

An Experimental Study of Multi-stage Retrieval Systems

Mohammed Yusuf Ansari
 Advisor: Mohammad Hammoud

Information Retrieval

- Information Retrieval (IR) is concerned with *search* over large *unstructured data* like web pages, emails, and image libraries, among others
- Any IR system has two major objectives:



Speed (or Efficiency)



Accuracy (or Effectiveness)

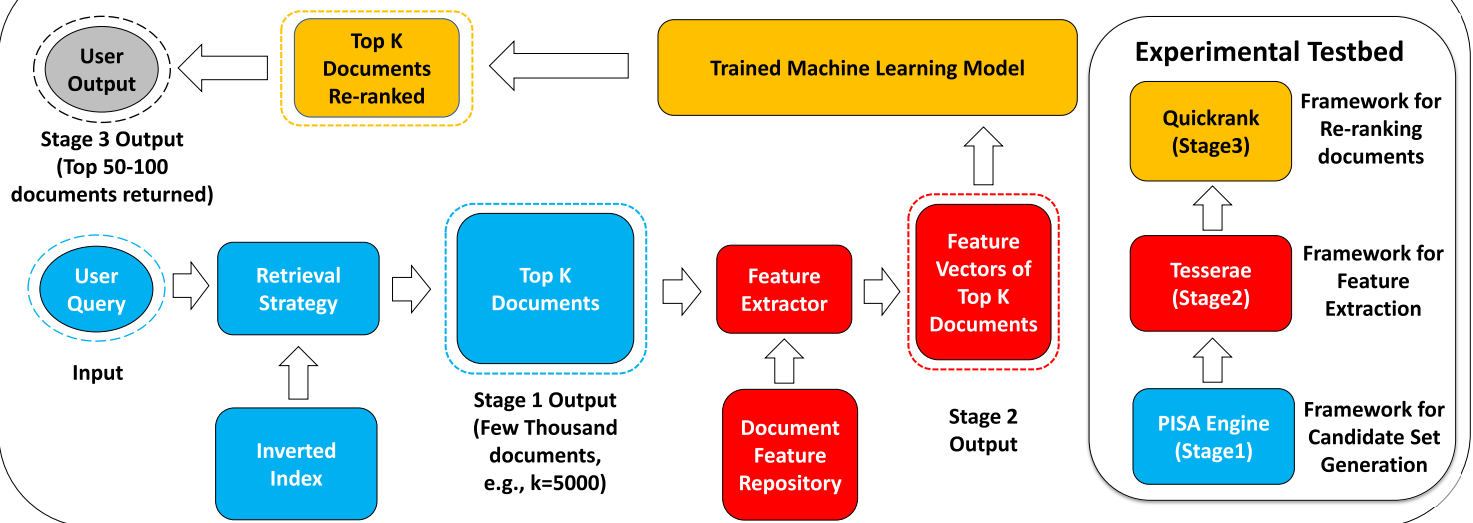
Problem Description

Observation: Past work aimed at optimizing a particular configuration parameter of a single stage in standard three-stage retrieval systems

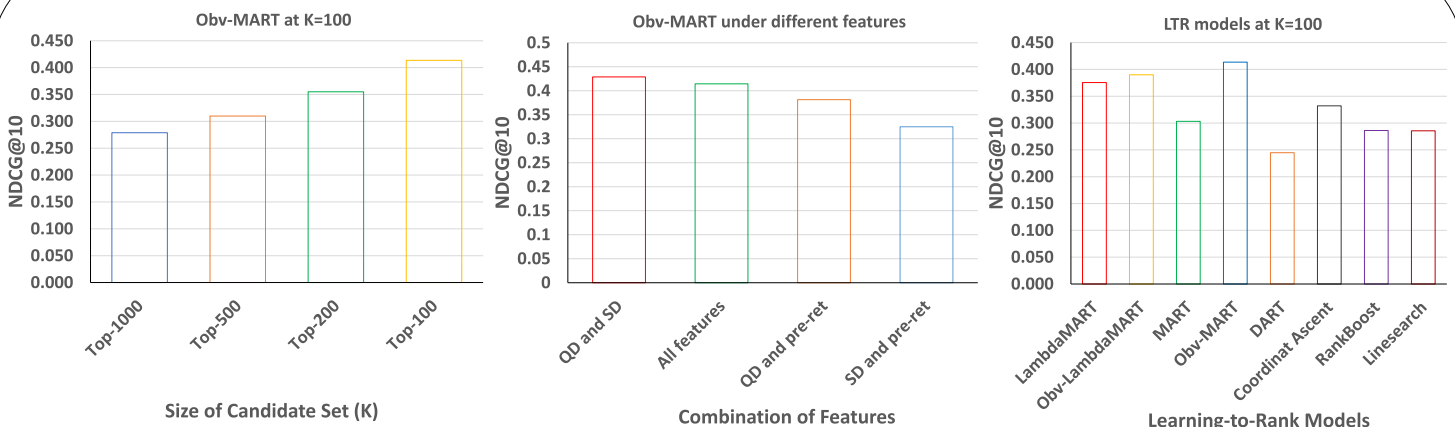
Problem: Learning-to-rank models in stage 3 were developed in obliviousness to the configurations of the first two stages

Solution: A comprehensive empirical study to examine and leverage the performance correlation between the three different stages of IR systems

Three Stages of Multi-stage Retrieval Systems



Experimental Evaluation



- ⇒ For **candidate set generation stage**, we recommend candidate set size (K) equal to 100
- ⇒ For **feature extraction stage**, we recommend the combination of query-dependent and static-document features
- ⇒ For **candidate set re-ranking stage**, we recommend Obv-LambdaMART or Obv-MART model