The new frontiers of Web search: going beyond the 10 blue links

Ricardo Baeza-Yates, Andrei Broder, Yoelle Maarek and Prabhakar Raghavan

Yahoo! Labs
Technology hot spot

OLD HOTSPOT

NEW HOTSPOTS

The Web

Web Spider

Indexer

Indexes

User

Search

© Yahoo! Research 2010
The evolution of commercial web search engines

- **First generation – Syntactic match**
  - Use only “on page”, text data, TF/IDF, etc

- **Second generation -- Use off-page, web-specific data**
  - Link (or connectivity) analysis
    - Sophisticated mathematical methods
  - Click-through data (What results people click on)
  - Anchor-text (How people refer to this page)

- **Third generation – Semantic analysis, answer “the need behind the query”**
  - Focus on user need, rather than on query
  - Integrates multiple sources of data,
  - Help the user
    - UI, spell checking, query refinement, query suggestion, syntax driven feedback, context help, context transfer, etc

- **What’s next? – Focus on task completion**

1994-1997 AV, Excite, Lycos, etc

From 1998. Made popular by Google but everyone now

Still evolving
Half-dozen challenges

1. Assisting the users in creating “better” queries
2. Integrating rich and complex data sources
3. Gathering and exposing Web derived knowledge → the “Web of Things.”
4. Creating tools that facilitate a superior post-search experience → “Direct answers”
5. Facilitating the integration of third party applications & social input
6. Implicit Search
Challenge #1: User assistance
Make the search box “reactive”

• Today: Query suggestions appear as you type in the search box
  → From the 2.3 words query to the 2.3 letter query?? 😊😊
• Tomorrow: Establish a dialog? Richer interaction? Voice?
Challenge #2: Data integration
Bring data from various sources

- Today: from maps/news/videos/livefeeds…

Google search for "geneva"
About 37,600,000 results (0.41 seconds)

Hotels in Geneva
www.FastBooking.com Charm, Design, Luxury or Family Up to 60% Discount! No prepayment
Hotels in Paris - Hotels in London - Hotels in Lisbon - Hotels in Rome

Visiting Geneva?
Geneva.TripAdvisor.co.uk Geneva Hotel Deals and Geneva Hotel Reviews!

Geneva Switzerland maps.google.com

Challenge #3: Web derived knowledge

“Fondue Palo Alto”

**Fondue Restaurants near Palo Alto, CA**

1. **Cold Stone Creamery** ★★★★★ (11)
   - coldstonecreamery.com
   - (650) 323-2102 - 9 Town And Country Vlg, Palo Alto, CA
   - Menu: fondue
   - 11 Reviews

2. **Crepes Cafe** ★★★★★ (11)
   - crepescafe.com
   - (650) 473-0506 - 1195 Merrill St, Menlo Park, CA
   - Menu: fondue
   - 11 Reviews

3. **Haagen-Dazs Shop** ★★★★★ (2)
   - haagen-dazs.com
   - (650) 326-3886 - 203 University Ave, Palo Alto, CA
   - Menu: fondue
   - 2 Reviews

47 More Local Results...

**Fondue near 2455 El Camino Real, Palo Alto, CA 94306 on ...**

Fondue near 2455 El Camino Real, Palo Alto, CA 94306...
In search of fondue: the good and the bad

Good:
- Fondue is on the (non-explicit) menu
- Geo-knowledge: Menlo Park is near Palo Alto
Fondue search: the good and the bad (cont)

Good: Local store associated to correct web site, although no list of stores available on site
Bad: Fondue is not on the menu, it is a recipe using ColdStone ice cream.

Tastemaster’s Tips

Ice Cream Fondue
Makes: 8 to 12 servings

Ingredients/Tools:

- 1 quart of your favorite Cold Stone Creamery ice cream flavor
- 4 bottles SMUCKER’S® Magic Shell Topping
- Toppings to taste—nuts, chocolate chips, coconut, candy, sprinkles or crushed OREO® cookies
- Toothpicks to serve

Follow these steps:

1. Pre-chill a large cookie tray in the freezer for 20 to 30 minutes.
2. Working quickly, so the ice cream doesn’t melt, use a melon ball scooper to make small round scoops of the ice cream and place them on the pre-chilled cookie tray. Insert a large
To blue links out of about 2,680,000 for Brett Favre
Searching for Brett Favre
Old: Find relevant pages → “web of pages”
Deconstructing the Web & creating a web of objects

Crawls, feeds: extract content/data

Identify potential objects, their attributes & possible relationships

Normalize to structured objects with relationships
Find relevant facts/entities & combine → “web of things”
Tomorrow: dynamically generate content: madonna

Madonna haunted by Ritchie’s ex

Video Premiere: “4 Minutes”

Related Topics:
- concert tickets
- organic gardening
- madison square garden
- madonna music
- raising malawi
- new york city

LastFM Top Tracks
1. Hung Up
2. Give It 2 Me
3. Candy Shop
4. Sorry
5. Like a Prayer
6. Miles Away
7. Heartbeat
8. 4 Minutes
9. Frozen
10. She’s Not Me
Tomorrow: dynamically generate content: Bank of America

Bank of America

Latest News

Bank of America settles Parmalat suit for $100M - AP via Yahoo News
36 minutes ago - Bank of America Corp. said Tuesday it will pay nearly $100 million to Parmalat SpA and its affiliates to settle lawsuits surrounding the Italian dairy company’s...

Details >>

Bank of America to close some branches - Yahoo News
3 hours ago - Bank of America Corp. said it plans to shrink its 5,109-branch U.S. network modestly over the next three to five years,...

Details >>

Deutsche Post Shares Upgraded to ‘Buy’ at Bank of America - Bloomberg
6 hours ago - Deutsche Post AG was raised to “buy” at Bank of America Corp., saying “we think the investment case is now more attractive”...

Details >>

Quote for Bank of America (BAC)

Last Trade: 13.34 ↑ 0.25 (1.91%)
Open: 12.95
Day’s Range: 12.91 - 13.43
52 wk Range: 2.53 - 39.50

Financial Blogs

Deal Journal Video: An SEC Trial of the Heart
Wall Street Journal by Stephen Cecere
July 29, 2005 - How did a love affair lead to convictions in an insider-trading case?
In his column today, Deal Journal’s Dennis K. Berman tackles the story of an Ernst & Young partner, who went in search of an affair and ended up with six counts of securities fraud...

Five U.S. Banks Are Too Big to Exist - Seeking Alpha by Jeff Nelson
July 27, 2005 - Fitch Ratings released a stunning report a few days ago which disclosed that just five banks (all U.S. banks) hold 80% of all derivatives risk. Bank of America (BAC), Goldman Sachs (GS), JP Morgan (JPM), Morgan Stanley (MS), and Citigroup (C)...
Challenge #4: Process search results

Scenario:

A user is planning a trip to Barcelona and issues a few related queries: “Barcelona cheap flights”, “Barcelona airport”, clicking each time on a few results.

When issuing the query “Barcelona airport transfer”, he might see…
Searchpad: auto notes, shareable, etc

Barcelona Airport Transport
www.barcelona-tourist-guide...en/airport/barcelona-airport-transport.html
type or paste new note

Hotel to Airport Transfer - Barcelona - Expedia
www.expedia.com/pubspec/scripts/eap.asp...
type or paste new note

How to Get From Barcelona Airport To City Center | eHow.com
www.ehow.com/how_2245123_barcelona-airport-barcelona-city-c...
type or paste new note
Challenge #5: Integrate third party applications & social input

Note source selection
Direct social input ➔ e.g. twitter
Third party app: search transfer
Indirect social input → buzz, trends, etc

Ranks of 'long-term unemployed' soar

A historic number of hard-hit job seekers highlights the depths of the current crisis. How they're defined →

Recession-proof jobs
• Renewing Bush tax cuts?
• Mortgage program criticism

Get Tix and Showtimes - Ad Feedback
"trending now" searches

Different presentation driven by context – other "trending now"
Challenge #6: Implicit Search
Today mostly “Content match”
Challenge #6: Implicit Search

Recommender systems
Summary

• Web search is now 15+ years old → becoming a mature technology
• Focus of innovation moving towards the “before” and “after” web search
• Numerous challenges → lots of research and technology development opportunities
Thank you!

broder@yahoo-inc.com

BTW: We are hiring (US, Beijing, Barcelona, Haifa, Bangalore...)

http://labs.yahoo.com/Job_Opportunities
This talk is Copyright 2010. Authors retain all rights, including copyrights and distribution rights. No publication or further distribution in full or in part permitted without explicit written permission.