LING 573
Deliverable 3

RYAN ALDRICH | MATTHEW BENAK | NORAH HOGOBOOM

Approach

Focus on improving question processing.

- We first tried a few elementary answer processing changes from our D2 version. We originally pulled 250 chars after hit. We tried extracting a balanced number of characters after the hit.

- We next implemented a new question processor which included <Target> in the query and removed wh- and a small set of stop words.

- Next we used the Porter Stemmer although that has a mixed history.

- We made several bug fixes in several modules.

- Finally we tried query expansion by adding hypernyms in several configurations from WordNet.

Question Processing

-Based on "Question Classification with SVM and Error Correcting Codes" - Hacioglu, Ward

-Extend SVM to multi-class using error correcting codes

-Ex: DESC:def =   0 0 1 0 1 1 0 1 1 1 0 1 1 0 0 1 …

DESC:manner = 1 1 0 1 0 1 1 0 0 0 0 1 0 1 1 1 …

-Abysmal results compared to paper (14.2% vs 79.4%). Only major difference is feature lists wasn't trimmed.

Strict	Lenient
0.000000	**0.210000**
0.094605	0.146483

Query Expansion with WordNet

The first option used all hypernyms from the first set of returned hypernyms.

The second one added only the first of the first set of returned hypernyms.

"position play professional football"

1. *play field_game*
    dramatic_composition point
    dramatic_work football grownup
    adult professional position
    contact_sport

2. *field_game dramatic_composition*
    adult contact_sport point

Strict	Lenient
0.000000	**0.210000**
0.111496	0.176461
0.110532	0.168692

All Results

Run	Strict	Lenient
D2 (baseline)	0.00000	0.21000
run_id bugfix	0.12490	0.19281
Extract 125 chars before and after hit	0.09269	0.14976
New question processing, including <Target>	0.09461	0.14648
Case-sensitive extraction	0.13756	0.19477
Porter stemmer	0.13969	0.19639
Cleaned up indices (no changes)	0.13969	0.19639
Case-sensitive question processing (no changes)	0.13969	0.19639
Added hypernymns to query string from WordNet	0.11150	0.17646
Added only one hypernym per term from WordNet	0.11053	0.16869

Next Steps

Query Formulation
- We can continue to do refinements in query processing as we focused mostly on query expansion techniques.
  - Reformulate using web results

Document Retrieval
- Our initial document retrieval was difficult to improve upon after some obvious bug fixes.
  - Web boosting

Answer Processing
- We have the most room to improve in Answer Processing
  - Take advantage of question classification improvements already in place.
  - Explore Named Entity Recognition.
  - Different weighting techniques.