

An Experimental Study of Multi-stage Retrieval Systems

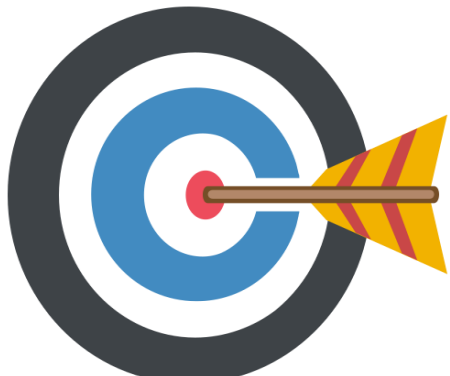
Mohammed Yusuf Ansari
Advisor: Mohammad Hammoud

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)

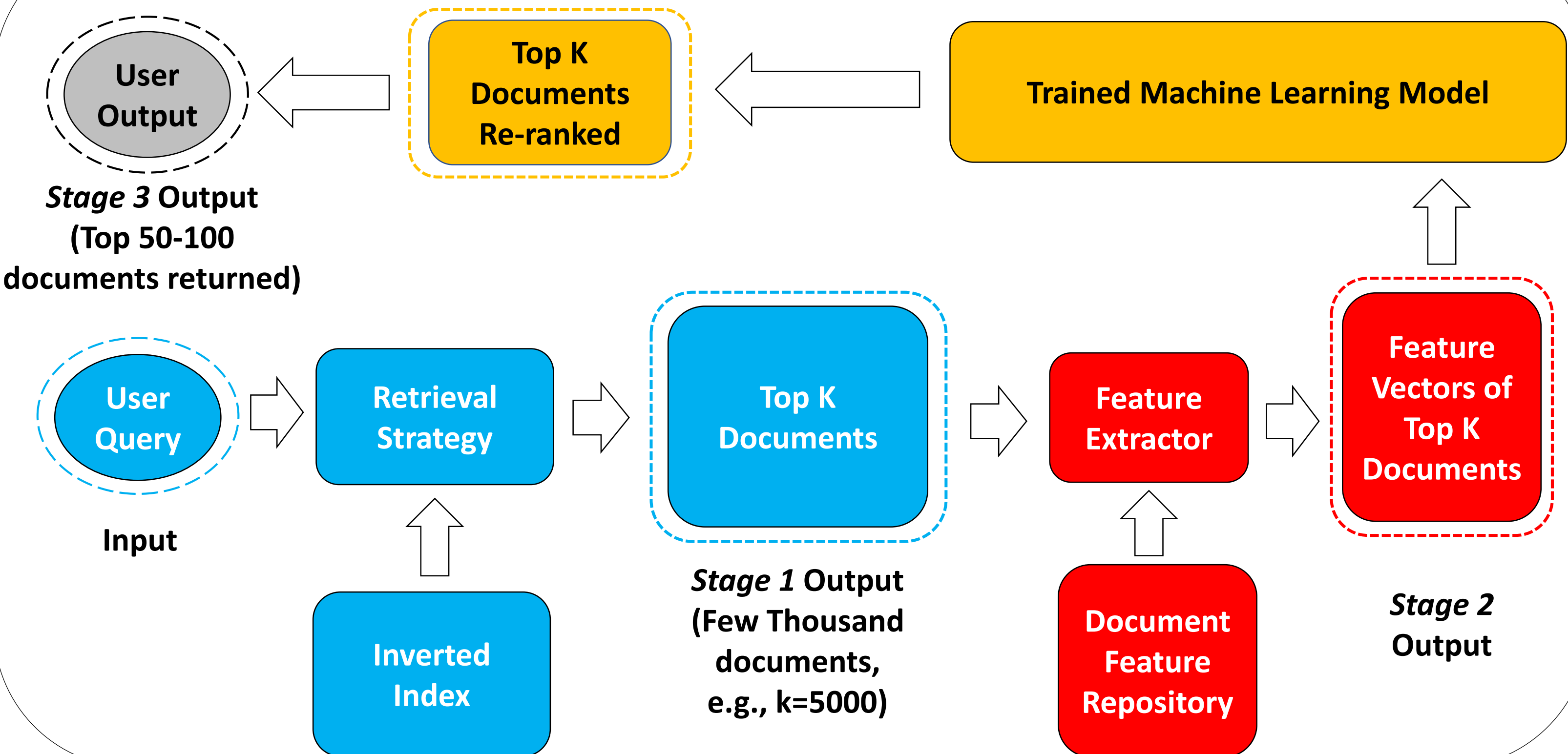
Problem Description

Observation: Past work aimed at optimizing a particular configuration parameter of a single stage in multi-stage IR systems

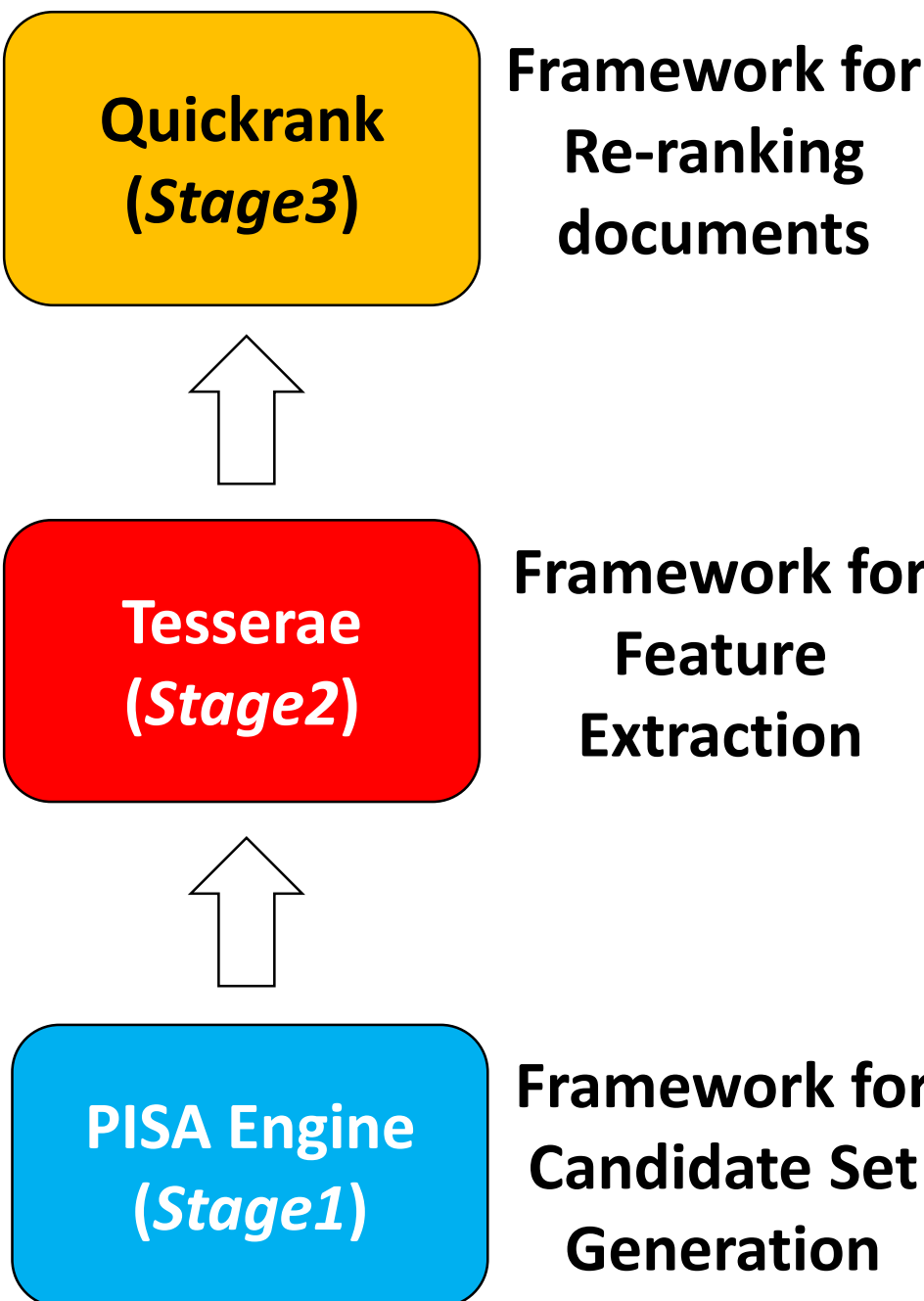
Problem: Learning-to-rank algorithms were developed in obliviousness to the configurations of different stages

Solution: A comprehensive empirical study to understand the performance correlation between the different stages of IR systems

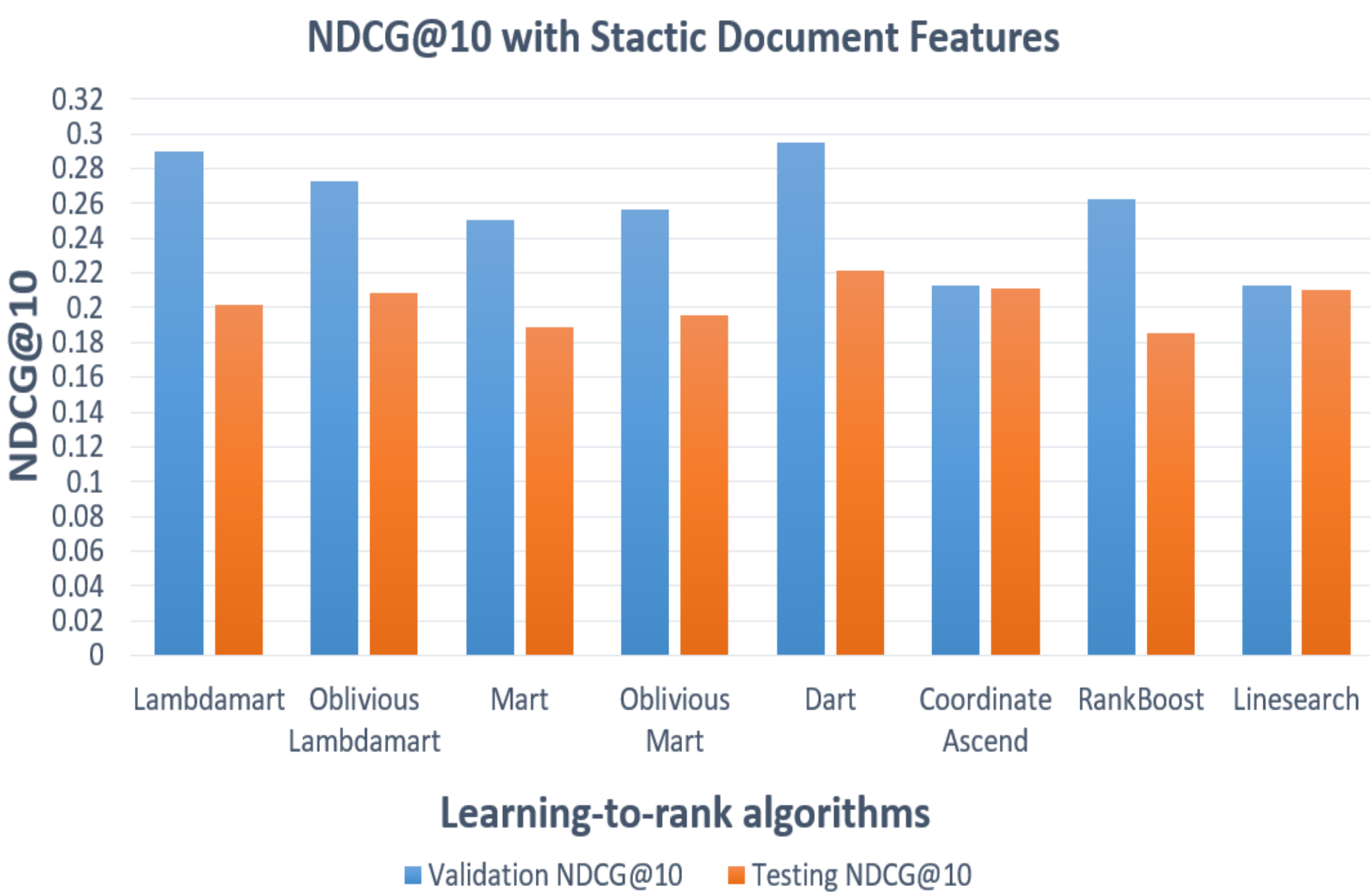
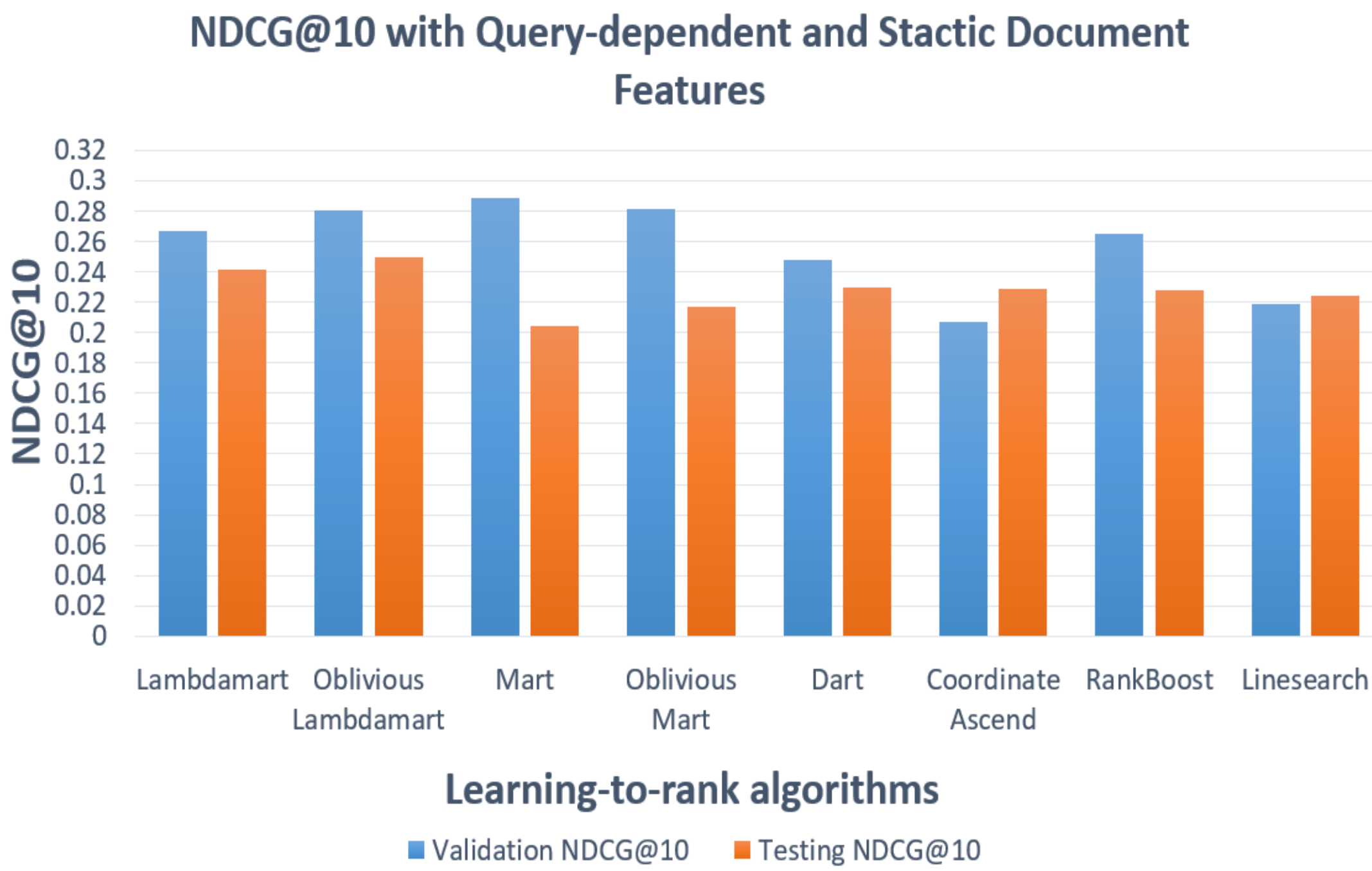
Three Stages of Multi-Stage Retrieval Systems



Testbed



Experimental Evaluation



⇒ Experimental results verify that with many features ensemble-based algorithms outperform other LTR algorithms

⇒ LTR models trained with query-dependent and static document features perform better than models trained with only static document features