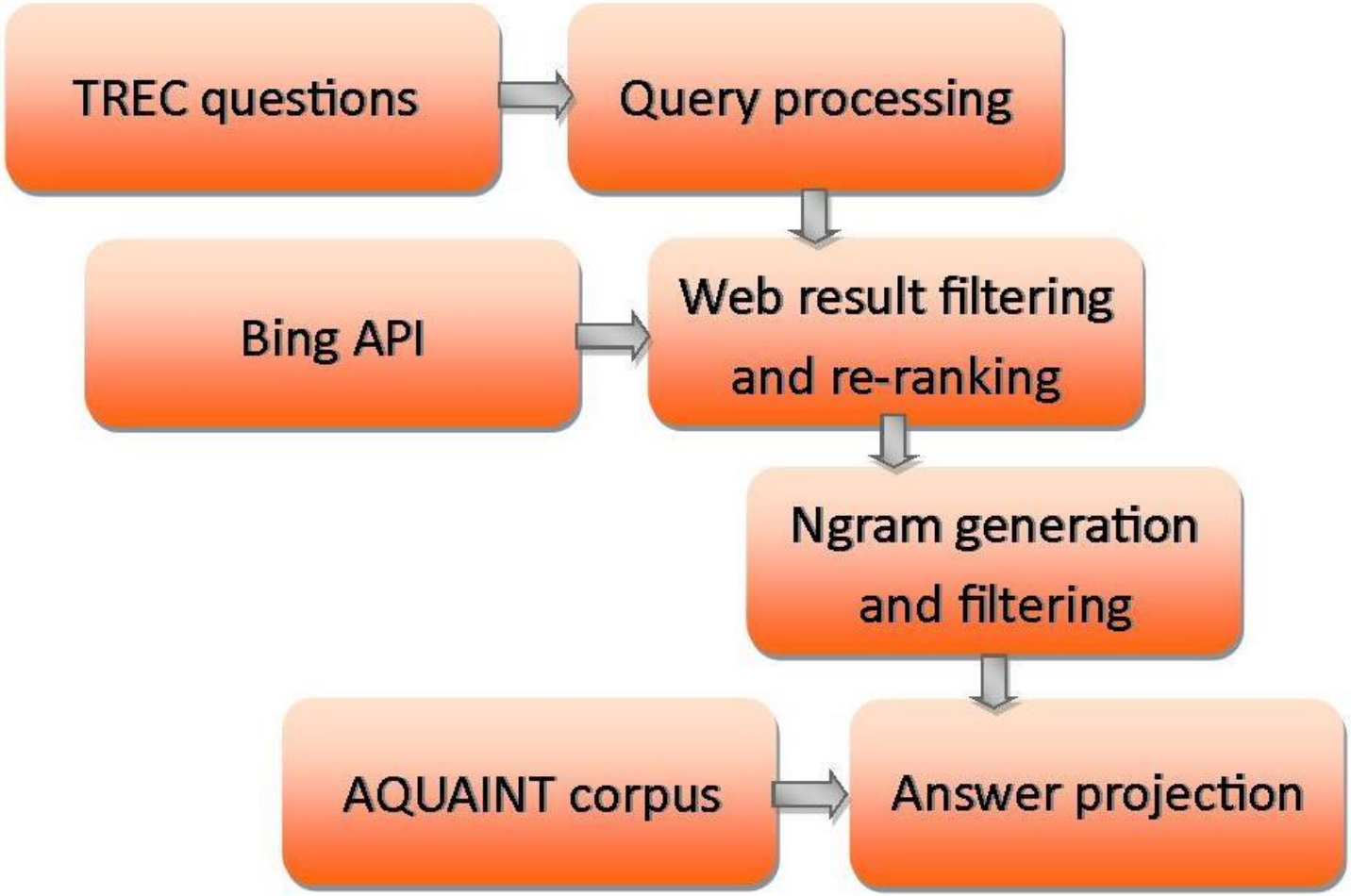
Improvements to our baseline question-answering system

Deliverable #3

Marie-Renee Arend ○ Josh Cason ○ Anthony Gentile

System overview



Changes to our approach over the baseline

- Prepared to treat the QA process as a classification task
  - We plan to use the MaxEnt algorithm to learn the features & weights
  - Did not have time to incorporate this for D3
- Tested the effect of query word sense disambiguation
- Added some preliminary answer pattern detection

Query processing module

- **Baseline module**: performed only the most basic processing tasks, such as stemming, removing stopwords & adding the target if not present in the question
- **Changes for D3**: Attempted to use our own version of the modified Lesk algorithm as described by Huang et al. (2008)
  - We chose to disambiguate every word in the query that could be found in WordNet, as opposed to just the headword
  - Intended to pull out both synonyms (for query expansion) & hypernyms (for use once we implement our machine learning approach)
- **Issues**:
  - WSD & synonym-generating code did not work as intended
    - Ended up with a lot of noise in the query due to the module adding a large number of relevant and irrelevant synonyms to the reformed query
    - Did not answer many questions that we successfully answered with our baseline system, so we ended up removing these changes for D3
  - Time – at the moment, the disambiguation code takes too long to process each query to be useful

Planned improvements to the query processing module

- **Improvements to our current code**:
  - Increase efficiency & accuracy of our modified Lesk algorithm
  - Reduce the # of possible synonyms to be added to the reformed query to cut down on noise
  - Try disambiguating just the headword of the query instead of all of the words in the query that could be found in WordNet
- **Other things to try for D4**:
  - Pivoting – try using translation via a pivot language for WSD of the headword
  - Possibly other WSD algorithms

Web result filtering & re-ranking

- This was not part of our baseline system
- **Changes for D3**:
  - To give priority to results that most likely have the answer we're looking for, we applied a weight to the results based on their initial web search ranking as retrieved from Bing
  - Searched for answer patterns based on the query, & if the pattern we're searching for appears, give that result another boost that will be used later in the pipeline
    - Three different answer pattern categories: temporal, location & numerical
- **Issues**:
  - SSL certification errors with requests library. Added some redundancy to ensure we retry if a web search fails.

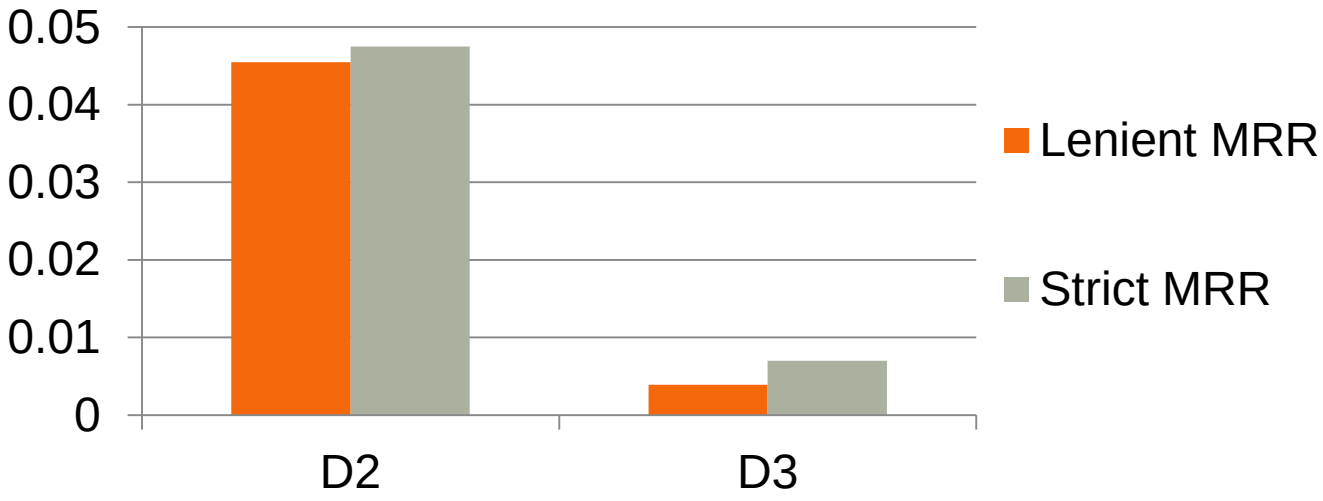
Planned improvements

- We want to use the answer pattern information as features once we incorporate our MaxEnt classifier
- Improve answer extraction methods [Ravichandran and Hovy, 2002]

Ngram generation & filtering

- **Baseline**: Generated ngrams from the web results and filtered them
  - 100 results chosen from results returned by Bing API
  - These results are tokenized, and from their title & snippet, we generated unigrams, bigrams, trigrams & quadrigrams
  - Ngrams removed from ranking based on specific criteria, and then re-ranked according to the formula in Lin (2007)
- **Changes for D3**:
  - Right now, since we don't have the classifier in place, this is the module in the pipeline where we are utilizing the weight boost given if the result matches the desired query answer pattern

Results



	D2	D3
Lenient MRR	0.0455	0.0475
Strict MRR	0.0039	0.0070

Useful readings & resources

- Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media.
- Graff, D. (Ed.). (2002). The AQUAINT corpus of English news text. Linguistic Data Consortium.
- Li, X., & Roth, D. (2006). Learning question classifiers: the role of semantic information. Natural Language Engineering, 12(3), 229-250.
- Lin, J. (2007). An exploration of the principles underlying redundancy-based factoid question answering. ACM Transactions on Information Systems (TOIS),25(2), 6.
- Hatcher, E., Gospodnetic, O., & McCandless, M. (2004). Lucene in action.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database*. International journal of lexicography, 3(4), 235-244.
- Resnik, Philip. (1995). Disambiguating Noun Groupings with Respect to WordNet Senses. Third Workshop on Very Large Corpora. Retrieved from http://acl.ldc.upenn.edu/W/W95/W95-0105.pdf

Questions?

Thanks for listening!