Information Retrieval 21-22

Exercises week 1

Deadline: 13 February 24:00

Exercise 1

Rewrite these Boolean queries in disjunctive normal form:

a) NOT (NOT Caesar OR roman) AND Brutus
b) (Euro OR NOT Dollar) AND NOT Yen

Exercise 2

Consider these documents:

Doc 1 new recipe for cookies

Doc 2 cookie dough recipe

Doc 3 new chocolate dough

Doc 4 recipe for chocolate cookies

a) Draw the term-document incidence matrix for this document collection

b) Draw the inverted index representation for this collection.

Exercise 3

For the document collection shown in Exercise 2, what are the returned results for these queries:

a) cookies OR cookie

b) recipe AND NOT (chocolate OR dough)

For the queries below, can we still run through the intersection time O*(x+y),* where *x* and *y* are the lengths of the posting lists for Brutus and Caesar? If not, what can we achieve?

    a) Brutus AND NOT Caesar

    b) Brutus OR NOT Caesar

Exercise 5

Recommend a query processing order for the query below, given the postings list sizes in the table:

(rocket OR launch) AND (sky OR universe) AND (sun OR orbit)

Term	Postings size
rocket	81300
launch	109999
sky	43123
universe	48491
sun	62687
orbit	577513