Why users need semantic search

Barney Pell, Ph.D.
Search strategist and evangelist @ Bing
Today's Discussion

Market Trends

Search Challenges

Semantic Opportunity
Online Information Explosion

- More and more sites...
  100K to 160M sites

- With more and more pages...
  26M to 1 Trillion URLs

- With richer and richer content...
  17 TB to 5PB
  - Images
  - Audio
  - Video

Bing
Users Depending More on Search
But satisfaction has not improved

People rely heavily on search for daily tasks...

- Visit social networking site: 13%
- Surf web for fun: 28%
- Research hobby: 29%
- Check weather: 30%
- Check news: 39%
- Online search: 49%
- Email: 60%

% who do this on a typical day

...in ever increasing numbers

- March 2008: 188 Million
- March 2009: 202 Million

8% Growth in Users

...using search more and more

- March 2008: 23 visits per searcher
- March 2009: 30 visits per searcher

30% Growth in Visits

...yet satisfaction with search could improve

- 2005: 68%
- 2008: 65%

Sources: Pew Research Center; ComScore; Harris Poll
Unmet Needs: The 3 Problems of Search

25% of queries result in quick click backs

Almost 50% of time spent searching is on these queries
<table>
<thead>
<tr>
<th>Time</th>
<th>Source</th>
<th>Search/Link</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:34</td>
<td>Yahoo</td>
<td>Sanyo Xacti VPC-E1075</td>
<td>1 min m1.2mdn.net/1447899/</td>
</tr>
<tr>
<td>14:17</td>
<td>Google</td>
<td>Sanyo Xacti VPC-E1075</td>
<td>5 mins <a href="http://www.walmart.com/catalog/">www.walmart.com/catalog/</a></td>
</tr>
<tr>
<td>14:22</td>
<td></td>
<td>Sanyo Xacti VPC-E1075</td>
<td><a href="http://www.ciao.co.uk/Sanyo_VPC_E1075_7006224">www.ciao.co.uk/Sanyo_VPC_E1075_7006224</a></td>
</tr>
<tr>
<td>14:22</td>
<td></td>
<td>Sanyo Xacti VPC-E1075 review</td>
<td>1 min <a href="http://www.letsgodigital.org/">www.letsgodigital.org/</a></td>
</tr>
<tr>
<td>14:23</td>
<td></td>
<td>Sanyo Xacti VPC-E1075 review</td>
<td>35 mins <a href="http://www.dell.com/">www.dell.com/</a>; accessories.us.dell.com/sna/</td>
</tr>
<tr>
<td>17:05</td>
<td></td>
<td>Kodak EasyShare M1033 Pink</td>
<td>33 mins reviews.cnet.com/digital-cameras/; <a href="http://www.walmart.com/">www.walmart.com/</a></td>
</tr>
<tr>
<td>Time</td>
<td>Source</td>
<td>Query</td>
<td>Duration</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>--------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>17:38</td>
<td></td>
<td><strong>Kodak EasyShare M1033 Pink</strong></td>
<td>11 mins</td>
</tr>
<tr>
<td>17:49</td>
<td></td>
<td><strong>Kodak EasyShare M1033 Pink</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Kodak EasyShare M1033 Pink</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Kodak EasyShare M1033 Pink</strong></td>
<td></td>
</tr>
</tbody>
</table>

**NEXT DAY**

<table>
<thead>
<tr>
<th>Time</th>
<th>Source</th>
<th>Query</th>
<th>Duration</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:12</td>
<td>Yahoo</td>
<td><strong>Kodak EasyShare M1033 Pink</strong></td>
<td>5 hrs 10 mins</td>
<td>cgi.ebay.com/Kodak-EasyShare-M1033-HD; <a href="http://www.walmart.com/browse/">www.walmart.com/browse/</a></td>
</tr>
<tr>
<td>20:22</td>
<td></td>
<td><strong>Kodak EasyShare Z1285-IS Black review</strong></td>
<td>4 mins</td>
<td><a href="http://www.cnet.com/1770-5">www.cnet.com/1770-5</a></td>
</tr>
</tbody>
</table>

15 Unique Queries
12 Partial Re-Queries and Refinements
2 Days
2 Search Engines
nausea headache pregnant what's wrong
nausea headache pregnant 3rd trimester
define:malaise
headache pregnant 3rd trimester
gestational diabetes
first signs of gestational diabetes
Typical Session: Health

16:15 Live
- physical therapy resource
  - 2 mins
  - www.ptcentral.com/resources

16:17 rehabilitation resources
- therapeutic exercise rehabilitation resources
  - 1 min
  - www.nmdinfo.net

16:18 therapeutic exercise rehabilitation resources
- physical therapy case studies
  - 1 min
  - pro-motionphysicaltherapy.com

16:19 physical therapy case studies
- the use of a single case research designs and rehabilitation studies
  - 2 mins
  - www.rehab.med.usyd.edu.au

16:23 young athlete herniated disc
- 1 mins
  - www.uptodate.com

16:24 young athlete herniated disc
- 9 mins
  - www.spineuniverse.com

16:33 young athlete herniated disc
- 8 mins
  - cc.msncache.com; www.baystatept.com/specialties; cc.msncache.com/cache

16:41 young athlete herniated disc rehab program
- 1 min
  - www.adapttraining.com/training/Bio.htm

16:42 young athlete herniated disc rehab program
- 3 mins
  - my.clevelandclinic.org/services/pediatric

16:45 young athlete herniated disc rehab program
- ??? mins
  - www.spinediscrehab.com

16 Unique Queries
8 Partial Re-Queries and Refinements
40 Minutes
<table>
<thead>
<tr>
<th>Time</th>
<th>Service</th>
<th>Query</th>
<th>URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>19:14</td>
<td>Google</td>
<td>Orbitz Travel</td>
<td><a href="http://www.orbitz.com">www.orbitz.com</a></td>
</tr>
<tr>
<td>19:15</td>
<td>Google</td>
<td>Cancun all inclusive vacation packages</td>
<td>cancun.bookit.com; <a href="http://www.travelocity.com">www.travelocity.com</a>; from.bookit.com</td>
</tr>
<tr>
<td>19:29</td>
<td>Google</td>
<td>Allegro Playacar-All Inclusive</td>
<td>from.bookit.com; <a href="http://www.occidentalhotels.com">www.occidentalhotels.com</a></td>
</tr>
<tr>
<td>19:31</td>
<td>Google</td>
<td>Occidental Grand Xcaret - All Inclusive Resort</td>
<td>from.bookit.com; search.bookit.com; <a href="http://www.occidentalhotels.com">www.occidentalhotels.com</a>; from.bookit.com</td>
</tr>
<tr>
<td>19:34</td>
<td>Google</td>
<td>NH Riviera Cancun - All Inclusive</td>
<td><a href="http://www.cancun.com/Hotels/NH-Riviera-Cancun">www.cancun.com/Hotels/NH-Riviera-Cancun</a>; <a href="http://www.tripadvisor.com">www.tripadvisor.com</a>; <a href="http://www.travel-center.com">www.travel-center.com</a></td>
</tr>
<tr>
<td>19:42</td>
<td>Yahoo</td>
<td>Cancun theme park</td>
<td><a href="http://www.cancun.com/Ecological_Parks/Xcaret">www.cancun.com/Ecological_Parks/Xcaret</a></td>
</tr>
<tr>
<td>19:43</td>
<td>Live</td>
<td>xcaret</td>
<td>search.live.com/results; search.live.com/video</td>
</tr>
</tbody>
</table>
Sessions are Decision Oriented

Users Relying More on Search

Q. In the past six months have you used a search engine to help inform your decisions for the following tasks?

- Product Purchase: 75%
- Local Activity: 62%
- Flight or Hotel: 45%
- Healthcare Research: 43%

66% of people are using search more frequently to make decisions.

Decision Sessions are Lengthy

Length of Sessions by Type

- Navigational: 1,90 queries/session, 2.50 average length (mins)
- Transactional: 4,76 queries/session, 8.87 average length (mins)
- Informational: 6,03 queries/session, 9.60 average length (mins)

Users need help with tasks and making decisions.

Complex task and decision sessions could be easier.

Sources: Microsoft Internal Research conducted by iPos 2009; Live Search Session Analysis
### Opportunities for Innovation

<table>
<thead>
<tr>
<th>Search Challenges</th>
<th>Opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Users are Dissatisfied</strong></td>
<td>Better Results</td>
</tr>
<tr>
<td>- Results not satisfying intent</td>
<td></td>
</tr>
<tr>
<td>- Many wasted clicks</td>
<td></td>
</tr>
<tr>
<td>- Repeated searches</td>
<td></td>
</tr>
<tr>
<td><strong>Users Want Help Searching</strong></td>
<td>Organized Experience</td>
</tr>
<tr>
<td>- Usage is in sessions, not queries</td>
<td></td>
</tr>
<tr>
<td>- Half of time is spent on long sessions</td>
<td></td>
</tr>
<tr>
<td>- Significant time spent refining or repeating queries</td>
<td></td>
</tr>
<tr>
<td><strong>Increasing Focus on Tasks &amp; Decisions</strong></td>
<td>Task and Decision Tools</td>
</tr>
<tr>
<td>- More reliance on search to make decisions</td>
<td></td>
</tr>
<tr>
<td>- Users focused on accomplishing tasks</td>
<td></td>
</tr>
<tr>
<td>- Search not optimized for tasks or decisions</td>
<td></td>
</tr>
</tbody>
</table>
Entity Centered Experiences

- Arise when the user’s intent is focused on a specific entity
- Entity Cards, Table of contents, related entities
- Key issues
  - Ambiguity in names requires entity resolution techniques
    - Name disambiguation dialog
  - Entities must be disambiguated to uniquely ID in massive DBs
    - Today: mixes different entities across a single result page
  - Hand-curated approaches are expensive. Requires:
    - Better structured data and knowledge management
    - Methods for extracting and synthesizing knowledge from disparate content
E-Card, TOC and Related Entities

2010 Winter Olympics - Lindsey Vonn

Country: USA  Sport: Alpine Skiing
Olympics: 2006, 2002
Height: 5'10"  Weight: 160  DOB: 1984-10-18
Events: Super-combined, downhill, super-G, giant slalom, slalom
Bio  Video  News  Photos

Source: nbcolympics.com

Lindsey Vonn - Wikipedia, the free encyclopedia  1 day ago
Lindsey Vonn (née Kildow, born October 18, 1984) is an alpine ski racer on the U.S. Ski Team. She is the first American woman to win back-to-back overall World Cup championships ...

en.wikipedia.org/wiki/Lindsey_Vonn  Wikipedia on Bing  Mark as spam

U.S. skier Lindsey Vonn  11 hours ago
Lindsey Vonn, a 25-year-old from St. Paul, Minn., is a three-time Olympian and the top U.S. female medal contender in alpine skiing.
Semantic Improvements to Core Search

**Semantic Retrieval & Ranking**
- Entity tagging & classification
- Improved precision/recall

**Semantic Query Understanding**
- Target: long, NL, phrase, entity-attrib
- Data: user, history, session, query sem.

**Presentation & Captions**
- Smarter text selection
- Structured data

**Answers & Question Answering**
- Structured & unstructured data
Semantic Retrieval & Ranking

- Better entity tagging
  - E.g. “Barack Obama”, “President Obama”, “the president”, “he”

- Derive semantic graphs from text
  - “Obama was endorsed by Senator Clinton after she conceded the race”
  - [sb:Clinton, rel:endorse, ob:Obama], [sb:Clinton, rel:concede, ob:race]
  - Allows precision of semantic roles to be enforced
  - Answers “who endorsed Obama”, not “who endorsed Clinton”

- Enables retrieval of fully instantiated triples from partially instantiated triples

- Semantic proximity a better indicator of relevance than textual proximity
Semantic Query Understanding

Term expansion
- Synonyms, acronyms, abbreviations

Flexible syntax
- {height empire state building}
- {how tall is the empire state building?}

Entity detection
- Carnegie Mellon University Robotics Club
- [ [ [ Carnegie ] [ Mellon ] University ] Robotics Club ]

Disambiguation in context
- Hilton locations v. Hilton appearances

Entities & attributes
- Barack Obama birthdate
Presentation & Captions

- Match the meaning of the query rather than just the keywords
- Provide answers and relevant meta-data directly in the caption
- Respect syntactic and semantic boundaries
- Structured data in captions
- Summarized captions and content preview (hover)
...85% of the population suffers from omega 3 fish oil deficiency.

...would qualify for a basic rebate of ...

...would qualify for a basic rebate of $600.

Query: bill gates parents


William Henry "Bill" Gates III (born ... Gates was born in Seattle, Washington, to William H. Gates, Sr. and Mary Maxwell Gates, of English, German, and ... over with his parents, who ...
The debate over the Economic Stimulus Package has been going on forever, but it hasn't even been a whole month.

It seems like...the debate over the Economic Stimulus Package has been going on forever...

The...debate over the Economic Stimulus Package has been going on forever...
Captions with Word Variation

**Query:** who mocked sarah palin

- **Sarah Palin** Prior to her appearance on the show, she had been **parodied** several times by Tina Fey, who was noted for her physical resemblance to the candidate. ... ↑ "Alaska Governor **Sarah Palin**".

- **Parodies of Sarah Palin** The comic strip Doonesbury **mocked Palin** through the use of Zipper Harris, a slacker college student. ... **Kelly Ripa impersonated Palin** on the Halloween edition of Live with Regis and Kelly (broadcast October 31, 2008), parodying phrases associated with **Palin** such as "Say it ain't so, Joe" and "Maverick", wearing **Palin**'s well-known red outfit, and speaking in **Palin**'s accent.

- **Public image and reception of Sarah Palin** See also: **Parodies of Sarah Palin** Soon after the RNC, **Palin** quickly became a favorite subject of satire and **derision**.
Captions with Structured Data

NOPA - Western Addition/NOPA - San Francisco, CA
560 Divisadero St (between Fell St & Hayes St), San Francisco, CA 94117
(Map) · User rating: 4/5 · 1484 reviews
www.yelp.com/biz/nopa-san-francisco · Cached page · Mark as spam

Download details: Internet Explorer Developer Toolbar
The Internet Explorer Developer Toolbar provides several features for exploring and understanding Web pages. These features enable you to: Explore and modify the ...
624 KB · Version 1.00.2188.0 · Published 5/9/2007
Recipes: chicken

**Chicken Schnitzel**
This traditional Austrian dish is a healthy, hearty meal of breaded chicken, potatoes, and cabbage.

- **Source**: Delish.com
- **User Ratings**: ⭐⭐⭐⭐⭐
- **Calories**: LOW

**Rosemary Chicken**

- **Source**: Delish.com
- **User Ratings**: ⭐⭐⭐⭐⭐
- **Calories**: LOW

**Pesto Chicken Pizzas**

- **Source**: Delish.com
- **User Ratings**: ⭐⭐⭐⭐⭐
- **Calories**: LOW

**Homemade Chicken Noodle Soup**

Our version of the ultimate feel-good food. Homemade Chicken Noodle Soup is chock-full of seasonal vegetables like leeks, carrots, celery, and onions. Slowly cooking a whole...

- **Source**: Delish.com
- **Calories**: LOW
Faceted Search

Medstory Health Search: low blood sugar - Windows Internet Explorer

Search: low blood sugar
Go

Feedback | Disclaimer | Help

Results for low blood sugar

Information that Matters™: click below to refine your search

Drugs & Substances
- Insulin
- Lantus
- Glucagon
- Glucophage
- Glyburide

Conditions
- Hypoglycemia
- Type 1 Diabetes
- Hyperglycemia
- Type 2 Diabetes
- Weight Gain

Procedures
- Glycosylated Hemo...
- Blood Glucose Level
- Triglyceride Level
- Insulin Pump
- Blood Glucose Sel...

In Clinical Studies
- Exenatide
- Detemir
- Glibenclamide
- Trahpane Insulin
- Pramlintide

Personal Health
- Exercise
- Diabetic Diet
- Low Carbohydrate ...
- Hospitalization
- Smoking

Nutrition
- Dietary Fiber

People
- Davis, Stephen N
- Hering, Bernard J
- Tamborlane, Willi...
- Umpierrez, Guillermo
- Okabayashi, Takehiro

The Web

The Web 1 to 10 of about 1,340,000

1. Hypoglycemia
Task-oriented Experiences

- User wants to make a decision and get something done
  - Decide and book movies, restaurants, flights, hotels
  - Research, buy and review products
  - Tasks can require many sessions and hours of work
- Trend: publishers exposing their services via APIs
- Help user string small steps together for larger tasks
- Guide users through available services & requirements
Conversational Assistant & Task Achievement
“Get me a table for two at Don Giovanni (at 8pm)”
### Summary: Semantics is key for reducing today’s search pain

| Deliver Great Results | Tools that provide faster access to essential information  
|                       | Inclusion of structured data and knowledge  
<table>
<thead>
<tr>
<th></th>
<th>Better document and query understanding</th>
</tr>
</thead>
</table>
| A richer, more organized experience | Better organized page results  
|                       | Faceted refinement  
|                       | Smart suggestions and anticipation of follow-on queries |
| Help user accomplish key tasks more easily | Bring user the best of shopping, local, travel and health  
|                       | Powerful decision making tools  
|                       | End-to-end experiences for key decision tasks and decisions |
Concluding remarks

- User needs have evolved
- Data and services are proliferating
- Search innovations are going live daily
- So now is a great time to work on semantics and search!