

Assignment 3 Milestone 1

Number of documents: 44536

Number of unique words: 155534

Program runtime: 0 hours, 0 minutes, 18 seconds

Index.json file size = 39,242 KB termsID.json size = 2,544 KB

Sample Index

doc1 = "Dogs are mean. They chase after birds."

doc2 = "Dogs are friendly, and cats are lazy."

doc3 = "Mice and birds are friends. They don't like dogs or cats."

##Map each doc to a unique docID

docIDs = { 1: doc1, 2: doc2, 3: doc3 }

##Map each term to a unique termID (ignore stop words)

term2termid = { "dogs": 1, "mean": 2, "chase": 3, "birds": 4, "friendly": 5,
 "cats": 6, "lazy": 7, "mice": 8, "friends": 9, "like": 10 }

##Map each docID to a map of term-frequency pairs

docFreqIndex = { 1: ["dogs": 1, "mean": 1, "chase": 1, "birds": 1],
 2: ["dogs": 1, "friendly": 1, "cats": 1, "lazy": 1],
 3: ["mice": 1, "birds": 1, "friends": 1, "like": 1, "dogs": 1, "cats": 1] }

##Map each termID to a list of docIDs-termFrequency pairs

reverseIndex = { 1: [1: 1, 2: 1, 3: 1],
 2: [1: 1],
 3: [1: 1],
 4: [1: 1, 3: 1],
 5: [2: 1],
 6: [2: 1, 3: 1],
 7: [2: 1],
 8: [3: 1],
 9: [3: 1],
 10: [3: 1] }