ASSIGNMENT 2 REPORT

Experimentally compare the performance of your implementations of the exact retrieval with that of inexact retrieval methods. What are your performance measures? How are they measured and compared? In your report, you will provide justifications for these measures. Use graphs, as appropriate, to compare performance and describe insights and conclusion drawn from your results. Also, include details for your implementation in the report that may have an impact on performance.

Performance measures:

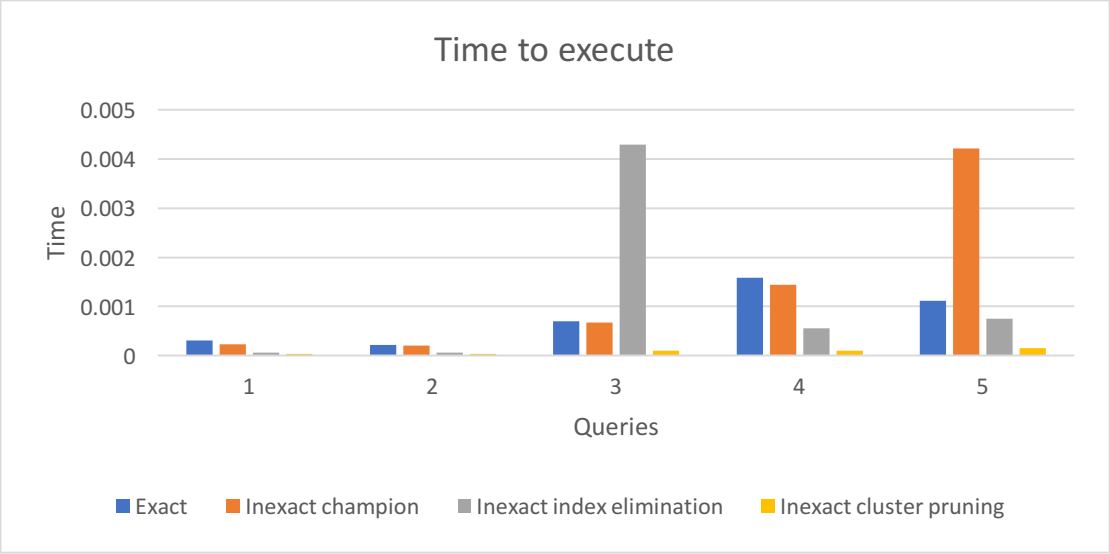
There are wo main components of performance measure

1. Retrieval time
2. Index time

1. Retrieval time

Query	Exact retrieval	Inexact retrieval		
		Champion List	Index Elimination	Cluster Pruning
['with', 'without', 'yemen']	0.0003151894	0.0002288818	0.0000619888	0.0000400543
['with', 'without', 'yemen', 'yemeni']	0.0002160072	0.0002031326	0.0000619888	0.0000381470
['rest', 'ready', 'more']	0.0006959438	0.0006790161	0.0042889118	0.0001080036
['with', 'ready', 'the', 'war']	0.0015799999	0.0014429092	0.0005629063	0.0001080036
['time', 'political', 'less', 'then']	0.0011160374	0.0042049885	0.0007560253	0.0001578331

Cluster Pruning < Index Elimination < Champion List < Exact retrieval



From above data and graph, we can say that Cluster Pruning is the fastest method to retrieve documents, then, index and champion list. And, direct retrieval takes too long time to fetch document based on given query.

2. Indexing Time:

Indexing	Time to index
tf-idf Index	0.5056769848
Champion Index	0.644336939
Cluster Pruning	0.3902151585

Here, also you can see that cluster pruning takes less time than other retrieval methods.

Reason:
The reason cluster pruning is always fastest because it compares find closest document with high cosine value. The time to perform this operation is less compared to calculating cosine scores of each document in exact retrieval.
For inexact index elimination, its takes more time because it also calculates cosines scores for half query terms, that's why it is also faster than inexact champion list.
Thus, **cluster pruning** is faster.