What Can Search Tell Us?
A Human-Centered Perspective

Alejandro (Alex) Jaimes
Yahoo! Research (Barcelona, Cat, Spain)

I. What is Search?
Information Retrieval
Information Retrieval
Information Interaction
Search

- Trying to find something or someone
- By seeking carefully, scrutinizing thoroughly

from late Latin circare ‘go around,’ from Latin circus ‘circle.’
NOT
A
Punctual Activity!
More than a query and a click

It is a series of activities around a specific goal
BUT
We often get distracted, our “goal” is often not specific, and we frequently change goals during the search process.
look for, seek, forage for, fish around/about for, look high and low for, ferret around/about for, root around/about for, rummage around/about for, cast around/about for, look through, hunt through, explore, scour, rifle through, go through, sift through, comb, go through with a fine-tooth comb; turn upside down, turn inside out, leave no stone, examine, inspect, check, frisk, hunt, look, quest, pursuit, manhunt, searching for, hunting for, seeking, looking for, on the lookout for, in pursuit of.
I. In order to *really* support search, we need to consider its complex, practical aspects
Main Points

- Not single query
- Not single session
- Context (implicit, explicit, internal, external)
Does in-network behaviour depend on how users “enter” the network?
Can we find patterns for different categories (e.g., social, search, aggregators, etc.)?
Session Length

Short sessions (aggregators)

Long sessions (mail)
Clusters
Main Observations 1

- The entry point of a user influences her behaviour
  - Session length
  - Type of page visited
  - Actions taken
  - Engagement
Main Observations II

- Mail
  - Longer sessions
  - Strong focus on managing contacts
- Search
  - Search task “moves” to Flickr
- Aggregators
  - Short sessions
  - Image browsing
We compare different ranking methods
II. How Can Search Analysis Be Useful?
Five Dimensions

- Personal (Personal Information Management)
- Algorithmic (improve results)
- Consumer activities
- Design (user experience)
- Macro (marketing and social aspects)
Personal Information Management

- Decision making
- Discovery
Algorithmic

- Ranking
- Better results (quality, relevance)
Consumer Activities

http://www.flickr.com/photos/shannonpaul/4623821529/
Design

- User experience
Parallel Search
Macro

- Social aspects
- Marketing
Search Behavior [with Ingmar Weber]

- 2.3 million “active” registered web search users
- Anonymous, aggregate analysis
- Who? What? How?
Search Behavior

Results I

• “Baby boomers”: older users, business & finance, more navigational
• “Liberal females”: women, shopping, more likely to have voted Obama, longer sessions
• “White conservatives”: older men, auto, business, cons. Electronics, less navigational
• “Challenged youth”: younger, lower income, music, more navigational
Search Behavior: Information Flows

Results II (w/Ingmar Weber, CIKM ’10)

• “sonia sotomayor”
  • hispanic -> everybody (burst) -> hispanic

• “avatar movie”
  • young/male -> everybody (burst) -> everybody
Yahoo! Clues (http://clues.yahoo.com)

Top Trends

Most Popular Search Queries

<table>
<thead>
<tr>
<th>Filters</th>
<th>Jun 13, 2011 to Today</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 India</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 philippines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Iran</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 China</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Japan</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 India

India's Nifty Stock Futures Advance as Valuations ..., Bloomberg
India's Infosys posts profit rise but shares fall, AFP
India Official: Last Fiscal Year Power Transmissio..., Nasdaq

2 philippines

Philippines: No word from 2 Americans' abductors, Boston Herald
Philippines says China sea claim shaky, AFP
Earthquake Off Philippines Measuring 6.2: Official..., ThirdAge

3 iran


Yahoo! Clues (http://clues.yahoo.com)

*Anonymous, based on census data*
III. What are the “hot,” “new” areas?
Key areas

- Mobile+multiscreen
- Multimedia
- Social (?)
Mobile
(it’s all mobile)
Multimedia
(conversations around media)
Social
(to be seen)
## Business Applications

APQC’s Process Classification Framework (PCF) – Operating Processes

<table>
<thead>
<tr>
<th>Process Category</th>
<th>Process group or Activity</th>
<th>Technical area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Vision &amp; strategy</td>
<td>Social Networking</td>
<td>SN Support tools</td>
</tr>
<tr>
<td>2. Products &amp; services</td>
<td>Product Recommendations</td>
<td>Recommenders</td>
</tr>
<tr>
<td></td>
<td>Social product search</td>
<td>Social search</td>
</tr>
<tr>
<td></td>
<td>Social CRM</td>
<td>SN support tools</td>
</tr>
<tr>
<td></td>
<td>Trend spotting</td>
<td>SN monitoring</td>
</tr>
<tr>
<td></td>
<td>Product quality</td>
<td>SN monitoring</td>
</tr>
<tr>
<td></td>
<td>Social marketing</td>
<td>SN support tools</td>
</tr>
<tr>
<td></td>
<td>Loyalty programs</td>
<td>Influence</td>
</tr>
<tr>
<td>3. Market &amp; selling</td>
<td>Direct marketing</td>
<td>Influence, communities</td>
</tr>
<tr>
<td></td>
<td>Advertising</td>
<td>Influence</td>
</tr>
<tr>
<td></td>
<td>Business intelligence</td>
<td>Churn, propagation, etc.</td>
</tr>
<tr>
<td></td>
<td>Churn prediction</td>
<td>Influence, propagation</td>
</tr>
<tr>
<td></td>
<td>Reputation monitoring</td>
<td>Monitoring</td>
</tr>
<tr>
<td>4. Delivery</td>
<td>Production scheduling</td>
<td>Mining of customer data</td>
</tr>
</tbody>
</table>
## Business Applications
### Management & Support Processes

<table>
<thead>
<tr>
<th>Process Category</th>
<th>Process group or Activity</th>
<th>Technical area</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Human Capital</td>
<td>Internal social networking</td>
<td>SN Support tools</td>
</tr>
<tr>
<td></td>
<td>Professional development</td>
<td>Expert routing</td>
</tr>
<tr>
<td></td>
<td>Recruiting</td>
<td>Social search</td>
</tr>
<tr>
<td>7. Information Technology</td>
<td>Resource allocation</td>
<td>Measurement</td>
</tr>
<tr>
<td></td>
<td>Information sources</td>
<td>Data preparation</td>
</tr>
<tr>
<td></td>
<td>Content Management</td>
<td>Privacy</td>
</tr>
<tr>
<td>8. Financial Resources</td>
<td>Customer &amp; product strategies</td>
<td>SN Mining</td>
</tr>
<tr>
<td></td>
<td>Customer-product mix</td>
<td>Community</td>
</tr>
<tr>
<td></td>
<td>Manage internal controls</td>
<td>Community</td>
</tr>
<tr>
<td>9. Property Management</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>10. Environmental issues</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>11. External Relationships</td>
<td>Public relations program</td>
<td>Monitoring</td>
</tr>
<tr>
<td></td>
<td>Legal and ethical issues</td>
<td>Privacy</td>
</tr>
<tr>
<td></td>
<td>Social networking</td>
<td>SN support tools</td>
</tr>
<tr>
<td>12. Knowledge Management</td>
<td>Knowledge sharing</td>
<td>Internal social networks</td>
</tr>
<tr>
<td></td>
<td>Strategic KM</td>
<td>SN Mining</td>
</tr>
</tbody>
</table>
IV. Big Data
New Data Paradigm

• Fragmentation

• Aggregation

• Multi-authorship
IT’S all ABOUT THE
CONNECTIONS,
CONNECTIONS,
CONNECTIONS,
CONNECTIONS!
On Implicit Links

“unapparent connections are stronger than ones which are obvious”

Heraclites
Fifth century BC
Main Points

1. Not single query, session; context
2. How can search analysis be useful
3. “Hot,” “new” areas
4. Big Data
Human-Centered Computing

Data → User → Hypotheses → Design → Context
Human-Centered Computing

- User Experience
- Data analysis
- Human aspects
- Machine Learning, User Modeling, Personalization
THANK YOU!
Alex Jaimes

Social Media Engagement Group
Neil O’Hare, Luca Aiello, Amin Mantrach, Pancho Tolchinsky, Jordi Vallmitjana, Ricard Sierra, Jordi Atserias, Luca Chiardini, Eduardo Graells, Michele Trevisol, Ruth Garcia.

Collaborators
Ingmar Weber, Barbara Poblete, Marcelo Mendoza
Papers listed at: http://www.jaimes-larrarte.com