

## Homework 2

**Student Number:**

**Name:**

**Problem 1.** (10 points) Why are skip pointers not useful for queries of the form  $x$  OR  $y$ ?

**Problem 2.** (20 points) In the permuterm index, each permuterm vocabulary term points to the original vocabulary term(s) from which it was derived. How many original vocabulary terms can there be in the postings list of a permuterm vocabulary term?

**Problem 3.** (20 points) Shown below is a portion of a positional index in the format: term: doc1: < position1, position2, . . . >; doc2: < position1, position2, . . . >; etc.

angels: 2: < 36,174,252,651 > ; 4: < 12,22,102,432> ; 7: <17>;

fools: 2: <1,17,74,222> ; 4: <8,78,108,458 >; 7: <3,13,23,193> ;

fear: 2: <87,704,722,901>; 4:<13,43,113,433>; 7: <18,328,528>;

in: 2: < 3,37,76,444,851> ; 4: <10,20,110,470,500>; 7: < 5,15,25,195> ;

rush: 2: < 2,66,194,321,702> ; 4: < 9,69,149,429,569>; 7: < 4,14,404 >;

to: 2: < 47,86,234,999>; 4:<14,24,774,944> ; 7: < 199,319,599,709>;

tread: 2:< 57,94,333>; 4: <15,35,155> ; 7: < 20,320>;

where: 2: < 67,124,393,1001 > ; 4:< 11,41,101,421,431>; 7: < 16,36,736> ;

Which document(s) if any match each of the following queries, where each expression within quotes is a phrase query?

a. “fools rush in”

b. “fools rush in” AND “angels fear to tread”

**Problem 4.** (20 points)

1. Find two differently spelled proper nouns whose soundex codes are the same.

2. Find two phonetically similar proper nouns whose soundex codes are different.

**Problem 5.** (30 points) Write the pseudo code showing the details of computing the Jaccard coefficient while scanning the posting of the k-gram index.