Experiments with Semantic-flavored Query Reformulation of Geo-Temporal Queries

Nuno Cardoso¹ and Mário J. Silva²

¹University of Lisbon Faculty of Sciences, LaSIGE, Lisbon, Portugal and SINTEF Language Technologies Group, SINTEF ICT, Oslo, Norway
²University of Lisbon Faculty of Sciences, LaSIGE, Lisbon, Portugal
ncardoso@xldb.di.fc.ul.pt
mj Silva@di.fc.ul.pt

Motivation

Simple queries work well with simple IR systems (term-match based document retrieval).

Query expansion (QE) helps...

More terms → matching odds increased → better retrieval results
... but sometimes not.

Bad selection of terms → drift from initial topic → noisy results

Why don’t we understand what the user want, instead of retrieving what the user said?
Why don’t we reason answers instead of guess terms? Is there a better approach for elaborated queries with geographic and temporal scopes?

Queries have entities, and entities have semantic information.

Statistics-based QE works at term level.

Reasoning-based QE requires working at entity level, where its semantic role is grounded.

Objectives

• **Build** a semantically-flavored query reformulation (SQR) approach, using external knowledge resources and reasoning approaches to reformulate queries at entity level.

• **Evaluate** how suitable is a SQR approach on retrieving documents for geographically-challenging queries.

System overview

1. Detect and ground entities in user queries and in the whole document collection

   requires a named entity recognition (NER) software.

2. Use external knowledge bases (Wikipedia, DBpedia, geographic ontologies) to access more information about entities.

Terms

<table>
<thead>
<tr>
<th>NEs</th>
<th>Entities</th>
<th>Geographic Entities</th>
<th>Temporal Entities</th>
</tr>
</thead>
</table>

3. Index terms and semantic information (NEs, entities, places and time expressions)

4. Extend a retrieval engine to cope with term / semantic indexes, reformulate queries to use against those indexes

Experiments and results

1. Baseline run, plain terms with no expansion
2. Automatic run, with DBpedia ontology lookup
3. Supervised run, with DBpedia ontology lookup
4. Extended run, with DBpedia abstract entities

We would like to thank Jorge Machado, for the support and development of LGTE. Diana Santos for comments in the paper, and SINTEF ICT. This work is supported by FCT for its LAASIE Multi-annual support, GReaSE II project grant PTDC/ETC/6314/2008 and a PhD scholarship grant SFRH/BD/45548/2008, and by the Portuguese Government, the European Union (FEDER and FSE) through the Linguaspaia project, under contract ref. POSC/3391.3/NC/MIC, UMIC and FCCN.