A Possible Future for The Future of Web Search

Ricardo Baeza-Yates
VP of Yahoo! Research for EMEA & LatAm

Yahoo! Research

• Mission: to develop the world-class science that will deliver the next generation of businesses to the company

• Established an open culture of collaboration with peers from academic and research institutions.
Conclusions

• Web search is no longer about document retrieval
  – Means for web-mediated goals
• New breed of search experiences
  – Demands search ecosystem combining content with intent
  – Exploting the Wisdom of Crowds behind the Web 2.0

Trends

• User Generated Content
  – Massive (quality vs. Quantity)
  – Social Networks
  – Real time (people + sensors)
• Impact
  – Fragmentation of ownership
  – Fragmentation of access (longer tail)
  – Fragmentation of right to access
• Viability
  – Business model based in advertising (?)
What we really want?

Search is Evolving

• Already, more than a list of docs
• Moving towards identifying a user’s task
• Enabling means for task completion

• New experiences based on the Web 2.0

• Challenges: on-line, scalability
Premise:
- People don’t want to search
- People want to get tasks done and get straight to their answers

I am craving for mussels & wine in Brussels

Net

- We move from a web of pages to a web of objects
- Objects are people, places, businesses, restaurants …
- Objects have attributes
  - Missing, noisy, etc.
- Intents are satisfied by presenting objects and attributes
- Attributes define faceted search
How do we get structured objects/attributes?

• Web Content
  – Metadata/Taxonomies/Folksonomies
  – Machine learning techniques
  – Classification/Extraction/Semantic Web

• Web Usage
  – Implicit relations

• Building out an open ecosystem
  – Publishers have incentives to contribute

The SearchMonkey Ecosystem

Publishers

XML / API / RDFa / DataRSS

Publisher Content Submission

SearchMonkey

Display

Personalized SERP

Users

Templates/
PHP

Developers

Developer Tool
Opening search - what does it mean?

• Clear win for: developers, site owners, users and Yahoo!
• Go from “to-do” to “done”

The Art of Pizza - Lakeview - Chicago, IL 60657
... that say they don’t like Chicago deep dish pizza have never had Art of Pizza... Top 10 Chicago meals, I would include an Art of Pizza Deep Stuffed ...
www.acmereviews.com/biz/the-art-of-pizza-chicago - 148k - Cached

BEFORE

Enhanced Results and Infobars:
Ratings and reviews, images, deep links, and other name-value pairs.

Site Owners

• Straight to the answers
• Increase in quality of site traffic
• Fosters loyalty and engagement

BEFORE

AFTER

BEFORE

AFTER
• James Surowiecki, a *New Yorker* columnist, published this book in 2004
  – “Under the right circumstances, groups are remarkably intelligent”
• Importance of diversity, independence and decentralization

  **Aggregating data**

  “large groups of people are smarter than an elite few, no matter how brilliant—they are better at solving problems, fostering innovation, coming to wise decisions, even predicting the future”.

– Crucial for Search Ranking
– Text: Web Writers & Editors
  • not only for the Web!
– Links: Web Publishers
– Tags: Web Taggers
– Queries: All Web Users!
  • Queries and actions (or no action!)
Endless image browsing

Faceted Image Search
Visual Diversity

- **Topical diversity**
  Query: “Jaguar”

- **Visual diversity**
  Query: “Jaguar X-type”

- **Other dimensions: spatial, temporal, social**

Exploit Visual Annotations

Flickr allows another kind of annotations (notes)

- Associate **text** with **visual area**
- Highly relevant to content
  → **Visual Annotation**

- Valuable to learn different visual representations of an object

Use visual annotations (text & image) and rank aggregation to improve retrieval
Tags are better than image features

But together they are much better as they are complementary

**SEMEDIA**

- Search Environments for Media
  - Focused on image & video search
  - www.semedia.org
“... keeps track of search query terms ... when it detects a trend, offers to save the result in an online document.”

CNET
Exploiting Microformats

---

Rich abstract

Related pages based on metadata

Events from personal calendar, Conferences, and bio from Linkedin

---

Bridging implicit and explicit metadata
For topics without a Wikipedia page, Correlator creates a “synthetic page” with an overview of the topic:

- Query: art deco chicago
- Synthetic page: Defines Art Deco, Defines Chicago, Shows relations between Art Deco and Chicago
Conclusions

• Web search is no longer about document retrieval
  Means for web-mediated goals

• New breed of search experiences
  – Ecosystem: SearchMonkey
  – Usage: SearchPad (beta)
  – Usage: TagExplorer (demo)
  – Usage & tagging: Faceted Image Search (beta)
  – Content: Correlator (demo)

• Virtuous cycle from people to people

Questions?

rbaeza@acm.org
http://search.yahoo.com
http://research.yahoo.com
http://sandbox.yahoo.com