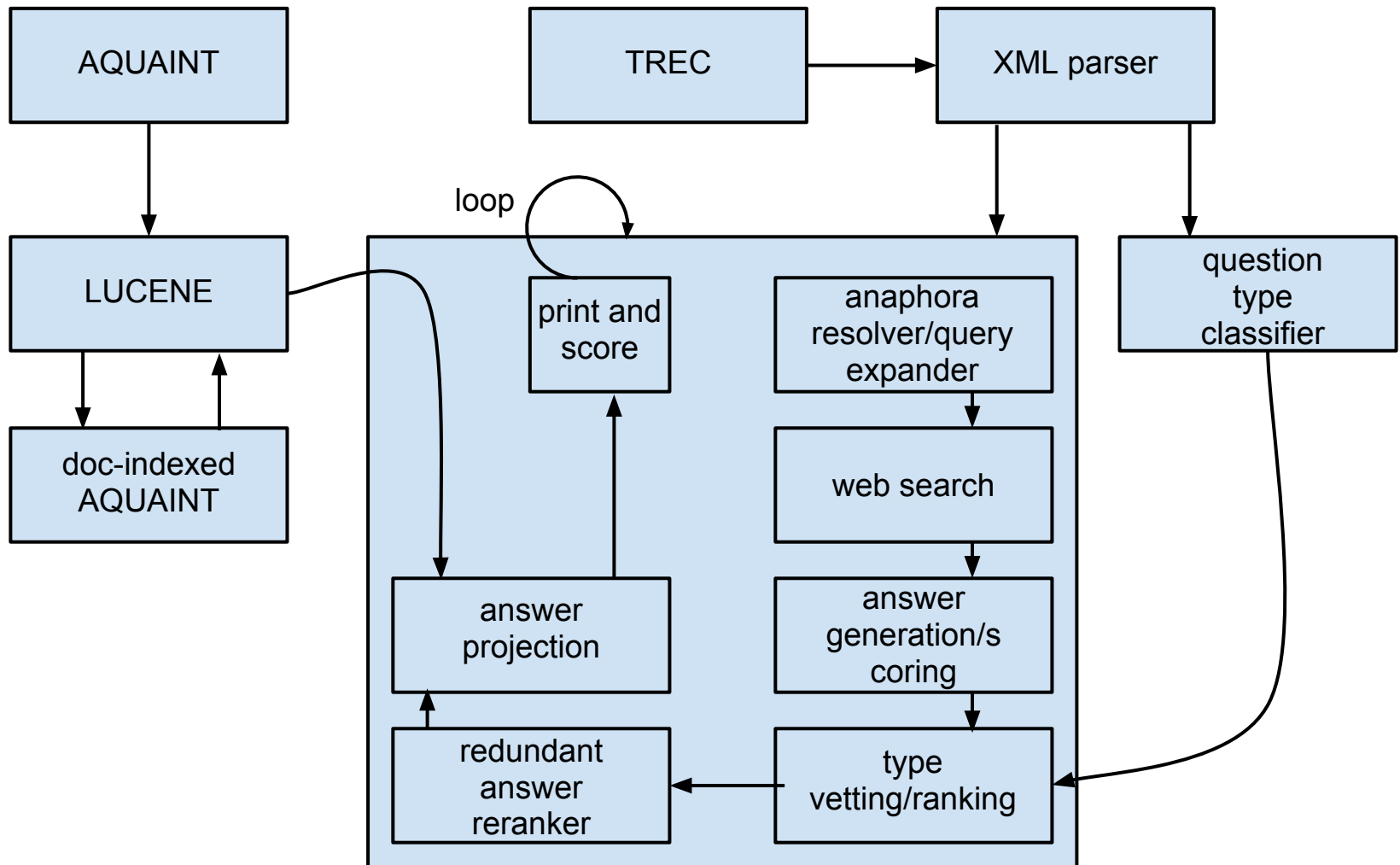

Deliverable 4

Stefan Behr, Tristan Bodding-
Long, Nick Waltner

System Overview



Results (No char-length difference)

Metric	2006	2007
Lenient	0.2559	0.2313
Strict	0.1256	0.0890
L. Accuracy	18.86%	15.86%
S. Accuracy	9.30%	5.17%

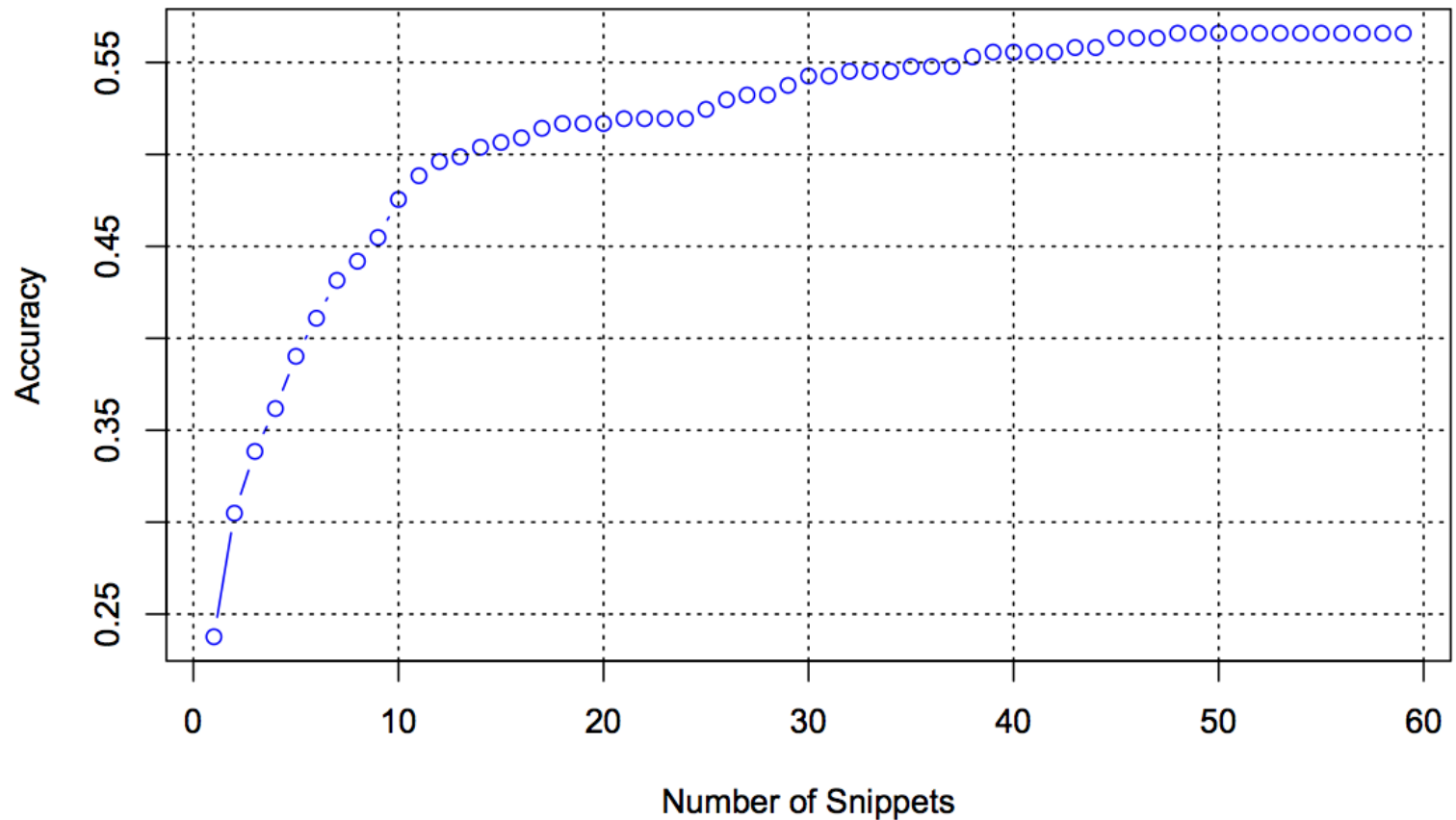
Answer Formulation

- After removing 0-val bookends

N	Lenient	Strict	L Accuracy	S Accuracy
1	0.1215	0.0720	0.0826873385	0.0516795866
2	0.1713	0.0862	0.1136950904	0.0568475452
3	0.1989	0.0879	0.1240310078	0.0568475452
4	0.2333	0.1177	0.165374677	0.0878552972
5	0.2559	0.1256	0.188630491	0.0930232558
6	0.2554	0.1204	0.180878553	0.0826873385
7	0.2538	0.1249	0.180878553	0.0904392765
8	0.2645	0.1231	0.1912144703	0.0878552972
9	0.2667	0.1155	0.1937984496	0.0801033592
10	0.2550	0.1212	0.180878553	0.0878552972

Evaluating Bing & Queries

Accuracy Answering Every Web Return



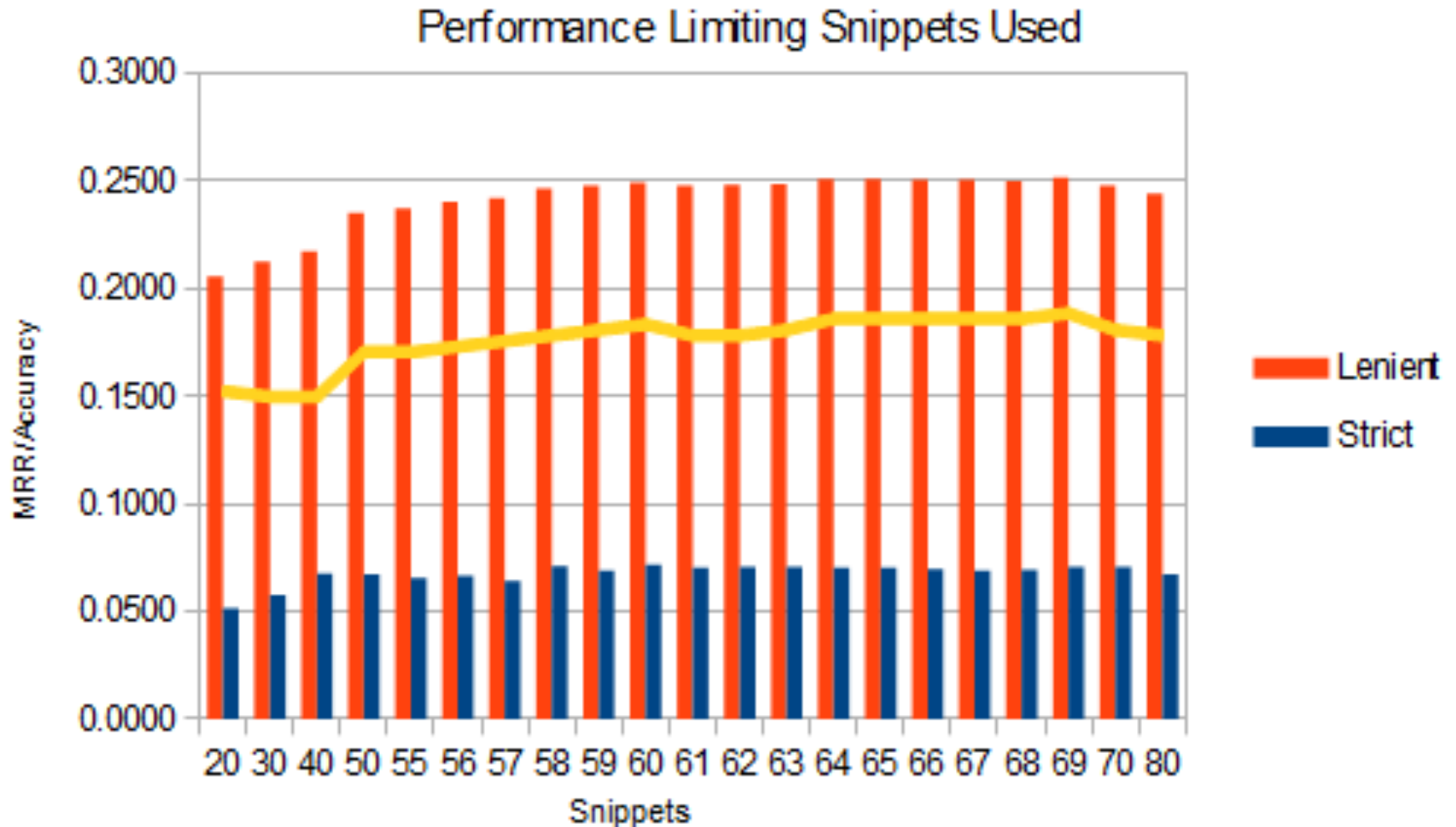
Queries & Snippets

- Maximum with perfect answer ranking: 65.37%
 - Average Snippets per Question: 90.2
 - 15% of correct answers we retrieved occurred for the *first time* in the 2nd half of answers
 - Redundancy approach has almost no chance at getting these answers
 - Including the 11th snippet/question adds only 5 correct new answers
 - No inclusion after the 12th snippet adds more than 2 correct new answers to the pool
-

Possible Solutions

- 'Better' Queries
 - Queries limited by the web's dynamicism
 - Question series information needs deep processing
 - Better retrieval
 - Non-Redundant approaches
 - Deep Processing Base-Corpus
 - Keyphrase / Named Entity Extraction across document collection
 - Algorithm driven constant setting
 - Resolve constants using classification
 - Limit Confounding Returns
 - Ensure correct answers, when found, are not confused by bad back-end returns
-

Decreasing Snippet Noise



Answer Re-ranking - II

Implemented R, Hovy & Och paper using SVM^{rank}

Used their four feature vector approach:

- Word frequency: Correct answer appears often. Use log of sum.
- Correct category: Build ME classifier using snippets and category guess. 0/1 variable. 67% test accuracy.
- Q-Word presence: Question words often appear near the answer. 0/1 variable.
- Overlap. Answers words overlap with question. 0-1 variable.

Lenient score dropped to 0.15, while strict was roughly the same.

Further, model tweaking could help.

Answer Projection

- D3 System
 - Boosted Answer + Bag of Topic
 - Bag of Answer + Topic
 - D4 System
 - Boolean Answer
 - Bag of Answer + Query

 - Roughly 40% boost in strict MRR
-