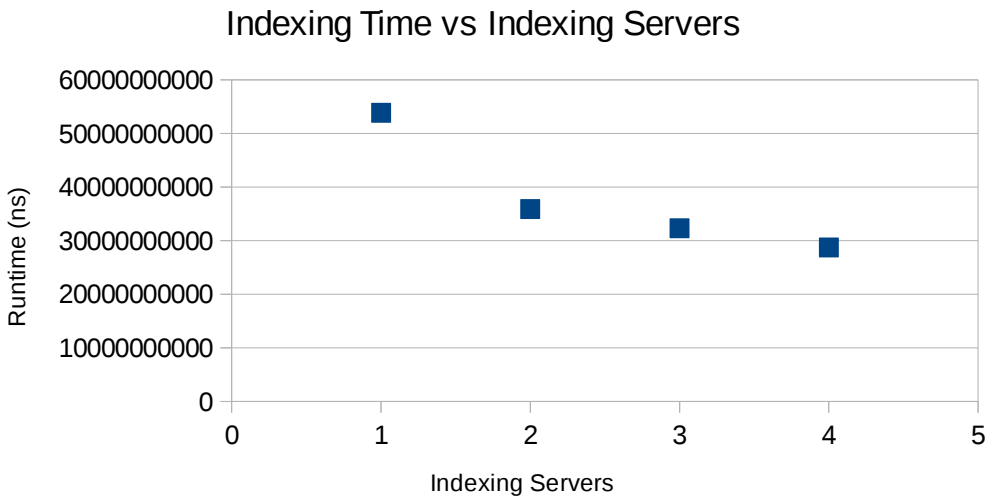
Forbes Turley
Jon Albert

# Part 1 Results

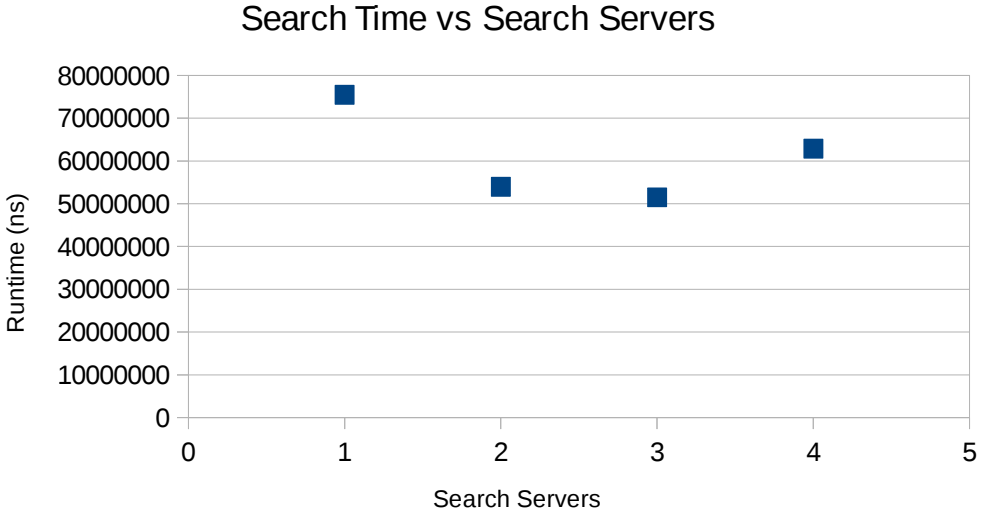
First, the indexer was tested against the largest of the given index files, War and Peace. This file is 3.2 mb, and was found to take well under a second to run on a single indexer. Adding additional indexers that would share the workload was tested, with comparisons made against 1,2,3, and 4 indexing servers. However, the runtimes were found to be virtually identical with regardless of the number of indexing servers parallelizing the work, and this was assumed to be because the time for communication between the Index Master and the Indexing Severs was large relative to the time required to index the file.




For these reasons, a much larger text file composed of many books concatenated was generated (1.6 gb). Then, the same experiment was performed, the results of which are shown below.

Clearly, adding more Indexing servers on such a large document makes a huge difference. Going from just 1 to two Indexing Servers nearly halves the time, because on a file this large the time required to index the file relative to the time required for communication is very large.

Next, the same set of experiments was performed with the Search Servers. However, no noticeable difference was observed, and so 4 files were indexed to run search queries against. For all tests, the words searched for were “yes, no, what, why”. The results are shown below.



The time required for the search initially decreases, then increases as more servers are added. This is likely because each additional Search Server requires more communication time with the Search Master, and because most of the inverted index can probably be stored in memory, performing the search is extremely fast.