Interesting Notes/Excerpts - Optional

1. Taming the Text (TT)

a. Synonym expansion - For each token, synonyms are looked up in a thesaurus and added to the index. This is often done on the query terms instead of the index terms, since updates to the synonym list can be accounted for dynamically at query time without the need to re-index.

b. Moreover, many search engines will also perform multiple queries for every user query and then collate the results for presentation back to the user. This approach allows the engine to use multiple strategies to find the best results.

c. A known item search happens when a user specifically knows the name of the item (or close to it) and just needs to know where that item is in the store. A category search is much more general and often involves only a few keywords: televisions or piano music .

   In the case of known item search, failing to return the specific item in the top few results is considered a failure of the system. For category search, there's more wiggle room in what to return since the terms are often fairly general.

d. What to display to the user?

   - What parts of the document should be displayed? The title is usually a given, if there is one. How about a summary or access to the original content?
   - Should you highlight keywords from the query in the display?
   - How do you handle duplicate, or near-duplicate, results?
   - What navigation links or menus can you provide to enhance your user's experience?
   - What if the user doesn't like the results or needs to expand or narrow the scope of their search?

e. Cache statistics—Many systems cache query results and documents and it's useful to know how often the cache is hit or missed. If there are routinely few hits, it may be faster to turn off caching. (Not as relevant for now, as our system is reasonably slow for many queries)

f. Before we begin, a warning is in order: tuning a search engine can take up large chunks of time and result in no measurable improvement. It's best to

double-check that improvements are needed by monitoring both the quality and quantity aspects of a system before undertaking any tuning.