of the index should be a function of the size of the input. Use the following variables to denote the various input size parameters:

N Number of reviews

M Total number of tokens (counting duplicates as many times as they appear)

D Number of different tokens (counting duplicates once)

L Average token length (counting each token once)

F Average token frequency, i.e., number of reviews containing a token You can Add additional variables as needed.

## 4 Theoretical Analysis of Runtime

Using the same variables, theoretically analyze the runtime of the functions of IndexReader. Include both the runtime and a short explanation.

* IndexReader(String dir):
* getProductId(int reviewId):
* getReviewScore(int reviewId):
* getReviewHelpfulnessNumerator(int reviewId):
* getReviewHelpfulnessDenominator(int reviewId):
* getReviewLength(int reviewId):
* getTokenFrequency(String token):
* getTokenCollectionFrequency(String token):
* getReviewsWithToken(String token):
* getNumberOfReviews():
* getTokenSizeOfReviews():
* getProductReviews(String productId):