

ASSIGNMENT 3

Due date: 3 Nov 2021 11.59PM (Hard deadline)

Marks: 10

For the corpus of documents you considered for your Assignment 2, apply the Best-Match Model variants and submit your solution for the following:

- Find $\text{sim}(d_j, q)$ using the original BM1, BM11 and BM15 models. What is the effect of the introduction of term frequency and document length normalization on the ranking? Demonstrate with relevant experiments.
- Find $\text{sim}(d_j, q)$ using the simplified BM1, BM11 and BM15 models. What is the effect of choosing the standard suggested values on the ranking? Report your observations in detail, by choosing any value other than the commonly used standard values.
- Apply the BM25 model to the corpus and observe the change in ranking, when values of K_1 and b are varied (Show the results for at least two such cases). Analyse your findings.

Note:

- Submit a detailed report on your observations and analysis, supported by the necessary code snippets w.r.t your program and results.
- Upload your report and code (well documented) on Moodle to the folder provided before the deadline of **11.59PM on 3rd November 2021 (No extension will be given).**

EVALUATION RUBRICS:

Part a: Code, demonstrative experiments +analysis - 15 marks

Part b : Code, demonstrative experiments +analysis - 15 marks

Part c : Code, demonstrative experiments +analysis - 10 marks

Detailed report on assignment - 10 marks