Introduction

Patent Prior Art Search: Finding previously granted patents or any published work relevant to new patent application.

Problem Definition: Generic IR techniques (e.g., different query reformulation) are ineffective for patent prior art search!

Previous Work: Mainly focused on different query expansion techniques to cope with significant term mismatch [1].


Oracular Term Selection

1. Relevance Feedback Score

Relevance feedback (RF) score for each term:

$$RF(t, Q) = Rel(t, Q) - Irr(t, Q) \quad (1)$$

where

$$Rel(t) \rightarrow \text{Avg. Term Frequency in Rel. Docs.}$$

$$Irr(t) \rightarrow \text{Avg. Term Frequency in Irr. Docs.}$$

2. Oracular Query Formulation

Formulate two oracular queries:

- Oracular Query = \{t ∈ top - 100|RF(t, Q) > τ\}
- Oracular Patent Query = \{t ∈ Q|RF(t, Q) > τ\}

Take Home Message

- Sufficiency of terms in baseline query
- Over-sensitivity of IR models to inclusion of negative terms (τ < 0)
- Need for precise methods to eliminate poor query terms (query reduction)

Query Reduction (QR): Approximating Oracular Query

1. Automated Reduction

QR Approaches:

1. Pruning document frequent (DF) terms ($DF(t) > τ$).
2. Pruning query infrequent terms ($QTF(t) \leq τ$).
3. Pseudo relevance feedback term selection ($PRF(t) > τ$).
4. Pruning IPC title general terms.

Take Home Message

- Automated QR methods fail to approximate oracular query.
- They cannot discriminate between positive and negative terms.

2. Semi-automated Interactive Reduction

Table 2: Performance of an Oracular Patent Query derived from only the top-k ranked relevant documents identified in the search results. We assume that the remaining documents in the top-100 are irrelevant.

<table>
<thead>
<tr>
<th>Baseline</th>
<th>PATATRAS</th>
<th>Oracular Patent Query (k=1)</th>
<th>Oracular Patent Query (k=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP</td>
<td>0.112</td>
<td>0.226</td>
<td>0.289</td>
</tr>
<tr>
<td>Recall</td>
<td>0.416</td>
<td>0.467</td>
<td>0.484</td>
</tr>
</tbody>
</table>

- MAP doubles over the baseline (0.112 → 0.289)
- Outperforms PATATRAS (0.226 → 0.289)

Take Home Message

- Interactive methods offer a promising avenue for simple but effective term selection in prior art search.

Acknowledgement

I would like to thank SIGIR for offering me a travel grant to attend the 2015 conference in Chile.

References


