Ha Tran 53409673

Leon Luo 72198827

CS 121

Project 3 M3

Query Set used to Test: **\*\* #1 – 10 are queries that did poorly \*\* #11 – 20 did fine \*\***

1. Computr – misspellings are accounted for. No fix - intentional
2. Text – broad and general term. This did poorly and had non-specific results. No fix.
3. 121 – dates and numbered queries did poor because we did not assign contextual information to them; No fix – intentional.
4. 10/31/2019 – dates queries did poor because we did not intentionally search for dates nor store them in our indexers; not fixed - intentional
5. L – random letters were picked up by our indexer and yielded resulted; fixed by not allowing terms less than 3 letters.
6. How to write a program in C++ - Symbols in our query were not included. Query translated to “How write program.” No fix - intentional.
7. @@@@ - breaks the code; fixed by only allowing alphabetical queries.
8. Computer science – Documents rank high because they have high frequencies of “computer” and “science.” No fix; we do not support proximity matching.
9. Cristina Lopes – It is not looking for documents with “Cristina Lopes”. It is looking for the best scores for “Cristina” and the best score for “Lopes.” A document with a high frequency of “Cristina” could rank very high for this query. We did not do proximity checking. No fix.
10. AAA – Yields results because it searches for terms that start with “AAA”. This is an irregular behavior. No fix.
11. Android Studio – This did well because the query is quite specific. This yielded results directly related to the query. Projects related to Android Studio in the case.
12. Wongma – Specific names did well because they yield importance and had more weight to the query and thus results.
13. AVL – although it is one-termed, it is specific and resulted in URLs that directly related to the query. Acronyms that are 3 letters or more are allowed.
14. Algorithm – General, but specific enough to yield good results
15. Java- General, but specific enough to yield good results.
16. Alfred – A general name, but significant enough in the dataset to yield good results, especially given the dataset.
17. Internships – General, but specific enough to yield good results.
18. Klefstad – specific names did well.
19. Boo – although this is seemingly random, it did well because it is a specific name well known to ICS students.
20. Hash Tables – specific topic that yielded related results