

A Test Collection for Entity Retrieval in DBpedia



Krisztian Balog, University of Stavanger, Norway
Robert Neumayer, Norwegian University of Science and Technology

<http://bit.ly/dbpedia-entity>

Entity retrieval

Addressing information needs that are better answered by returning specific objects instead of just any type of documents.

Action points

identified at the JIWES workshop at SIGIR'12*

1. Getting more representative information needs and favouring long queries over short ones.
2. Limiting search to a smaller, fixed set of entity types (as opposed to arbitrary types of entities).
3. Using test collections that integrate both structured and unstructured information about entities.

* K. Balog, D. Carmel, A. P. de Vries, D. M. Herzig, P. Mika, H. Roitman, R. Schenkel, P. Serdyukov, and T. Tran Duc. **The first joint international workshop on entity-oriented and semantic search (JIWES)**. SIGIR Forum, 46(2), December 2012.

DBpedia

- Effort to extract structured information from Wikipedia.
- One of the most comprehensive knowledge bases on the web, describing 3.64M entities (version 3.7).
- The DBpedia Ontology defines 320 classes, organized into a 6 levels deep hierarchy; 1.83M entities are classified within this ontology.

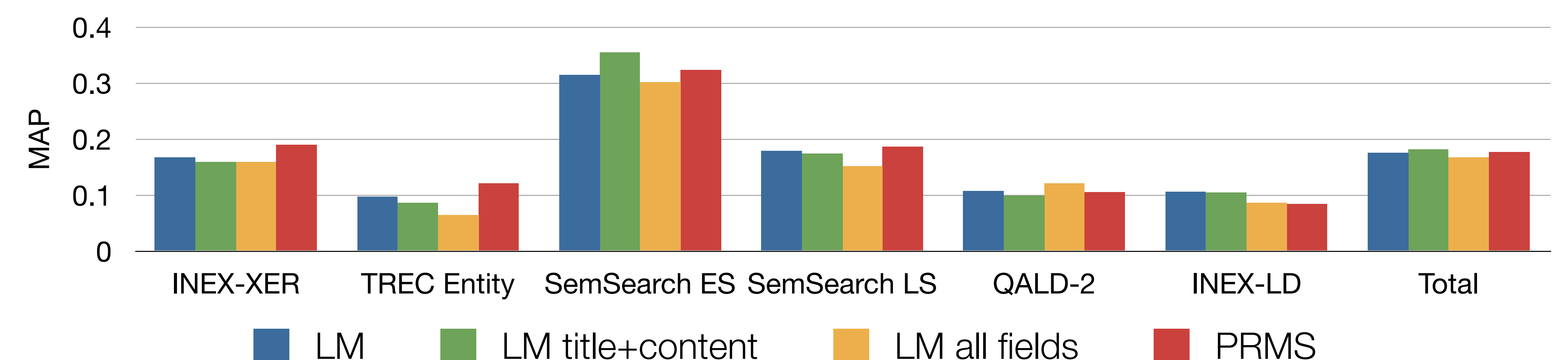
Our contributions

- Developed a test collection based on DBpedia, by synthesizing queries from previous benchmarking evaluation efforts.
- Provided a set of baseline results using language models and BM25.
- Performed an initial topic-level analysis.

Benchmarking evaluation campaigns

| Query set | #queries (orig) | #queries (used) | avg(q) | avg(#rel) |
|--|-----------------|-----------------|------------|-------------|
| INEX-XER "US presidents since 1960" | 55 | 55 | 5.5 | 29.8 |
| TREC Entity "Airlines that currently use Boeing 747 planes" | 20 | 17 | 6.7 | 13.1 |
| SemSearch ES "Ben Franklin" | 142 | 130 | 2.7 | 8.7 |
| SemSearch LS "Axis powers of World War II" | 50 | 43 | 5.4 | 12.5 |
| QALD-2 "Who is the mayor of Berlin?" | 200 | 140 | 7.9 | 41.5 |
| INEX-LD "England football player highest paid" | 100 | 100 | 4.8 | 37.6 |
| Total | | 485 | 5.3 | 27.0 |

Baseline results



- Various query sub-sets exhibit different levels of difficulty (as expected).
- Most differences are not statistically significant.
- A large number of topics is affected, but the number of topics helped/hurt is about the same.