CSCI 572 ASSIGNMENT 3 RUBRICS

Inverted Index

Total Marks: **10 marks**

Submission Requirements	Marks Allotted
unigram_index.txt	4 mark/s
selected_bigram_index.txt	4 mark/s
Two screenshots of the output folder	1 mark/s
Two source files (unigrams and bigrams)	1 mark/s

Rubrics – Inverted Index

- **Unigram Index (4 points)**
    - If this file is missing: -4 marks
    - If the format is incorrect: -2 mark
      -Just check if the (word, DocID) are present.
    - If results are incorrect: (-0.25 for each mismatch, cap at -2 marks)
        - Students may have chosen a smaller (dev) OR larger dataset to produce the results. Please randomly select a few (2-3) words from the student's submitted text file and compare them with the following results. Outputs for those (2-3) words must match entirely within one of the below files.
            - Smaller dataset results
            - Larger dataset results
        - Please note that for each word, the order of the docID in the student's text file might be different than the expected results. Please check thoroughly before docking points.
- **Bigram Index (4 points)**
    - If this file is missing: -4 marks
    - If the format is incorrect (Just check if (word1, word2, DocID) are present): -2 marks
    - If results are incorrect: (-0.25 for each mismatch, cap at -2 marks)
        - Please copy the words one by one from the "Bigram results" section mentioned on page 2 and search for them in the student's text file. Verify that the results for each word are correct. (Again, the student may have chosen either a smaller or larger dataset, and their result must match one of the expected results)
        - Please note that for each word, the order of the docID in the student's text file might be different than the expected results. Please check thoroughly before docking points.
- **Screenshots of the output folder (1 point)**
    - If this unigramJaya screenshot is missing: -0.5 mark

- If this bigram screenshot is missing -0.5 mark
- **Code Submission (1 point)**
  - If this unigram source code java file is missing: -0.5 mark
  - If this bigram source code java file is missing: -0.5 mark

Bigram results

Bigger data set

bruce willis	5722018301:4	5722018235:8			
computer science	5722018301:11	5722018508:597	5722018496:16	5722018235:62	5722018101:104
information retrieval	5722018496:1	5722018235:3	5722018508:8	5722018101:11	
los angeles	5722018101:200	5722018301:85	5722018496:318	5722018235:300	5722018508:174
power politics	5722018496:1	5722018301:6	5722018101:3	5722018235:3	5722018508:3

Shortened data set

bruce willis	5722018301:3	5722018235:4			
computer science	5722018496:1	5722018508:1	5722018235:1	5722018301:1	
information retrieval	5722018496:1	5722018235:1	5722018508:2		
los angeles	5722018496:2	5722018101:3	5722018301:2	5722018235:6	5722018508:13
power politics	5722018496:1	5722018508:1	5722018301:3	5722018235:3	