Overview

University of Hyogo at NTCIR-11 TaskMine by Dependency Parsing

Takayuki Yumoto (Graduate School of Engineering, University of Hyogo, Japan)

1. Collection

sources = seed pages + detailed pages

(1) Collecting seed pages by query expansion

query string + “方法(method)”

Search Engine

seed pages

(2) Collecting detailing pages

conditions:
- same domain
- all query keywords in anchor text (and a title of the seed page)

detailed pages

2. Extraction

We use a dependency parsing (Cabocha)

(1) Finding operation (and negation)

マスク(mask)を / つけて(wear) / 出かける(go out)

1. the end of sentence is the first candidate

2. the declinable chunk depending to the other candidates is also a candidate

(2) Finding target and postpositional particle

the indeclinable chunk depending to operation

マスク(mask)を / つけて(wear) / 出かける(go out)

3. Ranking

We compared two ranking methods

(A) Site frequency of quadruplets : Order by

(マスク(mask), を, つける(wear), -)

(B) Site frequency of target and operation

Order by Min( , , ), Max ( , , )

(マスク(mask), を, つける(wear), -)

Synonyms by Wikipedia corpus

w1 and w2 are regarded as synonyms

Result

given : total 50 queries of 4 categories

<table>
<thead>
<tr>
<th></th>
<th>nDCG@1</th>
<th>nDCG@5</th>
<th>nDCG@10</th>
<th>nDCG@50</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>0.109</td>
<td>0.150</td>
<td>0.171</td>
<td>0.191</td>
</tr>
<tr>
<td>(B)</td>
<td>0.098</td>
<td>0.119</td>
<td>0.132</td>
<td>0.166</td>
</tr>
<tr>
<td>baseline</td>
<td>0.013</td>
<td>0.040</td>
<td>0.053</td>
<td>0.096</td>
</tr>
</tbody>
</table>

Problems

- the part where subtasks are extracted e.g. the page describing methods to prevent a cold and mechanism to catch a cold
- same subtask in different expressions e.g. “wear a mask” = “use a mask”
- limitation of model : multiple targets are sometimes needed in a single subtask