Terabyte Track
Donald Metzler, Trevor Strohman, Yun Zhou, W.B. Croft

Efficiency Task
- Machine configuration:
  - CPU: Intel Xeon 2.4 Ghz (2)
  - Memory: 30B
  - Two processes
- single machine
- distributed across 6 machines
  - Scaling speed roughly 1.5Gbits/sec
  - End-to-end query processing time:
    Avg. query time: 1.8s (single)
    Avg. query time: 0.8s (distributed)
  - Threaded query processing:
    Avg. query time: 1.7s (single)
    Avg. query time: 0.2s (distributed)

<table>
<thead>
<tr>
<th>run id</th>
<th>Setup</th>
<th>Index Time (sec)</th>
<th>Total Q (avg)</th>
<th>#P</th>
<th>P/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ind48Stq</td>
<td>Single</td>
<td>2079</td>
<td>71.90</td>
<td>1.43</td>
<td>0.5320 0.5490</td>
</tr>
</tbody>
</table>

Ad Hoc Task

Dependence Model
- Term proximity important for large collections
- Helps reduce noisy single term matches
- Dependence model combines various features
- Three types of features:
  - single term
  - related window
  - context window

Pseudo-Relevance Feedback
- Rank original query (Qnum) and construct relevance model from top documents
- Construct expanded query (Qsys), using top expansion terms from relevance model
- Index query feedback:
  - RankPF(0.0 - 1.0) - rankPF

Results

Robust Track
Donald Metzler, Fernando Diaz, Trevor Strohman, W.B. Croft

Mixture of Relevance Models
- Combining evidence from multiple sources proved effective in past Robust Tracks
- Rather than using external web knowledge, we use a large collection of newswire documents called BIGNEWS
- Create relevance model for each collection and mix them together
- Expand query using most likely terms from model
  \[ P(w | Q) = \sum_{r} \frac{P(c)}{R} \sum_{i=1}^{M} P(w | D) P(Q | D) \]

Results

<table>
<thead>
<tr>
<th>run id</th>
<th>MAP</th>
<th>GMAP</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>ind48Stq</td>
<td>0.2159</td>
<td>0.1394</td>
<td>1.4290</td>
</tr>
<tr>
<td>ind48StqHtT</td>
<td>0.3029</td>
<td>0.1967</td>
<td>2.5777</td>
</tr>
<tr>
<td>ind48StqHt</td>
<td>0.3029</td>
<td>0.2001</td>
<td>2.6030</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>run id</th>
<th>MAP</th>
<th>GMAP</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>ind48StqHt</td>
<td>0.2159</td>
<td>0.1394</td>
<td>1.4290</td>
</tr>
<tr>
<td>ind48StqHtT</td>
<td>0.3029</td>
<td>0.1967</td>
<td>2.5777</td>
</tr>
<tr>
<td>ind48StqHt</td>
<td>0.3029</td>
<td>0.2001</td>
<td>2.6030</td>
</tr>
</tbody>
</table>

Title
- ind48StqHtT: dependence model
- ind48StqHtT: dependence model
- ind48StqHtT: dependence model

Description
- ind48StqHt: dependence model
- ind48StqHt: dependence model

MAP vs. GMAP

Q