

Information retrieval is concerned with finding relevant documents that satisfy user information needs. The user is only interested in the few of the most important document (top k). That is out of all the documents only few are useful for the user satisfying all the user query. For getting the top k document we need a score for that. For the performance measures we can use the precision and recall. Also, we can see the time taken by the output of different top k query method, which means that less time taken for retrieving top k documents according to the user query and getting the relevant document, more the good the method is. The index elimination is taking the least time.

Output of the program with the time taken to retrieve the output(top k) for different methods like Exact retrieval, index elimination, champion list and cluster pruning.

```
-----BUILDING INDEX-----
Retrieved in 4.360517978668213 sec.
-----Exact Top K query Result-----
exact query result:
Doc id: 298
Doc id: 67
Doc id: 187
Doc id: 409
Time it is taking is for exact top k 0.0040035247802734375 sec.
-----Inexact top k result-index elimination-----
Index Elimination top k:
Doc id: 298
Doc id: 67
Doc id: 26
Doc id: 1
Time it is taking for index elimination is 0.01564502716064453 sec.
-----Inexact top k result-champion result -----
top k from champion list:
Doc id: 298
Doc id: 67
Doc id: 47
Doc id: 49
Time it is taking for champion list is 0.6773524284362793 sec.
```

Exact Top k retrieval:

In exact top k tf-idf for documents and query are calculated. Cosine score is taken and top k are retrieval.

Index Elimination:

Everything is same as exact top k but the selected documents are selected.

Champions List:

A new list of documents is constructed with the r docs of highest weight in posting list.

Cluster Pruning:

In this clusters with leaders are formed and a query is checked in all the documents of cluster and score is computed.