Mining Test Oracles of Web Search Engines

Wujie Zheng¹, Hao Ma², Michael Lyu¹, Tao Xie³, and Irwin King¹,⁴

¹CUHK, ²Microsoft Research, ³NCSU, ⁴AT&T Labs

Research Problem

• How to check relevance of Web search results?

Related Work

• Multiple implementation testing: search engines often disagree with each other
• Mining specifications: often in implementation level
• Challenge: how to mine reliable test oracles from multiple search engines’ information

Our Approach

• Mapping queries and search results to itemset database
• Mining rules from existing queries and search results
• Detecting violations of rules for new queries and search results

Evaluation Setup

• Google and Bing
• 3432 queries
• 390000 ranked lists of search results in 4 months (top 10 results for each query in each day)

Example Mined Rules

1. top10:quotes.nasdaq.com, => top10:finance.yahoo.com, = 314/314 = 1.0
2. top10:finapps.forbes.com, => top10:finance.yahoo.com, = 262/262 = 1.0
3. top10:absoluteastronomy.com, => SE:bing, = 7657/7657 = 1.0
4. Q:facebook, => top1:facebook.com, = 182/182 = 1.0

• Rules 1-2: implications between Websites
• Rule 3: different opinions of search engines to certain Website
• Rule 4: the best top 1 search results

Example Violations

• April 1st, 2011, Bing violated the following rule:
  Q: where to login to john carroll university email, => top1:mirapoint.jcu.edu, = 172/180 = 0.96
• The top 1 search result of Bing was http://www.jcu.edu/index.php, from which one cannot find email entrance easily

Our New Idea

• Mining input/output specifications across multiple search engines
• Key point: a set of common input/output properties across different search engines in different time

<table>
<thead>
<tr>
<th>Category</th>
<th>Item Description</th>
<th>(Example) Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query</td>
<td>The query.</td>
<td>Q:se 2011</td>
</tr>
<tr>
<td>Query</td>
<td>A word in the query.</td>
<td>Q:Case</td>
</tr>
<tr>
<td>Query</td>
<td>The query type (not, common).</td>
<td>NotQ</td>
</tr>
<tr>
<td>Query</td>
<td>The number of words in the query.</td>
<td>OneWord</td>
</tr>
<tr>
<td>Query</td>
<td>The number of words in the query.</td>
<td>TwoWords</td>
</tr>
<tr>
<td>Search Result</td>
<td>The domain of the top 1 search result</td>
<td>top1:continuinged.ku.edu</td>
</tr>
<tr>
<td>Search Result</td>
<td>The domain of a top 10 search result</td>
<td>top10:continuinged.ku.edu</td>
</tr>
<tr>
<td>Search Result</td>
<td>The Alexa Rankings of the top 10 results’ top private domains are all greater than 1,000.</td>
<td>ALLGE1K</td>
</tr>
<tr>
<td>Match</td>
<td>The whole query does not appear in the title of any top 10 result.</td>
<td>NoTitleHasQ</td>
</tr>
<tr>
<td>Search Engine</td>
<td>The search engine that returns the search results</td>
<td>SE:google</td>
</tr>
</tbody>
</table>