ASSIGNMENT 2

DOCUMENTATION

def nextPermutation():

Generates a random sequence of integers between 0 to numDocs-1.

def Minhash(data): returns a 2d matrix

Computes the signature matrix in sigmatrix and returns it.

def JaccardSimilarityOrig(docID1, docID2): returns an integer

Computes Jaccard similarity between two documents using the original membership matrix

def JaccardSimilaritySig(docID1, docID2): returns an integer

Computes Jaccard similariy between two documents using the signature matrix

def findCandidatesLSH(docID): returns a list of integers

returns a list of candidate documents using LSH which need to be tested for similarity

def findCandidatesCosine(docID): returns a list of integers

returns a list of candidate documents using Cosine similarity which need to be tested for similarity

def outputResults():

tests each candidate in the list of candidates for similarity and prints m most similar documents