Volunteer?
What’s the population of Raleigh?
How many people live in Raleigh?
Paraphrasing Invariance Coefficient
-- Measuring Para-Query Invariance of Search Engines

Tomasz Imielinski, Jinyun Yan
Outline

- **Motivation**
  - Why do we want to measure the semanticity of search engines
- **Methods**
  - How to measure
  - What do we need to prepare
- **Future Work**
  - what should be solved
Search Engines are becoming semantic

- Google’s Eric Schmidt on the topic of the long term goals of Google search tells [Techcrunch](https://techcrunch.com): (2009-09-03)

  “Wouldn’t it be nice if Google understood the meaning of your phrase rather than just the words that are in that phrase? We have a lot of discoveries in that area that are going to roll out in the next little while.”

- Semantic Search: Search beyond words
Are they really semantic?

- **Semantic Search** is hot
  - Microsoft acquired semantic search engine PowerSet in 2008
  - Baidu published “box computing” in 2009
  - Facebook seeks to build the semantic search engine in 2010

---

**Facebook Seeks To Build The Semantic Search Engine**

 Posted by Nick O’Neill on April 21st, 2010 2:34 PM

It's an insanely ambitious project but Facebook has two things to their advantage: a user base of more than 400 million users and a suite of easy to implement social services which will help collect structured data from around the web. Overnight, Facebook has essentially launched the semantic search engine of the

Many search engine companies use the word “semantic”. Are they really semantic?
Today: Users have to ask nicely

- Search Engines are still quite sensitive to the way queries are formulated
- Semantically equivalent queries usually will get different results!
- Example:
  - 1. why do I get canker sores
  - 2. what causes canker sores
  - 3. why canker sores
  - 4. the reason of canker sores
Every URL is different. Such result is not semantic.
Why?

- **What does Semantic mean?**
  - Understand the meaning behind words
- **Should Semantic Search Engines understand the meaning of query?**
  - Yes
- **If they understand the meaning of query, are they supposed to recognize semantically equivalent queries?**
  - Yes
- **What should be the corresponding response when seeing semantically equivalent queries?**
  - Return the same answer!

⇒ It’s necessary for Semantic Search Engines to recognize paraphrases, although not sufficient!
Measure Metric

- **Semantic Invariance**
  - Be invariant to semantically equivalent queries

- **Equivalent queries => para-queries**
  - Alternative ways to express the same search desire
    - “Population of Raleigh” vs “How many people live in Raleigh”

- One feasible method to measure the semanticity
  - Return same results for para-queries.

- Necessary but not sufficient
  - Semantic $\Rightarrow$ Semantically invariant
  - Semantically invariant $\not\Rightarrow$ Semantic
Data Collection

• Objective tests need a large amount of para-queries
  – How to?

• Paraphrase generation
  – Para-query is a sub-class of paraphrase
  – Paraphrase: restatement of one phrase using different words or same words in different orders.
  – Borrow an idea from paraphrase generation
    • Lexical Variation: Numeric Synonym replacement (1623 pairs)
      – Top 10 songs ⇔ Top ten songs
      – Top 8 myspace ⇔ Top eight myspace

• Para-query has its own characteristics
  – short, few content words
Data Collection

• Extra Information
  – Tom Hanks ⇔ Tom Hanks actor
  – 8 categories extracted from wikipedia, 7198 pairs

• Query Reformulation
  – Similar to query reformulation: expand query using related words
  – Query reformulation doesn’t promise equivalent meaning
Collective Wisdom

- Rephraser Game
  - Duration: 15 minutes for each round
  - Round: 430 rounds
  - Input: a start query, a hidden phrase which is not visible to players
  - Goal: paraphrase the start query to gain score
  - Win policy:
    - If your phrase is proposed by others later: Votes
    - If your phrase matches the hidden phrase: Jack pot
  - Player: no limitation to the number of players
Rephraser Game
Rephraser Game
Amodal Perception

• In our game, every player independently plays the game
  – Can’t see others’ input
  – Can’t chat with other players

• Why such independence is important?
  – Some games with similar purposes provide assistant information
  – Amodal theory in psychology
    • Extra information will change the interpretation of the motion of objects
Visible Objects
Visible Objects
Occluded by a Pipe
Occluded by a Pipe
Generate para-Queries from the Game

• Votes
  - For each phrase proposed by players, it has a corresponding Votes score, which displays how many other players agreed on it
  - Filtering by Votes, we collect para-queries with high probability
    • Votes $\geq 3$ human acceptability ratio = 89%
      - Human acceptability ratio = $\frac{\#(agreed\ persons)}{\#(voting\ persons)}$
    • Stronger confidence
      - Absent with all choices (human acceptability ratio)
      - Not satisfy $p, 1-p$ distribution (Kappa)
      - $K$ players independently come up with a phrase $p$, in total $M$ phrases, the confidence of $p$?

• Templates
  - Select start queries to form templates
    • Who is the governor of [Alaska] now?
      - who is the governor of [X] now?
    - Find substitutions for argument slots
Measurement

Goal: Are search engines semantically invariant?

Experimental Set-Up

• Total of 19,518 pairs of (start query, para-query)
• Measured by paraphrasing invariant coefficient and entropy
• Tested on top 4 search engines:
Paraphrasing Invariant Coefficient

- **Associated Queries**
  - Queries are associated by Search Engine Return Page
  - Query q1 and q2 can have the same SERP, or not

- **Para-Queries**
  - Suppose q1 and q2 are para-queries, if they have the same returned top K URLs by search engine S, we assume:
    - Search engine S recognized that they are para-queries
  - Relax to consider the top result URL

- **Paraphrasing Invariant Coefficient**
  - Given a pair of para-queries, what’s the probability for a search engine S to return the same top result URL.

\[
PIC(S) = Pr(q_1 S q_2 | q_1 \approx q_2)
\]

  - It’s estimated from our test set:  \#(pairs with same top URL) / \#(pairs)
• quite low => high probability that you get different results for para-queries
• varies in different domains
Entropy

- **Uncertainty**
  - A standard in Information Theory to measure the level of uncertainty
  - Use it to measure how uncertain a search engine’s results are for a set of para-queries (q1...qn)

- **Calculation**
  - Given a set of para-queries (q1...qn), it has a corresponding top URL set (u1...um) returned by search engine S
  - If m = 1: S recognized that they are para-queries (best condition)
  - If m=n: every query got different result (worst condition)
  - If m<n: some queries are considered as para-queries
**Result**

<table>
<thead>
<tr>
<th>Templates</th>
<th>Google</th>
<th>Bing</th>
<th>Yahoo</th>
<th>Ask</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>what causes a <strong>Headache</strong></td>
<td>2.60</td>
<td>2.53</td>
<td>2.70</td>
<td>2.42</td>
<td>3.17</td>
</tr>
<tr>
<td>when was <strong>Mark Twain</strong> born</td>
<td>1.24</td>
<td>2.19</td>
<td>2.16</td>
<td>1.16</td>
<td>3.7</td>
</tr>
<tr>
<td>who wrote <strong>THE BIBLE</strong></td>
<td>1.84</td>
<td>1.53</td>
<td>2.10</td>
<td>2.14</td>
<td>3.17</td>
</tr>
<tr>
<td>where can I buy <strong>VIAGRA</strong></td>
<td>1.95</td>
<td>2.0</td>
<td>2.04</td>
<td>1.88</td>
<td>2.32</td>
</tr>
<tr>
<td>currency of <strong>RUSSIA</strong></td>
<td>1.69</td>
<td>1.67</td>
<td>1.97</td>
<td>1.79</td>
<td>2.32</td>
</tr>
<tr>
<td>how much do <strong>TEACHERS</strong> make a year</td>
<td>1.35</td>
<td>1.75</td>
<td>1.94</td>
<td>1.42</td>
<td>2.8</td>
</tr>
<tr>
<td>where is <strong>RUSSIA</strong> located</td>
<td>2.58</td>
<td>2.03</td>
<td>2.43</td>
<td>2.40</td>
<td>2.8</td>
</tr>
<tr>
<td>Is <strong>COPS</strong> on tonight?</td>
<td>1.55</td>
<td>1.4</td>
<td>1.58</td>
<td>1.83</td>
<td>2.6</td>
</tr>
<tr>
<td>fun things to do in <strong>NYC</strong></td>
<td>2.92</td>
<td>2.94</td>
<td>3.26</td>
<td>1.74</td>
<td>3.7</td>
</tr>
<tr>
<td>what are the health benefits of <strong>MILK</strong></td>
<td>1.95</td>
<td>2.40</td>
<td>2.60</td>
<td>2.06</td>
<td>3.17</td>
</tr>
<tr>
<td><strong>Daytona 500 pole winner</strong></td>
<td>1.74</td>
<td>1.82</td>
<td>2.13</td>
<td>1.71</td>
<td>3</td>
</tr>
<tr>
<td>What’s the oldest <strong>CITY</strong> in the <strong>U.S.A</strong></td>
<td>1.73</td>
<td>1.63</td>
<td>1.85</td>
<td>1.93</td>
<td>2.8</td>
</tr>
<tr>
<td>Who is the governor of <strong>Alaska</strong> now?</td>
<td>1.14</td>
<td>1.0</td>
<td>1.53</td>
<td>1.49</td>
<td>3.17</td>
</tr>
</tbody>
</table>

- The best value for entropy is 0
- The last column demonstrates the result for the worst condition
- Close to the worst condition, far away from best condition
Conclusion

• Semantic Search Engines need to understand the meaning of queries.
• If they could understand the meaning of queries, they should recognize para-queries.
• However, right now, they can’t.
Future Work

- How to recognize para-queries?
- What’s other measure metrics for semantic search engines?
  - Recognizing para-queries is necessary but not sufficient.
Thanks for your attention

- Questions?