

Summary & D4

What was tried

- Lots!
 - Stanford coref
 - Closed class answer sets
 - Query reformulation: MRS
 - Query expansion: WordNet, pseudo-relevance feedback
 - Query classification: Word-based, SVM ECC, MaxEnt
 - Answer extraction:
 - MRS matching, answer type filtering
 - Web-boosting/redundancy
 - Clean-up, bug fixes

What helped

- Big wins:
 - Clean-ups, bugfixes
 - MRS answer extraction
 - Web-boosting
- Small improvements:
 - Coref
 - Answer class filtering

Things that didn't work

- Sometimes:
 - WordNet expansion/reformulation
 - Maybe due to over-production
 - Other expansion
- Some query classification

Observations

- Expansion/reformulation ‘sensitive’
 - If over-aggressive, can hurt
 - Restrained approaches can have small gain

- Engineering is an issue:
 - Parallelizing
 - Caching
 - Web-scraping delays/blocking/quirks

Results!

- Major improvements:
 - Best MRR scores (including post-deadline):
 - Lenient: 0.345-0.37
 - Strict: 0.26
- Looking forward to even more!

D4: Test Data

- New questions:
 - evaltest directory
- New documents:
 - AQUAINT-2
 - On patas as LDC corpus
 - Format change
 - Orig: <DOC>
 - <DOCNO> NYT19981101.0001 </DOCNO>
 - <DOCTYPE> NEWS STORY </DOCTYPE>
 - New:<DOCSTREAM>
 - <DOC id="AFP_ENG_20041001.0001" type="story" >
 - <HEADLINE>

Goals

- Improve QA effectiveness
 - Focus:
 - Error analysis
 - Identifying contributors to errors
 - Implement strategies to improve
 - Each team member should 'own' some aspect
- Emphasize answer extraction