

TextRank Algorithm for Document Summarization

1. Open Text file with doc as f.
2. Divide Text in f into list of composing sentences using nltk tokenizer.
3. Pass sentence list to graph builder function.
4. Use each sentence in the list passed in step 3 as a node.
5. Calculate all node pairs and store in list nodePairs.
6. For pair in nodePairs
 - string1 = pair[0]
 - string2 = pair[1]
 - dist = Levenshtien distance between string1 and string2
 - Add edge to graph between pair[0] and pair[1] with weight = dist
7. Return graph
8. Use pagerank algorithm on graph to calculate the most significant nodes and store this in list sentenceRank.
9. Sort sentenceRank in descending order.
10. sentenceLimit = set the number of sentences in summary.
11. Join elements of list sentenceRank upto index sentenceLimit-1.
12. Write summary to new file → outputSummary.txt