|  |  |
| --- | --- |
|  | **Rochester Institute of Technology**  **Golisano College of Computing and Information Sciences**  **Department of Information Sciences and Technology** |

**Lab 2**

**Positional index construction and phrase query processing**

#### Overview

This lab consists of two major tasks:

* Building a positional index, and
* Use the positional index to process phase queries with any number of query terms

#### Resources

* You should have read Chapter 2 of Introduction to Information Retrieval.
* Carefully read the lecture examples of weeks 3 and 4 to understand the technical details.
* Go over the lecture notes of weeks 3 and 4.

**Tasks to be completed**

* Task1: Complete the constructor of the class that builds the positional index.
* Task2: Implement the intersect method that takes in two postings and output a merged postings.
* Task3: Implement the phraseQuery method that takes in a phrase query with multiple terms and return a list of DocId objects (the DocId class is given in the Appendix).
* Task4: Design and test phrase queries with 2-5 terms.

**Appendix**

**class** DocId{

**int** docId;

ArrayList<Integer> positionList;

**public** DocId(**int** did)

{

docId = did;

positionList = **new** ArrayList<Integer>();

}

**public** DocId(**int** did, **int** position)

{

docId = did;

positionList = **new** ArrayList<Integer>();

positionList.add(**new** Integer(position));

}

**public** **void** insertPosition(**int** position)

{

positionList.add(**new** Integer(position));

}

**public** String toString()

{

String docIdString = ""+docId + ":<";

**for**(Integer pos:positionList)

docIdString += pos + ",";

docIdString = docIdString.substring(0,docIdString.length()-1) + ">";

**return** docIdString;

}

}