R·I·T	**Rochester Institute of Technology** **Golisano College of Computing and Information Sciences** **School of Information**

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

- Task 1 (10 points): Complete the constructor of the class that builds the positional index.
- Task 2 (30 points): Complete implementing the intersect method that takes in two postings and output a merged postings list.
- Task 3 (20 points): Implement a phraseQuery method that takes in a phrase query with multiple terms and return a list of DocId objects (use the DocId class from the Positional Index coding lecture).
- Task 4 (40 points): Design and test phrase queries with 2, 3, 4, and 5 terms. Use the document collection from Lab 1 to design and test the queries.

Lab Submission Instructions – Create a PDF document showing all of your test cases for Task 4 and the resultant output. Your output should also show the postings for each term involved in a given query so it's obvious that your output is correct. Create a zip file called Lab2.zip that includes the PDF document with your results, the Lab1_Data folder with all documents contained therein, and your code and submit it to the Lab 2 dropbox prior to the due date. **Please do NOT use file paths that are specific to your machine. Your code should include a hard-coded path to the Lab1_Data directory, which will be in the current directory where the code resides. Failure to do this will result in significant point loss.**