ENGLISH RETRIEVAL FOR OTHER LANGUAGES

J. Dadashkarimi, R. Zhang, J. Kasai, W. T. Ting, C. Westerfield, H. Taggar, and D. Radev

1. Search English queries
2. Documents are indexed in other languages
3. The most common approaches are query translation and document translation
4. Query translation by Dictionary and document translation by Machine Translation

Problem

QT and Dictionary

Fig. 1 The proposed phrase-based query translation. \( t_{ij} \) is the \( j \)-th translation for \( q_i \). \( t_{12} \) and \( t_{22} \) are phrases in target language.

1. build all combinations and take the maximum
2. \( \text{max}(\#1(\#1(t_{11}, t_{22}), t_{32}), \#1(t_{12}, t_{22}), t_{31}) \)

DT and Proximity

Fig. 2 The proposed query hit proximity approach. A query hit should have \( q_1 \) and \( q_2 \) adjacent to each other. Based on how far are the query terms are located in each document \( f(q,d) \) gets lower score. This metric keeps the relative order of the query terms for matching.

1. Find different combinations of query terms’ positions
2. Find query hits that keep the order of the query
3. take the maximum possible hits
4. \# max\([1, 2, 23], [1, 2, 29], [15, 35, 71]\)
5. we can also have partial matching where a part of the query appears in the document
6. then we can combine \( p(q|d) \) based on language modeling:

\[
s(d) = \alpha \cdot p(q|d) + (1 - \alpha) \cdot f(q,d)
\]

\[
f(q,d) = -\log(\sum_{i=0}^{\#q} ||q_i, q_{i+1}||_d)
\]

Experimental Setup

1. We used Indri retrieval system
2. Europarl for CLEF collections and Bitext for surprise languages as training data for MT
3. We used hard cutoff of 20 for surprise languages and then ran a tuning from 1 to 20 to find a suitable cutoff for the collection
4. Wiktionary is used as a bilingual dictionary
5. Neural Machine Translation is used for document translation
6. Indri scores, word embedding, IDF, document length, and query length are features for soft cutoff
7. We used normalization and lower casing for all collections

Table 1. Collection Statistics

<table>
<thead>
<tr>
<th>ID</th>
<th>Lang.</th>
<th>Collection</th>
<th>Queries</th>
<th>#docs</th>
<th>#qrels</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>Spanish</td>
<td>EFE 1994</td>
<td>CLEF 2002, topics 91-140</td>
<td>215,738</td>
<td>2,854</td>
</tr>
<tr>
<td>de</td>
<td>German</td>
<td>Frankfurter Rundschau 94, Der Spiegel 94-95</td>
<td>CLEF 2002-03, topics 91-140</td>
<td>225,371</td>
<td>1,938</td>
</tr>
<tr>
<td>fr</td>
<td>French</td>
<td>Le Monde 94, SDA French 94-95</td>
<td>CLEF 2002-03, topics 251-350</td>
<td>129,806</td>
<td>3,524</td>
</tr>
<tr>
<td>sw</td>
<td>Swahili</td>
<td>Analysis</td>
<td>300 constrained queries</td>
<td>471</td>
<td>390</td>
</tr>
<tr>
<td>tl</td>
<td>Tagalog</td>
<td>Analysis</td>
<td>300 constrained queries</td>
<td>462</td>
<td>233</td>
</tr>
</tbody>
</table>

Experiments

1. Document translation gets consistently better results than Query translation
2. Phrase based retrieval gets as same result as synonym operator
3. Cutoff is very important for AQWV but not MAP
4. Wiktionary provides multiple choices for matching but MT provides only one option
5. System combination is a necessary step
6. We should consider variance around the best cutoff for each collection
7. Dictionary coverage is very important since a lot of queries have not translations

Table 2. Experimental Results

<table>
<thead>
<tr>
<th>Lang</th>
<th>ANALYSIS/DEV</th>
<th>QT/DT</th>
<th>Phrase</th>
<th>Morph</th>
<th>Reranking</th>
<th>MAP</th>
<th>P@10</th>
<th>AQWV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SW</td>
<td>ANALYSIS</td>
<td>QT</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>0.213</td>
<td>0.0805</td>
<td>0.1025</td>
</tr>
<tr>
<td>SW</td>
<td>ANALYSIS</td>
<td>DT</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>0.3162</td>
<td>0.1085</td>
<td>0.1877</td>
</tr>
<tr>
<td>SW</td>
<td>ANALYSIS</td>
<td>QT</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>0.212</td>
<td>0.0727</td>
<td>0.1313</td>
</tr>
<tr>
<td>SW</td>
<td>ANALYSIS</td>
<td>DT</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>0.3303</td>
<td>0.107</td>
<td>0.2606</td>
</tr>
<tr>
<td>TL</td>
<td>ANALYSIS</td>
<td>QT</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>0.2469</td>
<td>0.0656</td>
<td>0.202</td>
</tr>
<tr>
<td>TL</td>
<td>ANALYSIS</td>
<td>DT</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>0.5285</td>
<td>0.1286</td>
<td>0.3389</td>
</tr>
<tr>
<td>TL</td>
<td>ANALYSIS</td>
<td>QT</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>0.24</td>
<td>0.0641</td>
<td>0.2324</td>
</tr>
<tr>
<td>TL</td>
<td>ANALYSIS</td>
<td>DT</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>0.5621</td>
<td>0.1296</td>
<td>0.4282</td>
</tr>
</tbody>
</table>

Table 2. Experimental Results