

CSE 508, Information Retrieval

Assignment 1

Due Date: 29 January 2016

- Implement EDIT Distance Computation using Dynamic Programming.
- Consider the following document D, taken from a collection C.

“World’s fifth seed Rafael Nadal crashes out of the Australian Open after losing in the first round against compatriot Fernando Verdasco.”

“Rafael Nadal loses to Fernando Verdasco and crashes out of the Australian Open”

1. Develop the inverted index for the documents
 2. Assuming that stemming is used, compute the answers to the query: “Loses”
 3. Show the bi-gram for both the documents and prepare a consolidated bigram-inverted index to answer query “ Rafeal” and “Australia”
-
- Download 5 books from archive.org and write an indexer to build an inverted index. Remove stop words.
 - Develop query support to answer exact queries.
 - In each assignment we will add more components to this indexer and query engine to refine it iteratively.
 - Use language of your choice. Preferred in Java and Python primarily due to availability of lot of open source kits which can help in next assignments.