Here I’ll illustrate the general methods we use and the conclusions we got.

To obtain the concrete test cases and results, you can refer to the attachment.

Test case 1:

Input: the name of several works.

Result and conclusions: All the works related are acquired, but for some queries, some works that are unrelated are also gained. That is to say, it has a pretty good recall, but the precision is not so perfect;

Test case2:

Input: some characters in a work

Result and conclusions: Almost all the works that include the character are given, but some works whose character have a same name with the one I give are also given, That is to say, the recall is great, but the precision is not so perfect;

Test case 3:

Input: some stop words

Result and conclusions: inputting all the stop words have no output;

Test case 4:

Input: some words in different tense and voice

Result and conclusions: Most words of different tense and voice have the same result, but some excepted, for example, irregular verbs;

Test case 5:

Input: some exact sentences in the original text

Result and conclusions: most queries have one or two result which include the article we want, but a few queries output relatively more result. The recall is great, and the precision is a little bit lower;

Test case 6:

To test the running time of queries of different length

Input: the queries that 1, 2, 5, 10 and 15 words

Result and conclusions:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 单词数 | 1 | 2 | 5 | 10 | 15 |
| 单次耗时（s） | 0.001093 | 0.001673 | 0.006776 | 0.012228 | 0.012891 |
| 0.000919 | 0.001671 | 0.00646 | 0.011842 | 0.013533 |
| 0.000926 | 0.001693 | 0.006066 | 0.007089 | 0.013623 |
| 0.00096 | 0.001747 | 0.005976 | 0.0071 | 0.01557 |
| 0.000972 | 0.001883 | 0.006789 | 0.012063 | 0.015983 |
| 0.000898 | 0.001671 | 0.007626 | 0.010846 | 0.013697 |
| 0.000895 | 0.001865 | 0.007755 | 0.007805 | 0.013518 |
| 0.000885 | 0.001331 | 0.00465 | 0.010015 | 0.011285 |
| 0.000941 | 0.001717 | 0.005607 | 0.011157 | 0.01188 |
| 0.00091 | 0.009837 | 0.006558 | 0.010901 | 0.01314 |
| 平均时间(s) | 0.00094 | 0.002509 | 0.006426 | 0.010104 | 0.013512 |

From our tests, the running time and the number of words in one query is positively related, however the concrete relationship is quite complex. In the hashing search, the running time and words quantity is linear dependent, and in the process of merging the results of each words, the running time and the total documents number is about a square relation, but the relation between the words number and document numbers is indeterminate;