City of Disguise: A Query Obfuscation Game on the ClueWeb

Search Engines—A Thread to Your Privacy?

There are two approaches for automatic query obfuscation.

**Semantic Query Obfuscation**
Replace terms in the sensitive query with terms representing more general concepts (E.g., with WordNet).

**Statistical Query Obfuscation**
Use a private private search engine to identify non-revealing queries that retrieve similar results to the sensitive query.

[Arampatzis et al., IRJ'15]

Can Humans Obfuscate Queries?

Overview of the Game

City of Disguise is a retrieval game that tests how well searchers can reformulate some sensitive query in a ‘Taboo’-style setup but still retrieve relevant results.

Details:
- Corpus of 0.6 million ClueWeb12 documents
- Documents rendered with the Wayback Machine
- 200 sensitive queries with one relevant document

Scoring Query Obfuscations:
- Position of the relevant document
- Length of the Query
- Recall (top-100 results of sensitive query)
- MAP (top-100 results of sensitive query)

Evaluation

Effectiveness of obfuscated queries in ChatNoir and the games’ document sample (‘Sample’) as MRR, the number of documents retrieved for the original query (‘Ori.’), and the number of retrieved relevant documents (‘Rel.’). We compare automatically obfuscated queries and four types of queries submitted by players.

<table>
<thead>
<tr>
<th>Obfuscated Queries</th>
<th>ChatNoir</th>
<th>Sample</th>
<th>Sensitive Web Track Queries</th>
<th>ChatNoir</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Length</td>
<td>Time</td>
<td>MRR</td>
<td>Ori.</td>
<td>MRR</td>
</tr>
<tr>
<td>Only Suggestions</td>
<td>130/21</td>
<td>2.42</td>
<td>40.50 s</td>
<td>0.093</td>
<td>5.223</td>
</tr>
<tr>
<td>Some Suggestions</td>
<td>556/125</td>
<td>4.53</td>
<td>42.45 s</td>
<td>0.046</td>
<td>4.667</td>
</tr>
<tr>
<td>No Suggestions</td>
<td>576/157</td>
<td>2.88</td>
<td>44.39 s</td>
<td>0.029</td>
<td>2.935</td>
</tr>
<tr>
<td>New Word</td>
<td>559/158</td>
<td>3.57</td>
<td>46.27 s</td>
<td>0.002</td>
<td>1.517</td>
</tr>
<tr>
<td>Automatic</td>
<td>1025/327</td>
<td>2.91</td>
<td>—</td>
<td>0.088</td>
<td>9.229</td>
</tr>
</tbody>
</table>

Can You Beat the Highscore?

demo.webis.de/city-of-disguise/
github.com/webis-de/ECIR-22