Programming Assignment 4 - Little Search Engine

Test Cases

---

getKeyword (pts: 3,3,3,2,2)

1. sWord
   Ans: sword
2. paraphrase;
   Ans: paraphrase
3. really?!?!
   Ans: really
4. Between,
   Ans: null
5. either:or
   Ans: null

loadKeywordsFromDocument (pts: 6,7,7)

1. Lots of noise - pohlx.txt
2. Lots of keywords - Tyger.txt
3. Repetition of keywords - jude.txt

insertLastOccurrence (pts: 2,3,2,3,2)

1. Extreme left insertion
   Frequencies: [82,76,71,71,70,65,61,56,54,51,48,45,41,36,34,30,25,20,20,18,17,17,14,12,85]
   Ans: [11,5,2,0]
2. Extreme right insertion
   Frequencies: [82,76,71,71,70,65,61,56,54,51,48,45,41,36,34,30,25,20,20,18,17,17,14,12,4]
   Ans: [11,17,20,22,23]
3. In between test 1
   Frequencies: [82,76,71,71,70,65,61,56,54,51,48,45,41,36,34,30,25,20,20,18,17,17,14,12,50]
   Ans: [11,5,8,9,10]
4. In between test 2
   Frequencies: [82,76,71,71,70,65,61,56,54,51,48,45,41,36,34,30,25,20,20,18,17,17,14,12,26]
   Ans: [11,17,14,15,16]
5. Insertion with a frequency match
   Frequencies: [82,76,71,71,70,65,61,56,54,51,48,45,41,36,34,30,25,20,20,18,17,17,14,12,17]
   Ans: [11,17,20]

mergeKeywords (pts: 7,9,9)

1. Single document - metamorphosis.txt
2. 2 documents with no intersection - Tyger.txt and metamorphosis.txt
3. 2 documents with a few intersections - pohlx.txt and pohly.txt

top5Search (pts: 3,4,7,8,8)

docs to use: doc1.txt, doc2.txt, doc3.txt, doc4.txt, doc5.txt, doc6.txt, doc7.txt and doc8.txt.

1. No matches for either
   Keywords: strange (no match), case (no match)
   Ans: no list
2. No match for 1 keyword, but more than 5 total matches
   Keywords: color (doc5,3,7,2,4,1), strange (No match)
   Ans: doc5,3,7,2,4 (1 pt penalty if > 5 results)
3. Matches on both keywords, no common docs, more than 5 total matches
   Keywords: orange (doc5,2,7,3,1), weird (doc6,8)
   Ans:doc5,6,2,8,7 (1 pt penalty if > 5 results)
4. 4 common docs, no common result frequencies, more than 5 total matches
   Keywords: red (doc3,2,4,1,7,6), orange (doc5,2,7,3,1)
   Ans: doc5,3,2,4,7 (2 pt penalty if > 5 results)
5. 2 docs in different lists with the same frequency, more than 5 results
   Keywords: red (doc3,2,4,1,7,6), car (doc1,3,7,6,4)
   Ans: (doc1,3,2,7,4) (doc4 freq matches doc6) (2 pt penalty if > 5 results)