

Information Retrieval and Text Mining Fall 2016

Quiz 1 (Total Marks = 10)

Roll No: _____ **Name** _____

Q1) What will be stemmed version of following words using Porter stemmer (2 marks)

- a) talking
- b) ping
- c) retrieval
- d) comparable

Q2) With the query “information retrieval” and the document “search retrieval information retrieval”, what is the tf-idf score for the query and document? (2 marks)

Note: there are 10000 total documents, “information” occurs in 1000 of them, and “retrieval” occurs in 100 of them

Q3) What are advantages and disadvantages of having more skips in skip lists for postings? (2 marks)

Q4) What is advantage of using Single Pass In Memory Index (SPIMI) as compared to Blocked Sort Based Index (BSBI)? (2 marks)

Q5) If a collection is so large that all the unique terms (dictionary) cannot fit into memory of your machine then can you use term-ids instead of terms in your index ? Why ? (2 marks)