

ReQuery

Lisa Gress

Kathryn Nichols

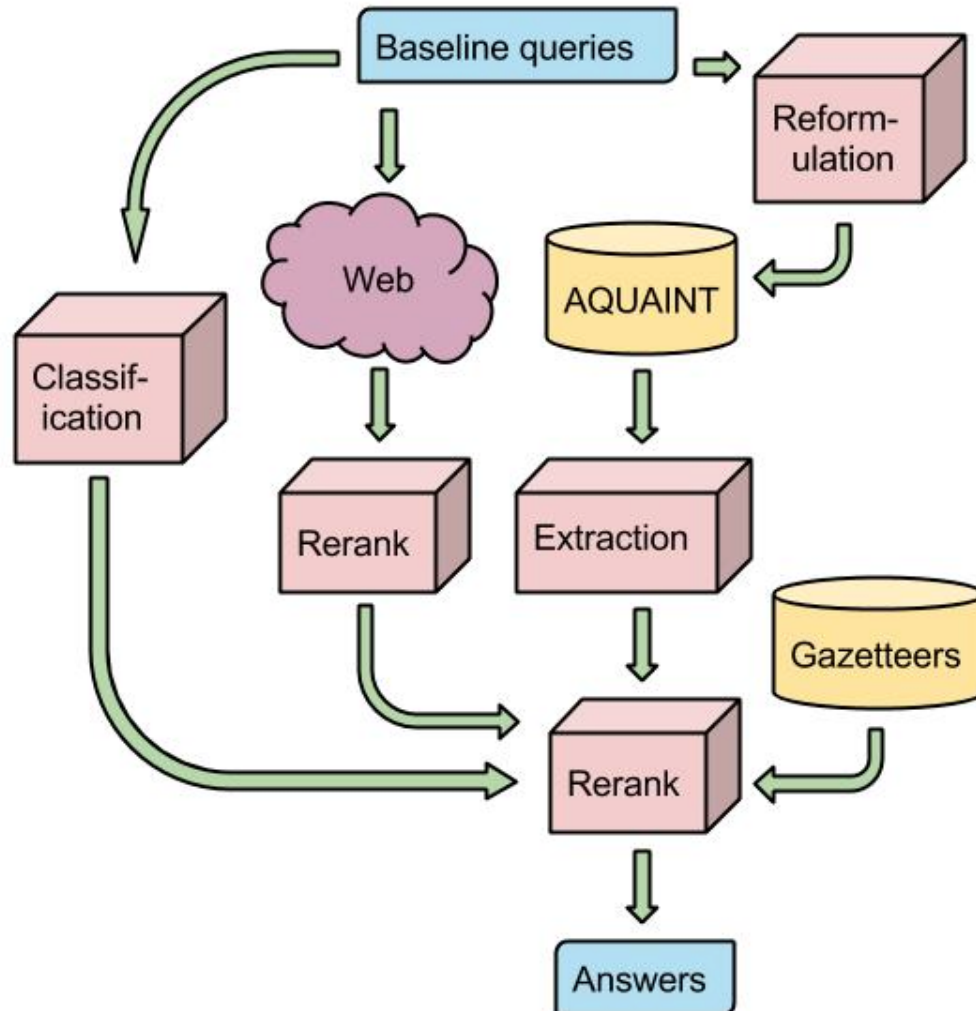
Shannon Watanabe

June 6, 2013

Background

- Overall system architecture
 - PowerAnswer-2: Harabagiu et al. (2005)
- Type taxonomy and classification
 - Li and Roth (2005)
- Redundancy
 - Lin (2007)

System pipeline



Question Classification

- Li & Roth taxonomy
- Additional fine types for pattern matching
 - Coarse types
 - Mallet MaxEnt classifier
 - ~86% accuracy on devtest
 - Fine types
 - Wh-word identification
 - 'When' → date
 - Headchunk pattern matching
 - 'What country ... ?' → country
- Supertype of a fine type overrides classifier when confidence below 85%

Retrieval - Web

Bing and Google caches

- xgoogle module
- Pattern module
- baseline queries (minimal reformulation)
- 100/50 snippets per question (Google/Bing)
- cached with pickle

For each snippet

- created all possible n-grams ($n=1, 2, 3, 4$)
- filtered, scored and ranked (Lin 2007)

Retrieval - AQUAINT

Passage retrieval

- 2 sentence sliding window; IDF + overlap
 - strict: 0.1263
 - lenient: 0.2282

Fragment retrieval

- top 20 fragments/question; Lucene scoring
- varied length of fragment, fragments/doc
 - strict: ~0.36
 - lenient: ~ 0.50

Retrieval - AQUAINT

Document retrieval - used in final system

- top 20 docs per question; Lucene scoring
 - strict: 0.4115
 - lenient: 0.5868

For each document

- created all possible n-grams (n=1, 2, 3, 4)
- filtered (same as web snippets)
- sent to scoring module

Answer Extraction

Web boosting

- Match any token of answer candidate to any token of webgram
 - if match, add inverse rank of webgram to score

Question-type boosting

- Gazetteers
- Regex

Top 20 n-grams were selected as answers

Answer Extraction - Q type

- Gazetteers:

- 28 gazetteers for select fine types
- Find type of ngram in gazetteer dictionary
- Gazetteer type = question type → boost

- Regex:

- Patterns for dates, digits, etc.
- Pattern type = question type → boost

33 questions (8%) had the correct type & predicted answers contained GS answer.

2006 (403 total):

- 144 predicted
- 131 correct (91.9%)
 - 60.7% gazetteer
 - 30.3% regex
 - 9% both

2007 (307 total):

- 85 predicted
- 76 correct (89.4%)
 - 63.2% gazetteer
 - 21.1% regex
 - 9.2% both

Results

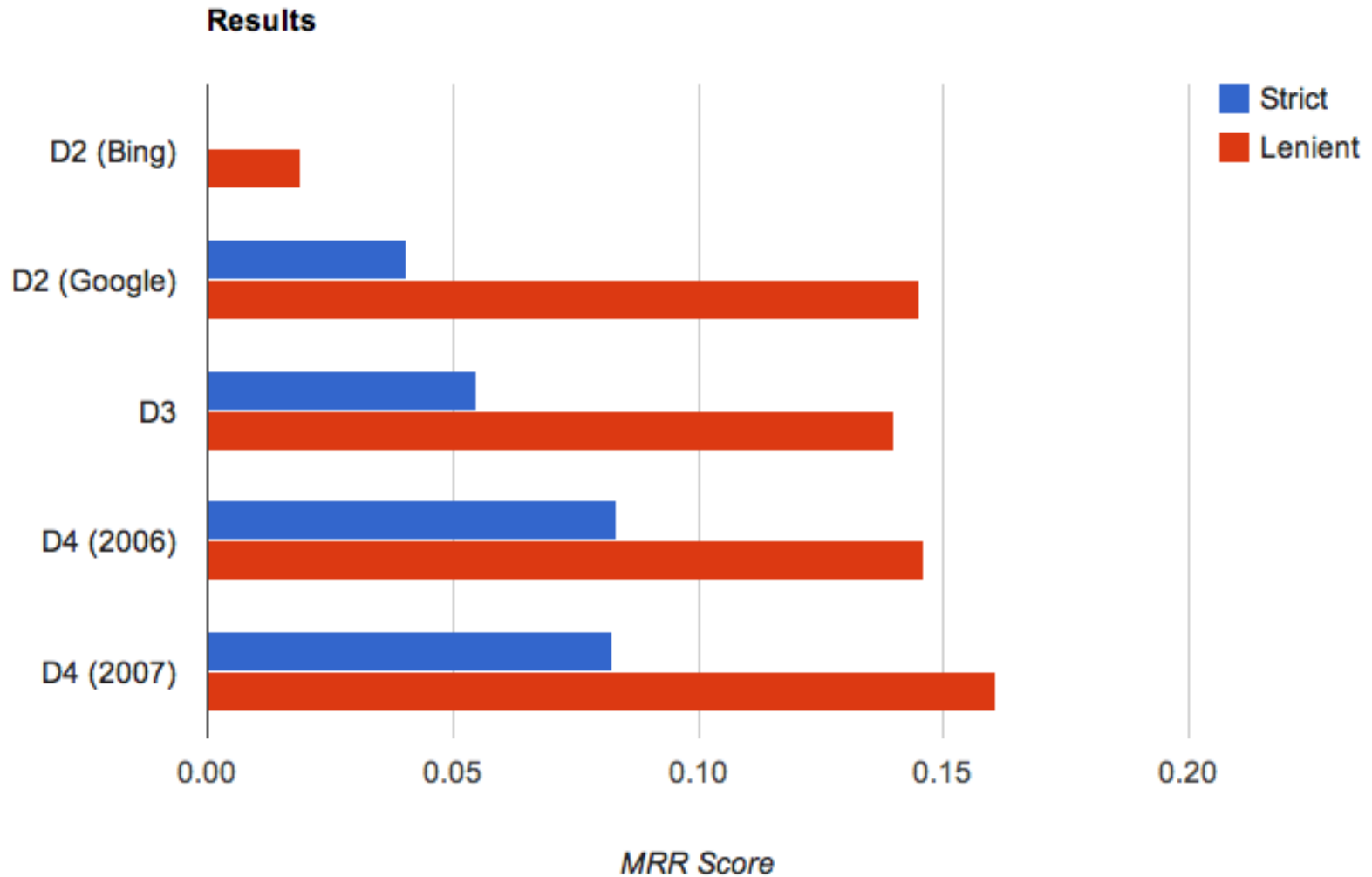
2006:

Char limit	Strict	Lenient
100	0.0833	0.1458
250	0.0833	0.1458

2007:

Char limit	Strict	Lenient
100	0.0826	0.1605
250	0.0826	0.1605

Results



Successes

(TREC 2006)

Queries and Lucene

- Retrieval:
 - 290 questions (72%) had GS answer in top 20 docs
- Ranking:
 - MRR for entire document:
 - strict: 0.4115
 - lenient: 0.5868

Issues and areas for improvement

Answer extraction and scoring:

- GS answer appeared in top 20 system answers for only 93 questions
 - 32% of 290
 - 23% overall
- MRR for web search alone (Google and Bing):
 - lenient: 0.2063

References

- Sanda Harabagiu, Dan Moldovan, Christine Clark, Mitchell Bowden, Andrew Hickl and Patrick Wang. 2005. Employing Two Question Answering Systems in TREC-2005. *Proceedings of the Fourteenth Text REtrievalConference*.
- Xin Li and Dan Roth. 2005. Learning Question Classifiers: The Role of Semantic Information. *Natural Language Engineering*.
- Jimmy Lin. 2007. An Exploration of the Principles Underlying Redundancy-Based Factoid Question-Answering. *ACM Transactions on Information Systems*, Vol. 25, No. 2, Article 6.