MIM—QA SYSTEM

With advanced Big Data, Machine Learning, Cloud and other hot words!

ANSWER RE-RANKING

- We collected bigrams from the original question, target, and the top-10 web bigrams.

- We sort our relevant passages by the number of bigram matches.

- Quite significant boost in accuracy.

QUERY EXPANSION WITH VERB INFLECTION

- We use a dictionary of verb inflections in NodeBox (covers ~10,000 verbs)

- Dictionary is converted from the UPenn English tree adjoint grammar

- We add past tense or 3rd person singular present tense to the query
  - depending on the auxiliaries in the question

SMALL TWEAKS

- Remove duplicate passages

- Observe and take advantage of NER tagger's error pattern
  - e.g. place names commonly mistagged as PERSON
  - We also tried better tagger (one that actually queries DBPedia!) but gave up on it since it inflates the running time

- New question classifier with fewer classes
  - Combine classes that we don't really differentiate in answer ranking

ERROR ANALYSIS: STATISTICS

DEV TEST

- Total Questions: 403
- Total Questions with patterns: 386
- Total Answers with non-zero score: 263
- Total Answers in Top-10 position: 224
- Total Answers in 1st position: 117

EVAL TEST

- Total Questions: 307
- Total Questions with patterns: 290
- Total Answers with non-zero score: 197
- Total Answers in Top-10 position: 182
- Total Answers in 1st position: 91

ERROR ANALYSIS: SOURCES

- What is Krugman's Academic Specialty? (Economics)
  - Tonnes of answers with "Economist"
  - None with "Economics"
  - 'Deeper' analysis needed
- When was CAFTA signed? (May 2004)
  - Official signing took place on May 28 (Web Snippets)
  - Bush signed the CAFTA...in May (Our answer)
- When was Rush Limbaugh born? (1951)
  - Loads of Cached Web Snippets contain correct answer
  - Bigram scoring of Web Snippets cause of failure

Observations and Conclusions

- Lots of low hanging fruit with extremely shallow processing

- Pareto principle at play; n-grams are double-edged swords

Folks, we all did a reasonable job!

- Everyone with scores > 0.2 gave a correct answer in the Top 5, on average!

- We are in the flat region of the curve now, and progress is harder