Information Processing and Retrieval

Master in Informatics Engineering and Computation at FEUP, U.Porto

Diogo Fonte
Rodrigo Figueiredo
Sofia Rodrigo
Vitor Cavaleiro

U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Improvements

- P-R curve formula correction

- Field boosts

- Semantic search

- Additional data processing

P-R Curve Formula Correction

dividing every step by the relevant results found up to that point

→

dividing by all relevant results

Field Boosts

different field boosts for each query → all queries are evaluated using the same boosts

Semantic Search

```json
{
  "name":"knn_vector",
  "class":"solr.DenseVectorField",
  "vectorDimension":"384",
  "similarityFunction":"cosine",
  "knnAlgorithm":"hnsw",
  "hnswMaxConnections":"10",
  "hnswBeamWidth":"40"
}
```

```json
{
  "name": "bert_vector",
  "type": "knn_vector",
  "indexed": true,
  "stored": true
}
```

schema.json

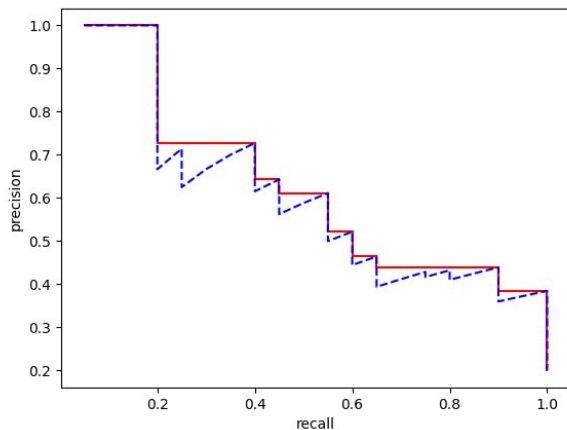# Additional Data Processing and Resizing

+

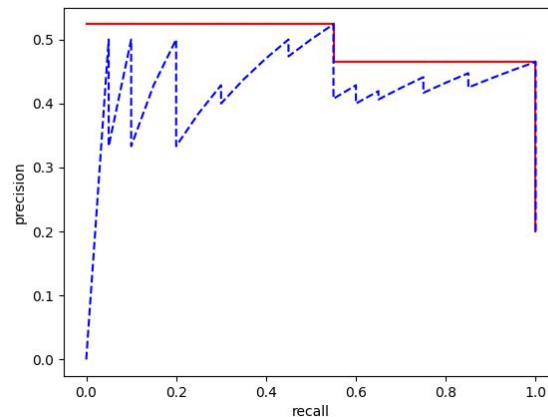Reducing 75% data ensuring we did not eliminate relevant news for our queries

Evaluation - Query 1

trump speaking on immigration crisis

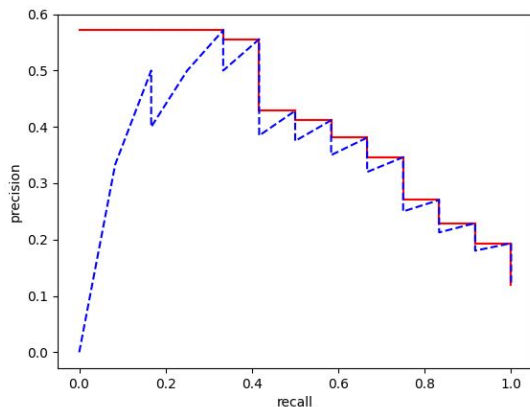| Metric	Without boosts	With boosts
AP	0.45	0.62
P@10	0.4	0.7

(simple schema without re-ranking)

(improved schema with re-ranking)

Evaluation - Query 2

lebron good performance
in games he won

Metric	Without boosts	With boosts
AP	0.39	0.39
P@10	0.5	0.3

(simple schema without re-ranking)
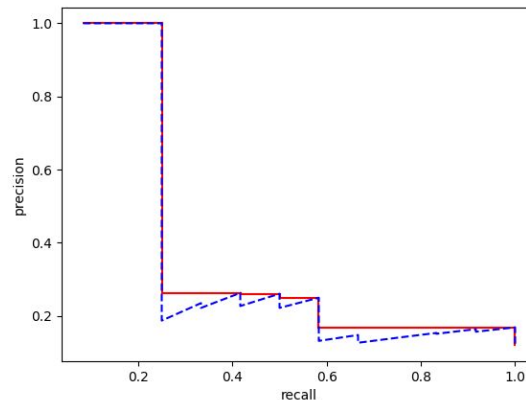
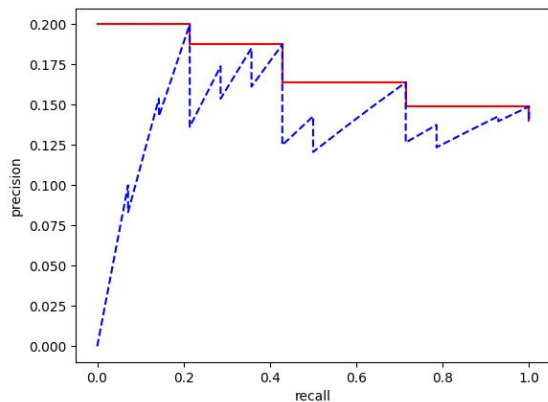
(improved schema with re-ranking)

Evaluation - Query 3

homicide investigated
by fbi in 2017

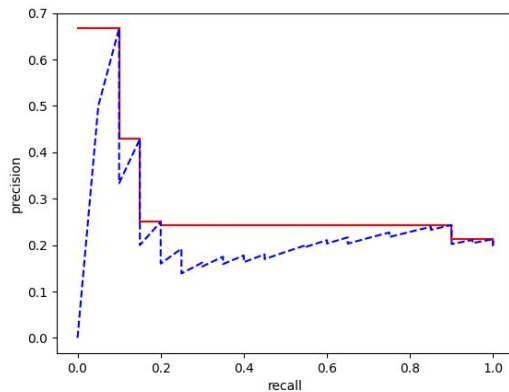| Metric	Without boosts	With boosts
AP	0.15	0.58
P@5	0.1	0.6

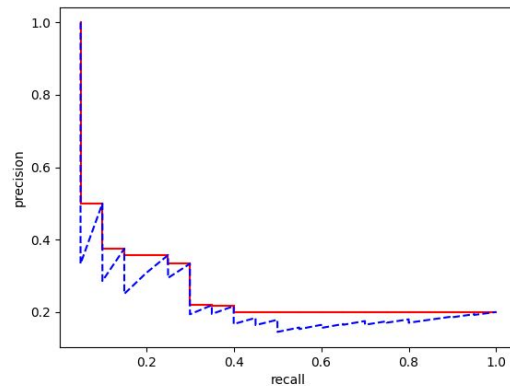

(simple schema without re-ranking)
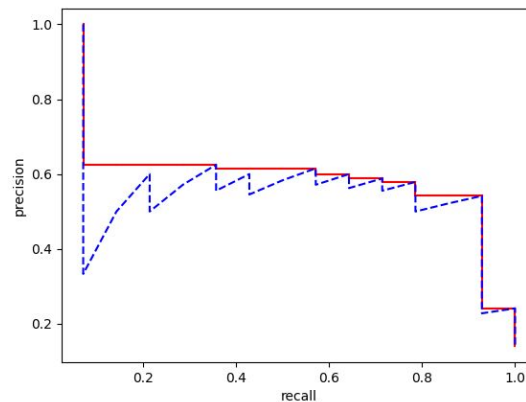
(improved schema with re-ranking)

Evaluation - Query 4

conflicts between republicans and democrats about gun ownership

Metric	Without boosts	With boosts
AP	0.25	0.27
P@10	0.3	0.3

(simple schema without re-ranking)

(improved schema with re-ranking)