Assignment 1

CS 5154 – Information Retrieval

# IR1A.py

## Code:

# IR1A.py CS5154/6054 cheng 2022  
# read lines from a text file  
# retrieve documents for boolean queries  
# Usage: python IR1A.py  
  
import time  
  
f = open("bible.txt", "r")  
t1 = time.process\_time\_ns()  
docs = f.readlines()  
t2 = time.process\_time\_ns()  
f.close()  
  
print(f"Time taken to read the file: {t2 - t1} ns")  
  
word1 = 'punishment'  
word2 = 'transgressions'  
  
t3 = time.process\_time\_ns()  
for i in range(len(docs)):  
 if word1 in docs[i] and word2 in docs[i]:  
 print(i, docs[i])  
t4 = time.process\_time\_ns()  
  
print(f"Time taken for boolean retrieval (without inverted index): {t4 - t3} ns")

## Results:

Time taken to read the file: 46875000 ns

Time taken for boolean retrieval (without inverted index): 15625000 ns

## Screenshot:

A picture containing text

Description automatically generated

# IR1B.py

## Code:

# IR1B.py CS5154/6054 cheng 2022  
# read lines from a text file  
# turn each line into a list of tokens  
# make the inverted index  
# retrieve documents using boolean queries  
# Usage: python IR1B.py  
  
import time  
  
t1 = time.process\_time\_ns()  
f = open("bible.txt", "r")  
docs = f.readlines()  
f.close()  
t2 = time.process\_time\_ns()  
print(f"Time taken to read the file: {t2 - t1} ns")  
  
t3 = time.process\_time\_ns()  
invertedIndex = {}  
for i in range(len(docs)):  
 for s in docs[i].split():  
 if invertedIndex.get(s) == None:  
 invertedIndex.update({s : {i}})  
 else:  
 invertedIndex.get(s).add(i)  
t4 = time.process\_time\_ns()  
print(f"Time taken for creating inverted index: {t4 - t3} ns")  
  
word1 = 'punishment'  
word2 = 'transgressions'  
t5 = time.process\_time\_ns()  
for i in range(0, 100):  
 for j in invertedIndex.get(word1) & invertedIndex.get(word2):  
 print(j, docs[j])  
t6 = time.process\_time\_ns()  
  
print(f"Time taken for boolean retrieval (with inverted index) for 100 iterations: {(t6 - t5)} ns")

## Results:

Time taken to read the file: 15625000 ns

Time taken for creating inverted index: 453125000 ns

Time taken for boolean retrieval (with inverted index) for 100 iterations: 78125000 ns

## Screenshot:

A picture containing calendar

Description automatically generated

… outputs lines 21665, 21647, 21650, 21653, 21655, 21660, 21663 hundred times each in total (including screenshots) …

Text

Description automatically generated with medium confidence