

DEPARTMENT OF COMPUTER SCIENCE - THIRD YEAR LICENCE (ISIL)

TD 03

Exercise 01

Write a merge algorithm in the style of x OR y.

Exercise 02

How should the Boolean query $x \text{ AND NOT } y$ be handled? Describe a naïve algorithm for this and then write an efficient solution

Exercise 03

Consider the following document collection:

Doc1: "The quick brown fox jumped over the lazy dog."

Doc2: "The brown dog chased the fox, but the fox was too quick."

Construct a positional index for this collection, where each term is represented with its corresponding document IDs and the position of the term in the document.

Exercise 04

Consider the following positional index:

Term	Document IDs and Positions
apple	Doc1: 1, 3, 5, Doc2: 2, 4
banana	Doc1: 2, 4, Doc2: 1, 3, 5
cherry	Doc1: 3, Doc2: 2, 4

1. Find all documents that contain the phrase "apple banana".
2. Find all documents that contain the phrase "banana cherry" with a maximum distance of 2 between the terms.

Exercise 05

Following is portion of positional index:

angels: 2: $\langle 36,174,252,651 \rangle$; 4: $\langle 12,22,102,432 \rangle$; 7: $\langle 17 \rangle$;
fools: 2: $\langle 1,17,74,222 \rangle$; 4: $\langle 8,78,108,458 \rangle$; 7: $\langle 3,13,23,193 \rangle$;
fear: 2: $\langle 87,704,722,901 \rangle$; 4: $\langle 13,43,113,433 \rangle$; 7: $\langle 18,328,528 \rangle$;
in: 2: $\langle 3,37,76,444,851 \rangle$; 4: $\langle 10,20,110,470,500 \rangle$; 7: $\langle 5,15,25,195 \rangle$;
rush: 2: $\langle 2,66,194,321,702 \rangle$; 4: $\langle 9,69,149,429,569 \rangle$; 7: $\langle 4,14,404 \rangle$;
to: 2: $\langle 47,86,234,999 \rangle$; 4: $\langle 14,24,774,944 \rangle$; 7: $\langle 199,319,599,709 \rangle$;
tread: 2: $\langle 57,94,333 \rangle$; 4: $\langle 15,35,155 \rangle$; 7: $\langle 20,320 \rangle$;
where: 2: $\langle 67,124,393,1001 \rangle$; 4: $\langle 11,41,101,421,431 \rangle$; 7: $\langle 16,36,736 \rangle$;

Which documents return as result if we have those phrase queries?

- a. "fools rush in"
- b. "fools rush in" AND "angels fear to tread"