

Programming Assignment 4 - Little Search Engine

Test Cases

getKeyWord

1. Word - 2
2. night, - 2
3. question?? - 2
4. Could - 2
5. test-case - 2

loadKeywords

1. Lots of noise - [pohlx.txt](#) - 5
2. Lots of keywords - [Tyger.txt](#) - 5
3. Repetition of keywords - [jude.txt](#) - 5

insertLastOccurrence

1. Extreme left insertion
Array: [(doc1.txt,20), (doc2.txt,15), (doc3.txt, 14), (doc4.txt, 12), (doc5.txt, 12),(doc6.txt, 10), (doc7.txt, 8), (doc8.txt, 21)]
Ans: [3, 1, 0] - 4
2. Extreme right insertion
Array :[(doc1.txt,20), (doc2.txt,15), (doc3.txt, 14), (doc4.txt, 12), (doc5.txt, 12),(doc6.txt, 10), (doc7.txt, 8), (doc8.txt, 1)]
Ans: [3, 5, 6] - 3
3. In between
Array: [(doc1.txt,20), (doc2.txt,15), (doc3.txt, 14), (doc4.txt, 12), (doc5.txt, 12),(doc6.txt, 10), (doc7.txt, 8), (doc8.txt, 13)]
Ans: [3, 1, 2] - 4
4. Insertion when there is a Match
Array: [(doc1.txt,20), (doc2.txt,15), (doc3.txt, 14), (doc4.txt, 12), (doc5.txt, 12),(doc6.txt, 10), (doc7.txt, 8), (doc8.txt, 10)]
Ans: [3, 5] - 4

mergeKeywords

1. Single document - [metamorphosis.txt](#) - 4
2. 2 documents with no intersection - [Tyger.txt](#) and [metamorphosis.txt](#) - 5
3. 2 documents with a few intersections - [pohlx.txt](#) and [pohly.txt](#) - 6

top5Search

docs to use: [doc1.txt](#), [doc2.txt](#), [doc3.txt](#), [doc4.txt](#), [doc5.txt](#), [doc6.txt](#), [doc7.txt](#) and [doc8.txt](#).

1. No matches for both
Keywords: yellow, blue

Ans: no list. - 2

2. No match for 1 keyword, but more than 5 total matches

Keywords: red, blue (No match)

Ans: doc1,2,3,4,5 - 3

3. Matches on both keywords, no common docs, more than 5 total matches

Keywords: green (doc1,2,3,4), bike (doc5,6,7)

Ans: doc1,4,5,7,2 - 5

4. 2 common docs, no common frequencies, more than 5 total matches

Keywords: apple (doc1,2,3,4), orange (doc2,3,5,6,7)

Ans: doc5,3,2,1,7 - 5

5. 2 docs in different list with the same frequency, more than 5 results

Keywords: Bus (doc2,3,4,5) Car (doc1,3,4,6,7)

Ans: (doc1,3,4,2,5) (doc7 freq matches doc5) - 5