The XLDB Group at CLEF 2005 ad hoc

Nuno Cardoso*, Leonardo Andrade*, Alberto Simões† and Mário J. Silva*
*XLDB Group – Computer Science Department, Faculty of Sciences, University of Lisbon
†Computer Science Department, University of Minho
{ncardoso, leonardo, mjs} @xldb.di.fc.ul.pt, ambs@di.uminho.pt
http://xldb.di.fc.ul.pt

Tumba at CLEF 2005 ad hoc

Objectives

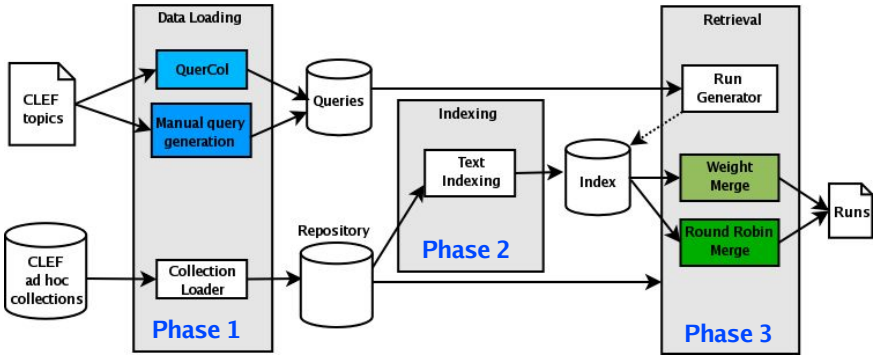
- Evaluate new methods and algorithms for the ad hoc task.

Improvements

- An automatic query generator, QuerCol.
- Implementation of the 'OR' operator on query strings.
- New methods for ranking and merging result sets.
- Topic translations for English to Portuguese bilingual runs.

http://www.tumba.pt

Architecture



Phase 1: Data Loading

CLEF collection loaded into repository.

QuerCol generates automatic queries from topics.

Additional queries created manually.

Phase 2: Indexing

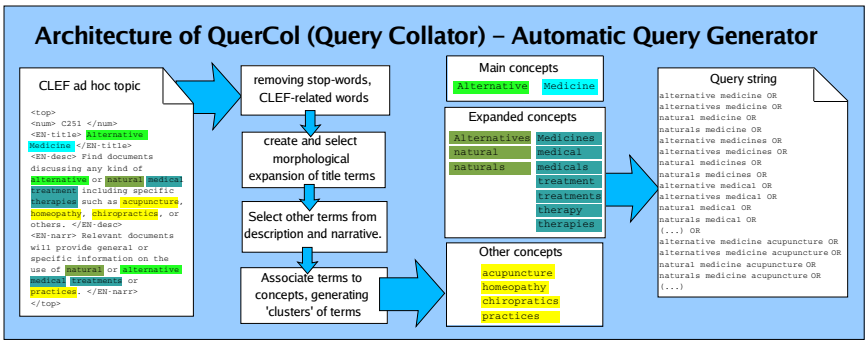
Documents are indexed, generating a term index.

Phase 3: Retrieval

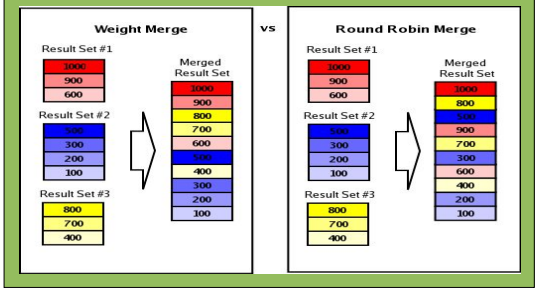
Queries processed.

Results ranked and merged using alternative methods. Runs generated.

Architecture of QuerCol (Query Collator) – Automatic Query Generator



Result Set Merging Approaches



CLEF 2005 ad hoc monolingual & bilingual subtasks

Experiments

Monolingual PT

Manual Query vs **Automatic Query** — Queries generated by QuerCol

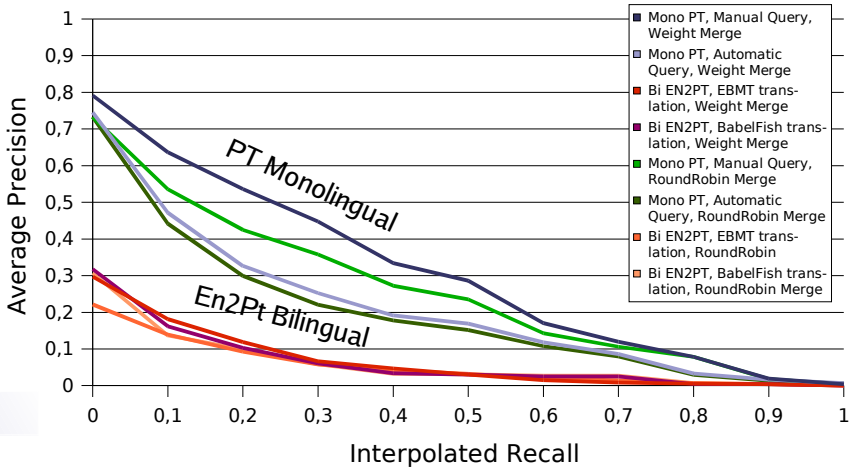
Weight Merge — Result set merging based on the sum of Okapi ranking values vs **Round-Robin Merge** — Result set merging by iteratively picking the best result of each set

Bilingual EN 2 PT

EBMT Translation — Example Based Machine Translation methods, based on parallel corpora vs **BabelFish Translation** — Translation by BabelFish, a web translation tool

Results

XLDB @ CLEF2005 ad hoc



Conclusions

- Weight Merge better than Round-Robin Merge
- Manual queries better, but automatic ones had good performance.
- EBMT better than BabelFish, with more relevant results with less retrieved documents.