

Department of Computer Science
The University of Melbourne
COMP90042 WEB SEARCH AND TEXT ANALYSIS (Semester 1, 2018)

Workshop exercises: Week 11

Discussion

1. Discuss the concepts of **query expansion** and **relevance feedback** and how they are related.
2. Discuss and give an example of the following forms of relevance feedback.
 - (a) User relevance feedback
 - (b) Pseudo relevance feedback
 - (c) Indirect relevance feedback
3. Is it possible to perform **query expansion** without **relevance feedback** and vice versa? Discuss!
4. Discuss the process of static **inverted index construction** and how it can be used to perform incremental index construction.
5. Why is a **logarithmic** index layout useful? What are the disadvantages of such an index structure?

Programming

1. Work on the project! :-)

Catch-up

- What is top- K retrieval and why is it useful?
- What are some of the reasons a search engine would want to offer a query completion service?

Get ahead

- The lecture links to a blog post discussing a more advanced index merging scheme used by Lucene. Read up on why Lucene uses this method compared to the logarithmic method discussed in the lecture.
- In general, similarity measures that we have discussed up till now are quite simplistic as they only take into consideration term frequency statistics. Modern search engines use many more advanced features to determine the relevance of a document. Can you think of some features that might be useful?