

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

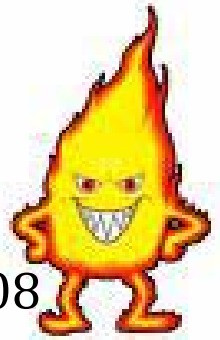
- Google's Eric Schmidt on the topic of the long term goals of Google search tells [Techcrunch](#): (2009-09-03)

“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



29

10 Comment:

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

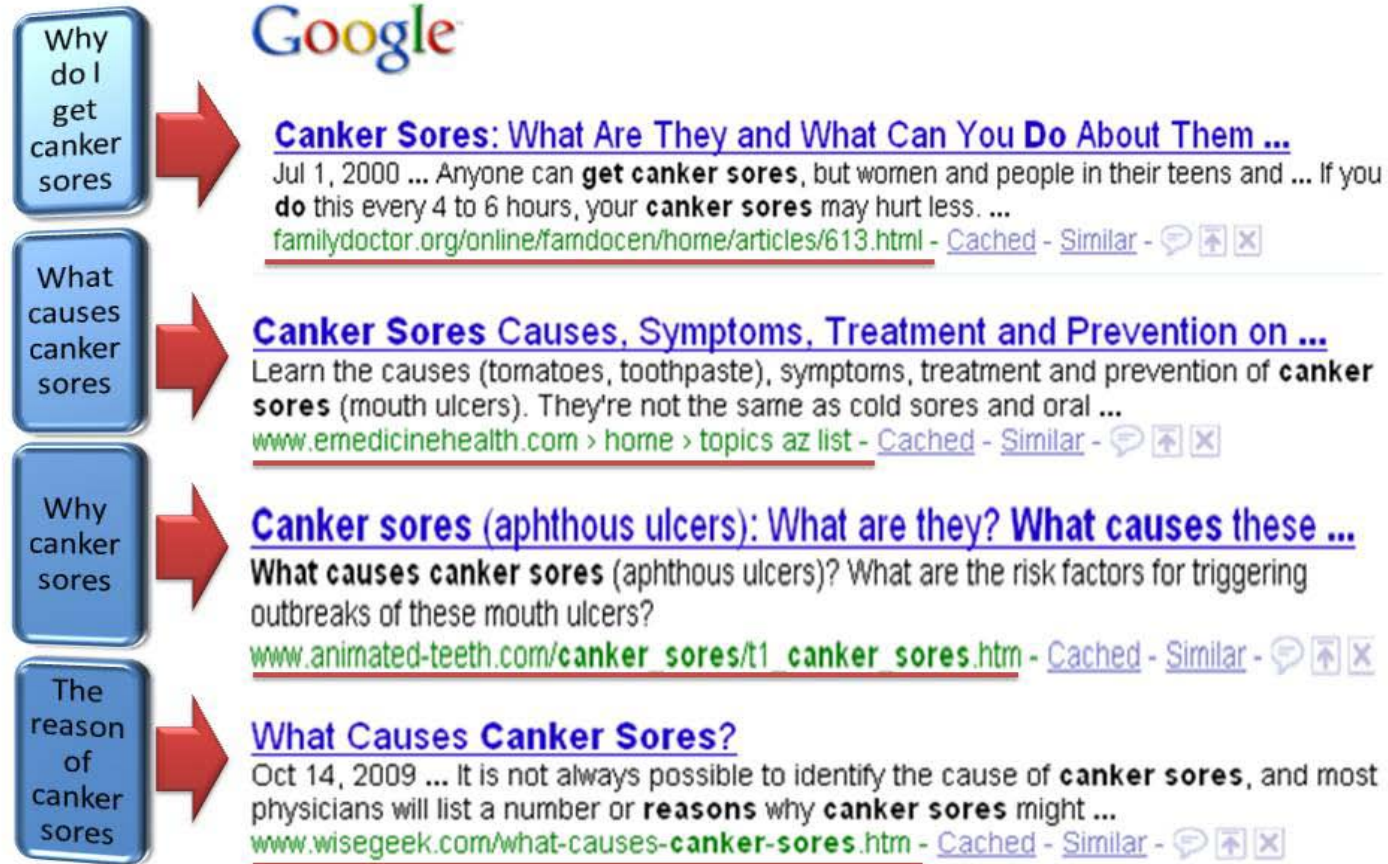
Many search engine companies use the word “semantic”. Are they really semantic ?

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



Returned Results



The image shows a Google search interface with the Google logo at the top. Below the logo, four search queries are listed in blue rounded rectangles on the left, each with a red arrow pointing to a search result. The search results are as follows:

- Why do I get canker sores** → [Canker Sores: What Are They and What Can You Do About Them ...](#)
Jul 1, 2000 ... Anyone can **get canker sores**, but women and people in their teens and ... If you **do** this every 4 to 6 hours, your **canker sores** may hurt less. ...
familydoctor.org/online/famdocen/home/articles/613.html - [Cached](#) - [Similar](#) -   
- What causes canker sores** → [Canker Sores Causes, Symptoms, Treatment and Prevention on ...](#)
Learn the causes (tomatoes, toothpaste), symptoms, treatment and prevention of **canker sores** (mouth ulcers). They're not the same as cold sores and oral ...
[www.emedicinehealth.com > home > topics az list](http://www.emedicinehealth.com/home/topics_az_list) - [Cached](#) - [Similar](#) -   
- Why canker sores** → [Canker sores \(aphthous ulcers\): What are they? What causes these ...](#)
What causes canker sores (aphthous ulcers)? What are the risk factors for triggering outbreaks of these mouth ulcers?
[www.animated-teeth.com/canker sores/t1 canker sores.htm](http://www.animated-teeth.com/canker-sores/t1-canker-sores.htm) - [Cached](#) - [Similar](#) -   
- The reason of canker sores** → [What Causes Canker Sores?](#)
Oct 14, 2009 ... It is not always possible to identify the cause of **canker sores**, and most physicians will list a number or **reasons** why **canker sores** might ...
www.wisegeek.com/what-causes-canker-sores.htm - [Cached](#) - [Similar](#) -   

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 para-queries, although not sufficient!

Measure Metric

- **Semantic Invariance**

- Be invariant to semantically equivalent queries

- Equivalent queries \Rightarrow 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 \neq 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 \Leftrightarrow 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

TALLADEGA

Joe Girard

Bill Franken

Jeremy Sinar

Celia Jay

Roman Giro

Phil Lafayette

Kerry French

Waiting for player to join

Waiting for player to join

TALLADEGA

How many species of SNAKEs are there?:

<< Return to front page

Score History

Chat here

© 2004 Rutgers University

Search for popular questions

Rephraser Game



	Jan Davis Patents 2 Score 2
	Bill Franken Patents 2 Score 2
	Jeremy Siner Patents 4 Score 4
	Celia Jay Patents 2 Score 2
	Roman Oles Patents 5 Score 25
	Phil Lafayette Patents 3 Score 3
	Kerry French Patents 1 Score 20
	Waiting for player to join.
	Waiting for player to join.



How many species of SNAKEs are there?: Go

« Return to front page Search for popular questions

Score History

- 3 how many different kinds of snakes are there
- 2 types of snakes
- 1 species of snakes

Repeated by: Phil Lafayette +1

Repeated by: Celia Jay +1

Scored to: Jeremy Siner 1

Chat here Chat

Game ends in 10m: 15s

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

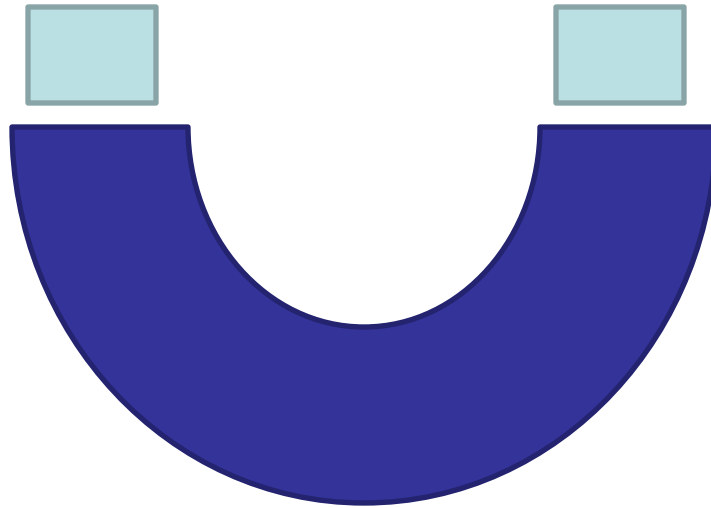
Visible Objects



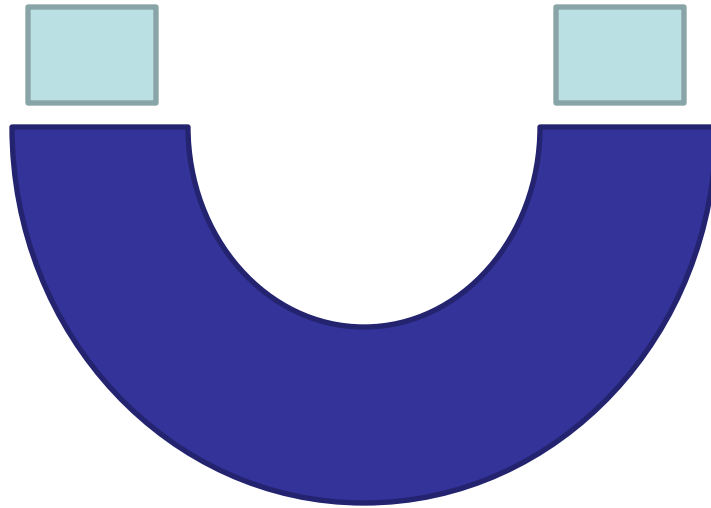
Visible Objects



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 ≥ 3 human acceptability ratio = 89%
 - Human acceptability ratio = $\#(\text{agreed persons}) / \#(\text{voting persons})$
 - 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:

The Google logo, featuring the word "Google" in its characteristic multi-colored font.The Bing logo, featuring the word "bing" in a blue, lowercase, sans-serif font.The Yahoo! logo, featuring the word "YAHOO!" in a red, uppercase, serif font.The Ask.com logo, featuring the word "Ask" in white, italicized, sans-serif font inside a red oval, with ".com" in a smaller font below it.

Paraphrasing Invariant Coefficient

- Associated Queries
 - Queries are associated by Search Engine Return Page
 - Query q_1 and q_2 can have the same SERP, or not
- Para-Queries
 - Suppose q_1 and q_2 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 : $\#(\text{pairs with same top URL}) / \#(\text{pairs})$

Result

Templates	Google	Bing	Yahoo	Ask
what causes a Headache	0.11	0.10	0.08	0.14
when was Mark Twain born	0.52	0.24	0.26	0.55
who wrote THE BIBLE	0.28	0.36	0.21	0.18
where can I buy VIAGRA	0.13	0.18	0.14	0.13
currency of RUSSIA	0.21	0.25	0.18	0.19
how much do TEACHERS make a year	0.38	0.26	0.19	0.35
where is RUSSIA located	0.15	0.17	0.13	0.10
Is COPS on tonight?	0.29	0.35	0.28	0.22
fun things to do in NYC	0.09	0.09	0.05	0.37
what are the health benefits of MILK	0.25	0.15	0.10	0.22
Daytona 500 pole winner	0.26	0.25	0.16	0.27
What's the oldest CITY in the U.S.A	0.26	0.29	0.23	0.22
Who is the governor of Alaska now?	0.50	0.53	0.35	0.39

- 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 ($q_1 \dots q_n$)
- Calculation
 - Given a set of para-queries ($q_1 \dots q_n$), it has a corresponding top URL set ($u_1 \dots u_m$) 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

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

- 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 ?