Exposing Inconsistent Search Results with Bobble

Nick Feamster
Georgia Tech

Wenke Lee, Xinyu Xing, Bilal Anwer, Dan Doozan
Georgia Tech

Alex Snoeren
UCSD
Motivation

- Search engines deliver inconsistent search results
- These inconsistent results may sway searchers’ opinions or judgment to products or political events etc.
Rex Agreement: Avoid This Like The Plague — Penguin Homes Blog
real estaleblog.wordpress.com/.../rex-agreement-avoid-this-like-the-pl...

Mar 17, 2008 — The Rex Agreement has to be one of the most abusive instruments I can imagine. It’s just like taking your home equity to the pawn shop.

Rex Agreement

FirstREX enables homeowners to convert the equity in their home into cash, with no debt, no interest, and no monthly payments — ever. A Rex Agreement is an...

Rex Agreement: A Great New Way to Reduce Debt
real estaleblog.wordpress.com/.../rex-agreement-packed-less...

The Rex Agreement is a real deal of any kind. It is not a real estate investment agreement that enables you to convert a portion of your home equity into immediate ....

Rex Agreement

The Rex Agreement is a way for homeowners to get cash from their home without monthly payments or interest.

Searches related to rex agreement
rex agreement reviews
rex agreement
Editorial: No on Prop. 39 - SFGate
Sep 21, 2012 – the Legislature recently came within a whisker of eliminating the loophole and devoting the new revenue to scholarships for middle-class students...
Goal: Understand the Nature of Inconsistencies

• Browser plugin, Bobble (http://bobble.gtisc.gatech.edu/)
  – allows users to see how the search results that Google returns to them differs from the results that would be returned to other users distributed around the world
  – record the user’s search query and repeating it from a variety of different vantage points

• Study how users’ Google search results vary based on their geographic locations and past search histories
  – 75,000 queries
  – 175 users
  – Nine months
Bobble Architecture
Requirements for Data Collection

- Effects of **personalization**
  - personalized and non-personalized search results of Google users

- Effects of **geography**
  - non-personalized search results from different regions
Challenges for Data Collection

- Non-intrusive data collection
- Measurement benchmark
Data Collection Platform

- A Chrome browser agent
- Browser agents on 308 PlanetLab nodes
Benchmark

• Use a 50Km-planetlab-node search result as a Google user’s non-personalized result
Benchmark Results

- Search results from planetlab node == search results from regular user’s machines
  - A proportion test shows no significant difference at p-value < .05
Statistics

- From 2012/1/17 – 2012/10/25 (9 months)
- 174 unique Google-user installation
- 100,451 queries
  - 13,974 queries issued by non-signed-in users
  - 86,477 queries issued by signed-in users
- 80,897 unique search terms
Geographic Distribution of Queries
Bobble Response Time

Number of PlanetLab Nodes per Task
\( \mu = 53.0, \sigma = 29.7321374946 \)
Query Categorization

Using dmoz.org query categorization
(How) Does Location Affect Search Results?

- Use dbscan algorithm to cluster PlanetLab nodes based on locations (cluster 1)
- Cluster Google search results based on the unique search result sets (cluster 2)
- Chi-square test:
  - ~95% of queries show high correlation in p-value (< 0.05)
Summary of Inconsistencies

• Not in user’s result set, but in Google top 3 elsewhere: 30.66%

• Not in user’s result set, but in Google top 10 elsewhere: 86.41%

• At least one result appears in Google’s result set but does not appear at other PlanetLab nodes: 1.88%
How Many Unique Sets of Results?
How Does Personalization Affect Results?

• For signed-in users
  – 33% of queries have at least one search result added as a result of personalization
  – 11% of queries have at least one search result removed

• For anonymous users:
  – 31% of queries have at least one search result added
  – 15% have at least one search result removed
Hoeffding Distance

- Way of characterizing inconsistencies across searches
  - Interpretable with respect to search algorithms retrieving ranked lists of different lengths
  - Models the increased attention users pay to top ranks over bottom ranks

- Zero: No difference between sets
  One: Completely different
Personalized Queries, Signed-in users

(a) Signed-in data set

<table>
<thead>
<tr>
<th>Category</th>
<th># of queries</th>
<th># of customized queries</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts</td>
<td>12,302</td>
<td>7,859</td>
<td>0.0829</td>
<td>0.0902</td>
</tr>
<tr>
<td>Business</td>
<td>7,284</td>
<td>4,602</td>
<td>0.0834</td>
<td>0.0902</td>
</tr>
<tr>
<td>Computers</td>
<td>13,159</td>
<td>8,642</td>
<td>0.0859</td>
<td>0.0905</td>
</tr>
<tr>
<td>Games</td>
<td>3,318</td>
<td>2,132</td>
<td>0.0847</td>
<td>0.0907</td>
</tr>
<tr>
<td>Health</td>
<td>2,000</td>
<td>1,243</td>
<td>0.0859</td>
<td>0.0930</td>
</tr>
<tr>
<td>Home</td>
<td>1,470</td>
<td>933</td>
<td>0.0842</td>
<td>0.0888</td>
</tr>
<tr>
<td>Kids and Teens</td>
<td>5,588</td>
<td>3,583</td>
<td>0.0821</td>
<td>0.0896</td>
</tr>
<tr>
<td>News</td>
<td>348</td>
<td>211</td>
<td>0.0795</td>
<td>0.0820</td>
</tr>
<tr>
<td>Recreation</td>
<td>3,055</td>
<td>1,946</td>
<td>0.0890</td>
<td>0.0959</td>
</tr>
<tr>
<td>Reference</td>
<td>2,127</td>
<td>1,402</td>
<td>0.0896</td>
<td>0.0957</td>
</tr>
<tr>
<td>Regional</td>
<td>9,361</td>
<td>5,840</td>
<td>0.0825</td>
<td>0.0882</td>
</tr>
<tr>
<td>Science</td>
<td>4,170</td>
<td>2,621</td>
<td>0.0838</td>
<td>0.0923</td>
</tr>
<tr>
<td>Shopping</td>
<td>3,239</td>
<td>2,052</td>
<td>0.0868</td>
<td>0.0937</td>
</tr>
<tr>
<td>Society</td>
<td>5,758</td>
<td>3,619</td>
<td>0.0836</td>
<td>0.0904</td>
</tr>
<tr>
<td>Sports</td>
<td>2,761</td>
<td>1,759</td>
<td>0.0817</td>
<td>0.0895</td>
</tr>
<tr>
<td>World</td>
<td>12,041</td>
<td>7,850</td>
<td>0.0869</td>
<td>0.0922</td>
</tr>
</tbody>
</table>
Other Applications: News

- News Agencies:
  - Reuters
  - ABC News
  - Aljazeera
  - CNN
  - Agence France-Presse
  - Agência Brasil
  - American Press Association
  - ANP (Netherlands)
  - Associated Press
  - ....

Internet

- AJC
- LA Times
- NYTimes
Data Collection
Lack of Sources in RSS Feeds

- 80-20 principle for English language edition countries.
- For many countries its 90% of articles from 10% of news sources.
- Same holds for Spanish, French and Arabic.
## Local BIAS (RSS Feeds)

<table>
<thead>
<tr>
<th>Country</th>
<th>Sources</th>
<th>Articles</th>
<th>Top Two Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>1,529</td>
<td>54,943</td>
<td>washingtonpost.com, reuters.com</td>
</tr>
<tr>
<td>UK</td>
<td>604</td>
<td>35,848</td>
<td>telegraph.co.uk, bbc.co.uk</td>
</tr>
<tr>
<td>Singapore</td>
<td>993</td>
<td>63,134</td>
<td>straitstimes.com, washingtonpost.com</td>
</tr>
<tr>
<td>Brazil</td>
<td>560</td>
<td>37,549</td>
<td>terra.com.br, dgabc.com.br</td>
</tr>
<tr>
<td>Israel</td>
<td>1,495</td>
<td>103,157</td>
<td>jpost.com, washingtonpost.com</td>
</tr>
<tr>
<td>Australia</td>
<td>429</td>
<td>36,308</td>
<td>abc.net.au, smh.com.au</td>
</tr>
<tr>
<td>France</td>
<td>423</td>
<td>40,934</td>
<td>20minutes.fr, nouvelobs.com</td>
</tr>
<tr>
<td>Ireland</td>
<td>287</td>
<td>29,055</td>
<td>google.com, irish-times.com</td>
</tr>
<tr>
<td>Canada</td>
<td>373</td>
<td>38,504</td>
<td>vancouversun.com, montrealgazette.com</td>
</tr>
<tr>
<td>Germany</td>
<td>325</td>
<td>39,522</td>
<td>focus.de, welt.de</td>
</tr>
<tr>
<td>New Zealand</td>
<td>202</td>
<td>26,008</td>
<td>stuff.co.nz, tvnz.co.nz</td>
</tr>
<tr>
<td>Switzerland</td>
<td>226</td>
<td>30,676</td>
<td>romandie.com, nouvelobs.com</td>
</tr>
<tr>
<td>India</td>
<td>266</td>
<td>39,278</td>
<td>indiatimes.com, hindustantimes.com</td>
</tr>
</tbody>
</table>
Conclusion

• Search inconsistency (and information manipulation) is pervasive
  – Geographic location introduces inconsistency in about 98% of queries
  – Personalization results in addition or removal of results more than 30% of the time

• We have also done this analysis for news stories (similar geographic conclusions)

• Next steps
  – More detailed study of how personas affect results
  – Countermeasures