# Experiment Design

In order to test the three different implementations of the Index search, a simple experiment conducted by the group is presented here. The three different implementations, simple Index, inverted index using hashmap and inverted index using treemap were compared in terms of the speed it took to retrieve the results of a search. To test different scenarios, three words with different characteristics were chosen:

* of: common word among websites;
* attend: less occurrences.;
* grouphisthebestgroup: no occurrences in any website.

Beyond that, the searches were conducted using three different files (varying in size) as the website database. The files used were tiny (10 KB), small (510 KB) and medium (9766 KB). The large file was not used because no query was managed to be completed due to time exception in the computer the experiment was conducted. Because the method “System.nanoTime()” present some variation at each occurrence, each result was recorded 5 times and the average value used as the parameter. The results were recorded in microseconds and are contained in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Word** | **Method** | **Database file** | | |
| **tiny** | **small** | **medium** |
| **of** | SimpleIndex | 368,8 | 7376,6 | 109340,2 |
| Inverted Index TreeMap | 309,8 | 4078,2 | 38068,6 |
| InvertedIndex HashMap | 320,4 | 5188,8 | 33097,6 |
| **attend** | SimpleIndex | 415 | 2852,4 | 38139,6 |
| Inverted Index TreeMap | 135 | 67,6 | 192 |
| InvertedIndex HashMap | 195,2 | 82 | 107,4 |
| **grouphisthebestgroup** | SimpleIndex | 464,6 | 1140 | 19755,4 |
| Inverted Index TreeMap | 138,2 | 50,2 | 67,6 |
| InvertedIndex HashMap | 127 | 44,6 | 98,6 |

First, the observations show that the simple index query time grows according to the size of the file, no matter what word is being searched. However,for the tiny file, the difference between the methods are not so pronounced yet, even though treemap and hashmap perform better for the two last words, for the word “of” there is basically no difference.

When the size of the database grows, the simple index presents a much worse performance than the other two methods. Treemap and hashmap have very similar performance for both words with little or no occurrences. However, the hashmap is still slightly faster for the word “of” which appears more in the file when the datafile is larger (medium datafile). Therefore, we conclude this experiment by choosing the Inverted Index Hashmap as the method used for the search engine.