Deliverable #4: The Whole Enchilada

Thomas Phan
Jason Shaw
Varden Wang
Lucene

- Followed standard Lucene practices and examples
- Index of TREC 2005 documents was 3.3 GB
  - Stored entire document body in index
- Added Porter stemmer (PorterStemFilter class)
- Removed duplicate words from query
- Experimented with different stopword lists
  - Range in size from 33-600
  - Settled on NLTK’s 127-word list
MultiText

- Implemented from scratch
- Based on the description in Clarke et al. (TREC 9 submission)
- True IDF measure used
- Find “covers” -- passages bounded by query terms
  - Length of “cover” not accounted for
  - Only look in top 100 documents
  - Stopping performed by Lucene
    - Law of unintended consequences
Reranking

- Classify questions and answers
- Compare classifications to determine relevance
  - Cosine similarity
- Scale MultiText metric accordingly
- Rerank all 100 passages, return top 20
Query Classification

- Unigrams
- POS tags
- WordNet senses list

- Accuracy on TREC 10: 88.0%
- Accuracy on TREC-2005: 83.5%
Answer Classification

- Automatically constructed training data
  - Labels: query classifications
  - Instances: correct answers from corpus
  - Features
    - Unigrams
    - WordNet senses
- Accuracy

<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>250</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coarse</td>
<td>59.4%</td>
<td>60.8%</td>
<td>58.4%</td>
</tr>
<tr>
<td>Fine</td>
<td>43.6%</td>
<td>43.1%</td>
<td>39.3%</td>
</tr>
</tbody>
</table>
Results

- Document Retrieval
  - Stopping lowers MAP (.301 to 0.291) but increases precision (P₅: 0.176 to 0.185)

- Passage Retrieval (Lenient MRR)

<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>250</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliverable 3</td>
<td>0.111</td>
<td>0.201</td>
<td>0.380</td>
</tr>
<tr>
<td>Stopping</td>
<td>0.072</td>
<td>0.142</td>
<td>0.296</td>
</tr>
<tr>
<td>Reranking</td>
<td>0.081</td>
<td>0.142</td>
<td>0.297</td>
</tr>
</tbody>
</table>

2004 reranking boost: 0.065 to 0.083
Lessons, questions, future directions

- Different tasks require different techniques
  - Document and passage retrieval need not share query processing methods

- Why did reranking only help on short passages?

- Improved answer classification
  - NER
  - POS?

- Dependency parsing for reranking