

11791 Report for Design and Engineering of Intelligent Information System and Software Methods for Biotechnology Homework 4

Engineering and Error Analysis with UIMA

Name: Trisha Black

ID: trishab

Baseline Retrieval and Analysis

For this assignment a prototype architecture provide a document retrieval and analysis system. The three components of this system are: 1) The *DocumentReader* composed the relevance scores and the query ID. 2) The *DocumentVectorAnnotator* identified the word and word frequency. 3) The *RetrievalEvaluator* computed the similarities and the performance metric of the output.

Error Analysis and Output

The following tasks were completed for the output below: 1) Extract bag of words feature vector from the input text collection. 2) Compute the cosine similarity between two sentences in the collection. 3) Compute the Mean Reciprocal Rank for all sentences in the text collection.

Score	Rank	Relevance Score (rel) Query ID (qid)	Sentence
Score:0.452267	rank=1	rel=1 qid=1	Classical music may never be the most popular music
Score:0.102062	rank=1	rel=1 qid=2	Climate change and energy use are two sides of the same coin.
Score:0.507093	rank=1	rel=1 qid=3	The best mirror is an old friend
Performance (MRR) Mean Reciprocal Rank ::1.0			
Total time taken: 2.762			