

Let's measure run time!

Extending the IR replicability infrastructure to include performance aspects

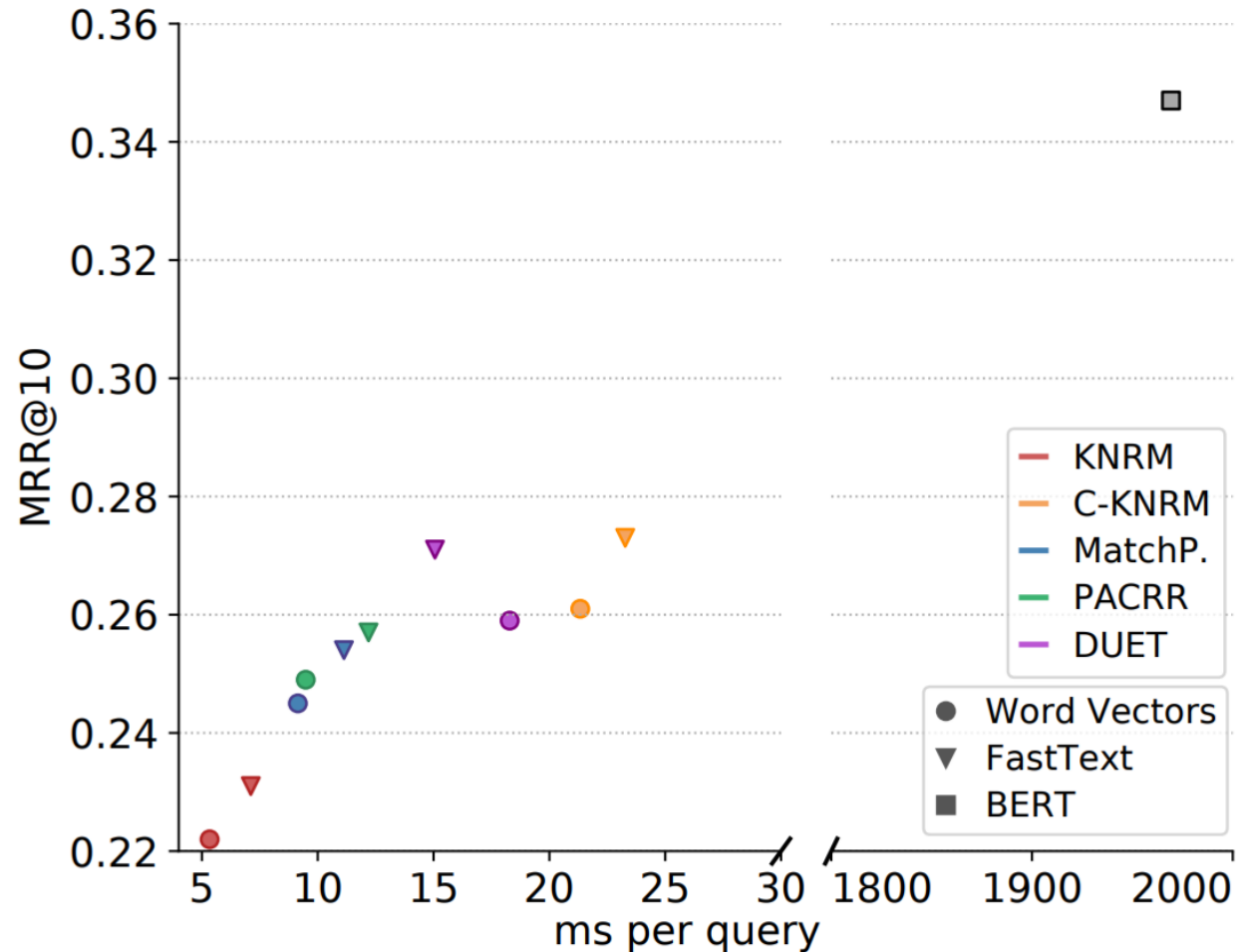
Sebastian Hofstätter
sebastian.hofstaetter@tuwien.ac.at

Allan Hanbury
hanbury@ifs.tuwien.ac.at

Performance vs. Effectiveness

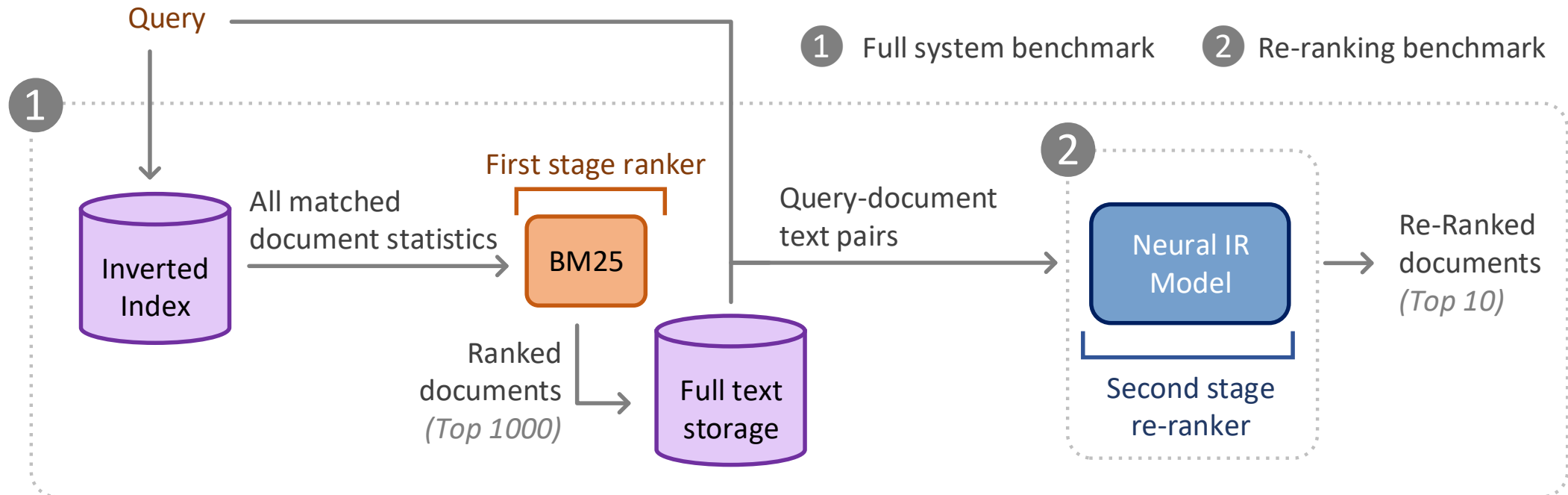
A case study of neural IR models

Performance-Effectiveness Tradeoff



- 2 classes of models:
 - BERT & non-BERT
- Fine-grained tradeoffs between non-BERT models
- FastText is a stable improvement over WV

Measuring Performance





Thank you

Let's measure run time!

- Docker-infrastructure allows evaluation beyond effectiveness measurements
- Let's extend the infrastructure to measure performance
- Especially needed for recent neural IR models
 - Essential trade-off considerations between architectures