CMPE 493, Introduction to Information Retrieval
Assignment 1 - A Simple Document Retrieval System for Boolean Queries

Utku SARIDEDE
2010400189

I have used well defined comments in source code. The one who wants to execute my source code should install some packages mentioned in readme file. Answers of the questions:

i.) a.) 2.745.153
b.) 1.850.520
c.) 92.267
d.) 64.799

e.) (u'will', 14654)
(u'its', 14724)
(u'from', 14939)
(u'that', 15189)
(u'is', 16755)
(u'pct', 17046)
(u'on', 17797)
(u'it', 18082)
(u"'s", 18438)
(u'dlrs', 20268)
(u'The', 24271)
(u'for', 25194)
(u'mln', 25524)
(u'a', 48369)
(u'in', 49949)
(u'said', 52887)
(u'and', 53410)
(u'to', 68539)
(u'of', 72194)
(u'the', 119923)

f.) (u'price', 7362)
(u'trade', 7434)
(u'net', 7703)
(u'inc', 7737)
(u'u.s.', 7979)
(u'market', 8369)
(u'would', 9233)
(u'ct', 9413)
(u'share', 10092)
(u'billion', 10696)
(u'compani', 11335)
(u'bank', 12149)
(u'year', 13365)
(u'vs', 14827)
(u'pct', 17962)
(u"'s", 19283)
(u'reuter', 20013)
(u'dlr', 24442)
(u'mln', 26675)
(u'said', 53089)

ii.) I have used hash-map approach to create inverted index. Hash-map keeps unique words and some unique dates(1985, 11.18.2002 etc.) as “key”. Values of these keys are lists (known as array in some languages) which include the indexes of keys. All other stuffs are written in the source code as comment outs.

iii.) Conjunctive and Disjunctive search query results are as the followings respectively;