Combining NLP and semantics to support a query expansion system

María G. Buey, Ángel Luis Garrido
IIS Department, University of Zaragoza and Heraldo de Aragón (Spain)

Problems in relational database searches

- **Term mismatch problem**: to retrieve data that exactly match the user keywords and indexers and users do often not use the same words.
- **"Documentary silence"**: losing relevant information that the user is trying to get (low recall).
- **"Documentary noise"**: needless data that are not interesting for the user (low precision).

Solution & Contributions

- **Automatic Query Expansion (AQE)**: to expand the original query with other words that best capture the actual user intent or that simply produce more useful query that is more likely to retrieval relevant documents.
- **Exploit linguistic and semantic techniques** to improve the quality of the keyword-based search process on relational repositories of the Heraldo Media Group.
- **SQX-Lib focuses on automatically and semantically expanding** the scope of the searches and fine-tune them by taking advantage of the Named Entities (NE) included in these queries.

1. Analysis of the query string

- Obtain logical query structure
- Morphological analysis of the query
- Obtain exact words
- Obtain Named Entities
- Eliminate Stop Words
- Select terms to process

2. Processing of terms

- Obtain lemmas
- Obtain synonyms
- Disambiguation
- Select the closest synonyms

3. Expansion of the query

- It reconstructs and expands the query from the relevant data extracted in each of the previous tasks while keeping the initial logical structure of the query

References


Acknowledgements

Thank you to Raquel Trillo, Sergio Ilarri and HERALDO DE ARAGON S.A.

Contact

General information: María Granados Buey (mgbuey@unizar.es)
Research group SID (http://sid.cps.unizar.es)