

ASSIGNMENT 3

Due date: Wednesday, 26th October 2022, 11.59PM (Hard deadline)

Note: Assignment evaluation will be on Thursday, 27th October 2022 during regular lab hours.

Consider the corpus of 100 articles from the Incredible India website (articles that you used for the Assignment 2). You may reuse the preprocessing performed on the corpus earlier, for this assignment. Consider two standard queries of lengths 3 words and 5 words, for your experimentation. For the document corpus represented in VSM using the standard TF-IDF weights, generate the ranked list for each of the sample queries using Cosine similarity (as done in Assignment 2). Now, implement Rocchio's algorithm to incorporate relevance feedback in the same. Recompute the ranked list using Cosine similarity for each of the reformulated queries to assess the effect of relevance feedback for at least three reformulated queries.

- Implement the positive feedback mechanism ($\beta = 1$) and analyse the changes in the ranked list for query lengths 3 and 5. How many terms are added/deleted and what is their effect on the original ranking of documents over each successive reformulation? Clearly demonstrate the effect.
- Implement the negative feedback mechanism ($\gamma = 1$) and analyse the changes in the ranked list for query lengths 3 and 5. How many terms are added/deleted and what is their effect on the original ranking of documents over each successive reformulation? Clearly demonstrate the effect.
- Implement the standard feedback mechanism ($\beta = 0.75$ and $\gamma = 0.25$) and analyse the changes in the ranked list for query lengths 3 and 5. How many terms are added/deleted and what is their effect on the original ranking of documents over each successive reformulation? Clearly demonstrate the effect.

Note:

- Submit a detailed pdf report on your observations and analysis, supported by the necessary code snippets w.r.t your program and results. **Plagiarised assignments will not be graded.**
- Upload your report and code (well documented) on Moodle to the folder provided before the deadline of **11.59PM on 26th October 2022 (No extension will be given).**

EVALUATION RUBRICS:

Part a: Rocchio's implementation and demo	- 10 marks
Part b: Implementation and analysis w.r.t Q.(a)	- 10 marks
Part c: Implementation and analysis w.r.t Q.(b)	- 10 marks
Part d: Implementation and analysis w.r.t Q.(c)	- 10 marks
Detailed report	-10 marks