

Problem statement:

Build the inverted index for a given set of documents (compute the term weights by TF-IDF as shown in slide 5 of Chapter 4, using base 10 logarithm). Ignore the letter case, i.e., consider all words as lower case.

Input files:

Each line is in format of “DocID DOC”, where DocID is the ID of the document, and DOC is the document content (a list of terms). For example:

0 Construct Inverted Index
1 using MapReduce
2 inverted index code
3 index index file

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro