What are they thinking?

Searching for the mind of the searcher

Joint Conference on Digital Libraries Vancouver, British Columbia

Daniel M. Russell June 27, 2007



Title: What are they thinking? Searching for the mind of the searcher

Abstract: We are in a new age of being able to understand what people are doing when trying to search. With internet search engines in common and constant use, we also have a new challenge to understand what people are really searching for, and what it is they want to do. Whatever people are doing, it's certainly not the same as the older models of search. How are people searching on Google? What are they thinking when they make certain queries? What is their intent? How can we discern what that intent really is?

In this talk I'll describe some of the ways we're working to understand what people are really doing, and why they're doing it that way. The goal of this work is to vastly improve the searcher use-experience by understands the minds of millions of searchers.

Bio: Daniel M. Russell is an Über Tech Lead for Search Quality & User Happiness at Google. In this job, Dan studies Google searcher behavior using a variety of methods to get closer to the real experience of searching. Most recently, Dan was a senior scientist and senior manager at the IBM Almaden Research Center in San Jose, California. He is best known for his work on IBM's Blueboard system (a large shoulder-to-shoulder collaboration system) and for establishing the basis of sensemaking theory while at Xerox PARC (work with Stu Card, Mark Stefik and Peter Pirolli). In addition to IBM and PARC, Dan has also worked in Apple's Advanced Technology Group, and taught at both Stanford and Santa Clara Universities. He enjoys word play, music, and long distance running, becoming disgruntled when all three can't be in one day.





```
[ jaguar ]
[ iraq ]
[ latest release Thinkpad drivers touchpad ]
[ ebay ]
[ first ]
[ google ]
[ brittttteny spirs ]
```



- How can we figure out what you're trying to do?
- The information signal is sometimes weak:



00:12 [actor most oscars]

- 00:10 So this is celebrity with most Oscars...
- 00:11 Actor... ah... most...
- 00:13 I'm just going to try that...most Oscars... don't know...
- 00:19 (reading) "News results for 'actors most Oscars' ... "
 huh..
- 00:25 Oh, then that would be currently "Brokeback"... "prior voices"... "truth in Oscar's relevance"...
- 00:32 ...now l know...
- 00:35 ... you get a lot of weird things..hold on...
- 00:38 "Are Filipinos ready for gay flicks?"
- 00:40 How does that have to do with what I just....did...?
- 00:43 Ummm...
- 00:44 So that's where you can get surprised... you're like, where is this... how does this relate...umm...
- 00:45 Bond...I would think...
- 00:46 So I don't know, it's interesting...
- 01:08 **Dan**: Did you realize you were in the News section?
- 01:09 Oh, no I didn't. How did I get that? . . .
- 01:10 Oooh... no I didn't.

1:15 [actor most oscars Academy]

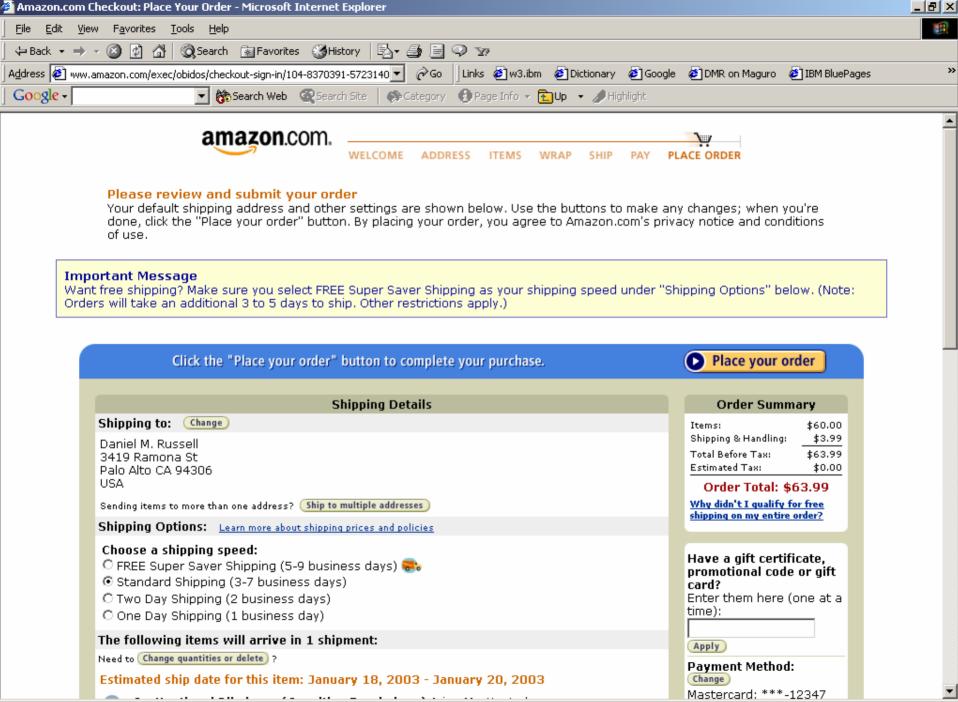


How to be literate user of a UI?

- How does one make sense of a user interface?
 - What's interactive? What's live?
 - What do various actions do?
 - What model does user have of UI?
 - Groups / Functions / Overall operation / Gestalt









No matches found.

Please check the Help page to make sure you entered the search string properly.





No matches found.

Please check the Help page to make sure you entered the search string properly.



Invisible UI elements



<u>Veb Images Video News Maps more»</u>

library database ui

Search

Advanced Scholar Search Scholar Preferences Scholar Help

Scholar All articles - Recent articles

All Results

H Korth A Silberschatz

A Sliberschatz

V Basili I Bruno

V Ogle

User interface modeling in UMLi - all 10 versions »

PP da Silva, NW Paton - Software, IEEE, 2003 - ieeexplore.ieee.org

... This article examines some of UML's user interface modeling facili- ties. ... can provide a combination of a book's properties for querying a library database. ...

Cited by 43 - Related Articles - Web Search - BL Direct

[воок] Database Design - all 2 versions »

GC Wiederhold - 1977 - McGraw-Hill, Inc. New York, NY, USA Cited by 163 - Related Articles - Web Search - Library Search

[воок] Database system concepts - all 2 versions »

HF Korth, A Silberschatz - 1986 - McGraw-Hill, Inc. New York, NY, USA

Cited by 461 - Related Articles - Web Search - Library Search

A digital library for geographically referenced materials - all 6 versions »

TR Smith - Computer, 1996 - ieeexplore.ieee.org

... library services; and the ingest facility, for storing and pro- cessing data from new holdings. CUI means graphi- cal user interface. ODBC means Open DataBase ...

Cited by 97 - Related Articles - Web Search - BL Direct



Why is chron sort so hard? Compare...

```
@article{gross96dec,
  title={{Demonstrating the Electronic Cocktail Napkin}},
  author={Gross, M.D. and Do, E.Y.L.},
  journal={ACM Human Factors in Computing-CHI},
 volume={96},
 pages=\{5--6\}
@article{mueller2005hod,
  title={{Hug over a distance}},
  author={Mueller, F. and Vetere, F. and Gibbs, MR and Kjeldskov,
  J. and Pedell, S. and Howard, S. },
  journal={Proc. CHI},
  volume={5},
 pages={1673--1676},
 year={2005} }
```



So... what do we do?

- How do we understand what people are doing?
- Between inattention and low-signal density...



Multiple views of user behavior

3M points-of-view:

Micro: lowest level details—milliseconds

Meso: mid-level observations—minutes to days

Macro: millions of observations—days to months



WHAT are people doing?

• Field studies (meso)

Getting out to see what reality is

2 Eyetracking studies (micro)

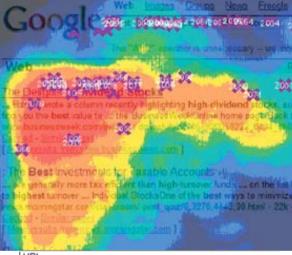
Studies in the microscopic

8 Sessions analysis (macro)

What are people doing in logs, bring outside behavior back to where we can see the signals







http://www.google.com/ig?hl=en

http://127.0.0.1:4664/search?q=wardrobe+in+NBC&num=2&shttp://www.google.com/search?hl=en&q=wardrobe+in+NBC http://ad.doubleclick.net/adi/N339.cbshealthwatch/B1809262.http://adopt.specificclick.net/adopt.sm?l=1013290775&sz=pohttp://www.cbsnews.com/stories/2004/07/01/entertainment/mhttp://127.0.0.1:4664/search?q=wardrobe+in+NBC&num=2&shttp://www.google.com/search?hl=en&q=wardrobe+in+NBC http://www.nbc.com/nbc/Will_&_Grace/wardrobe/frames/searchttp://www.nbc.com/nbc/Will_&_Grace/wardrobe/index.shtmlhttp://127.0.0.1:4664/search?q=wardrobe+in+NBC&num=2&shttp://www.google.com/search?hl=en&q=wardrobe+in+NBC http://to.0.1:4664/search?q=wardrobe+in+NBC&num=2&shttp://www.google.com/search?hl=en&q=wardrobe+in+NBC http://127.0.0.1:4664/search?q=will+and+grace+wardrobe+in-http://www.google.com/search?hl=en&lr=&safe=off&q=will+arhttp://www.nbc.com/nbc/footer/FAQ.shtml

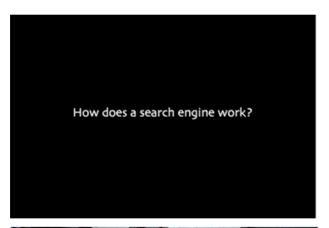
http://nbc.resultspage.com/display.php?w=wardrobe http://www.nbc.com/nbc/Will_&_Grace/wardrobe/frames/sear http://www.nbc.com/nbc/Will_%26_Grace/wardrobe/frames/se

http://www.nbc.com/nbc/Will & Grace/wardrobe/

Lies, Truth & Videotape — Field studies (meso)

Interviews held in situ…

- Workplace, home, coffee shopany place... must be search-place
- Place + context cueing effects
- Interested in natural use phenomena (ads, distractions, multiple tasks...)











What we learn from field studies...

- How people think…
- Mental models
- Qualitative approaches
- Emotional reactions
- Expectations (and violations)
- WHY we're getting certain behaviors:
 - Example: why are 50% of clicks to Advanced Search page short?



Eyetracking & usability studies at Google

~10-20 / week - typically 3 - 5 observers

Testing new, specific features of UI

Typical studies:

- How users perceive a UI change
- Eyetracking to get at deeper understanding

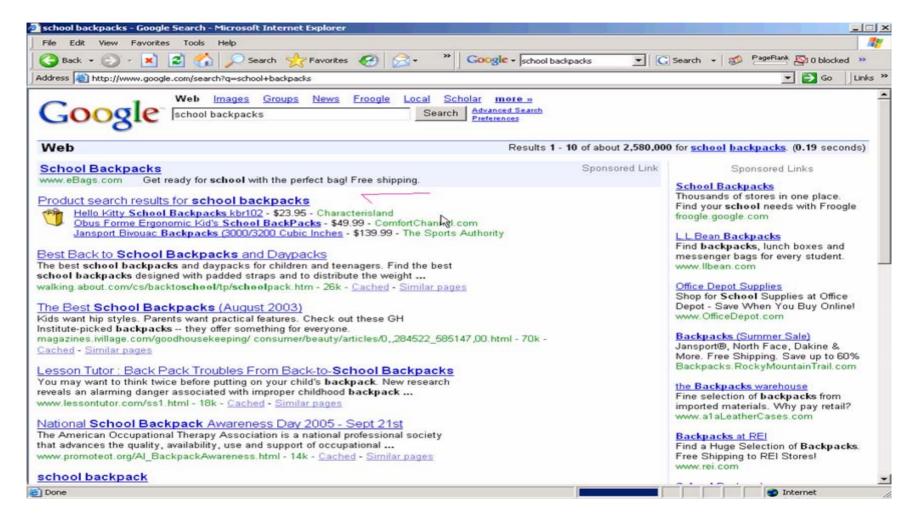




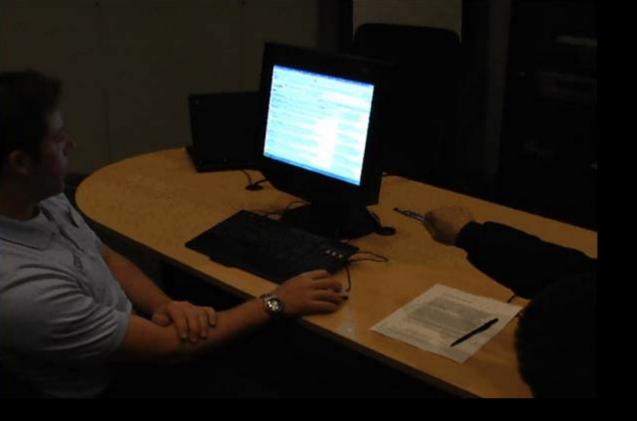




3 Google Users video







So.. Did you notice the FTD official site?

To be honest, I didn't even look at that.

At first I saw "from \$20" and \$20 is what I was looking for.

To be honest, 1800-flowers is what I'm familiar with and why I went there next even though I kind of assumed they wouldn't have \$20 flowers

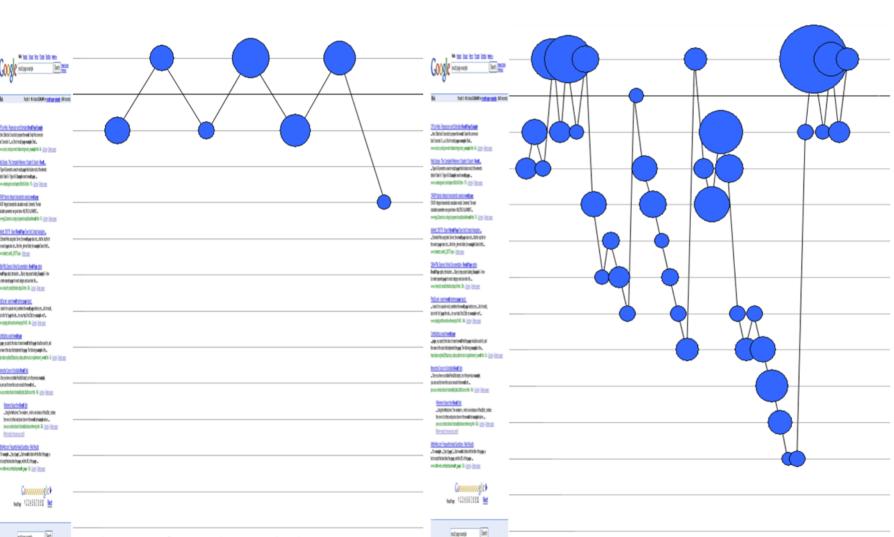
And you knew they were expensive?

I knew they were expensive but I thought "hey, maybe they've got some flowers for under \$20 here..."

But you didn't notice the FTD?

No I didn't, actually... that's really funny.

Interview video





Rapidly scanning the results

Note scan pattern:

Page 3: Result 1

Result 2

Result 3

Result 4

Result 3

Result 2

Result 4

Result 5

Result 6 <click>

Q: Why do this?

A: What's learned later influences judgment of earlier content.





How do users behave in search?

- Experiment conducted at Cornell [Gay, Granka, et al., 2004]
- Users:
 - Searched freely with any queries
 - Script removed all ad content
 - 5 info & 5 nav tasks given to participants
- Subjects (Phase I)
 - 36 undergraduate students
 - Familiar with Google

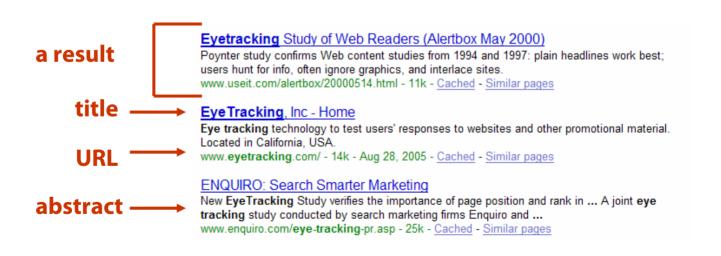
"zones" created around each result → eye-movements analyzed specific to the rankings





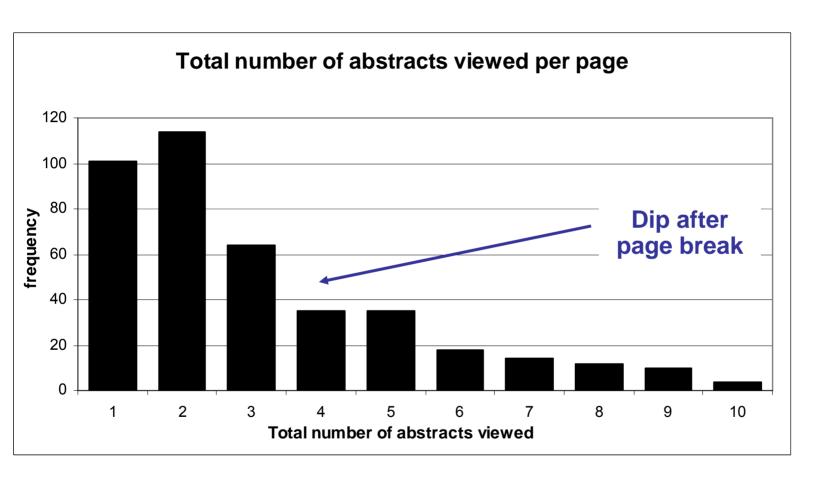
Skill of reading a SERP (search engine results page)

- How many results are viewed before clicking?
- Do users select the first relevant-looking result they see?
- How much time is spent viewing results page?





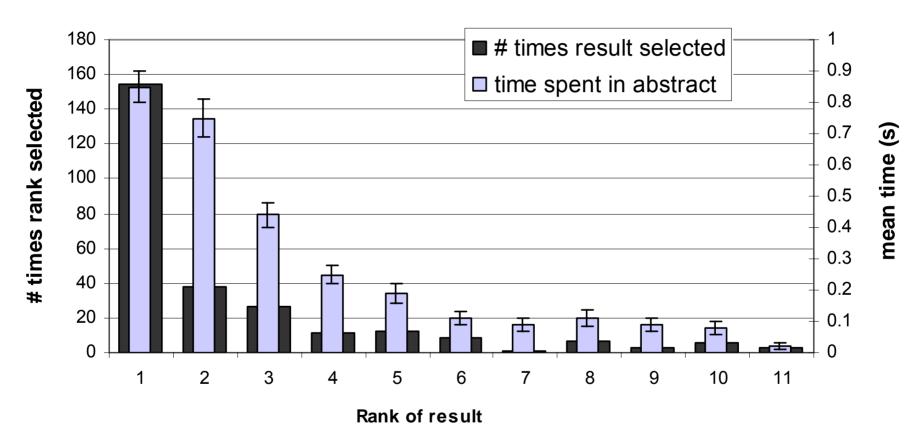
How many links do users view?



Mean: 3.07 Median/Mode: 2.00



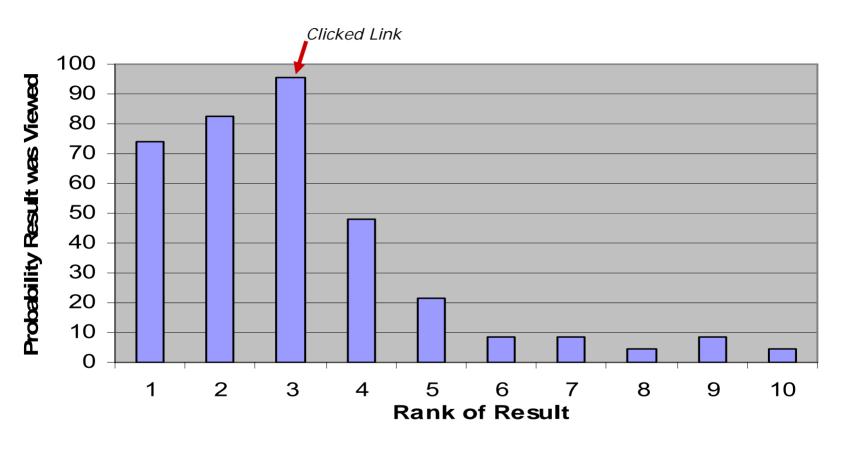
Looking vs. Clicking



- Users view results one and two more often / thoroughly
- Users click most frequently on result one



Which results are viewed before clicking?

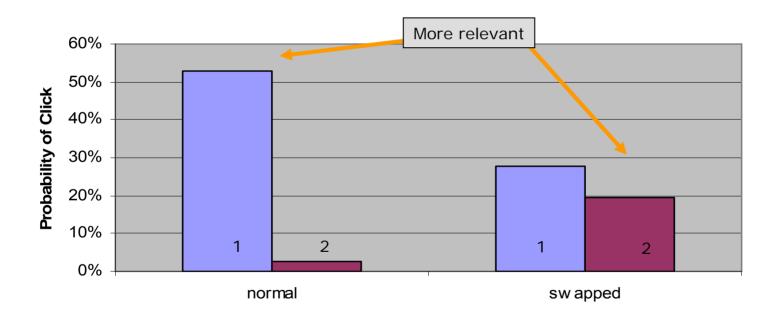


 Users typically do not look at lower results before they click (except maybe the next result)



Presentation bias – reversed results

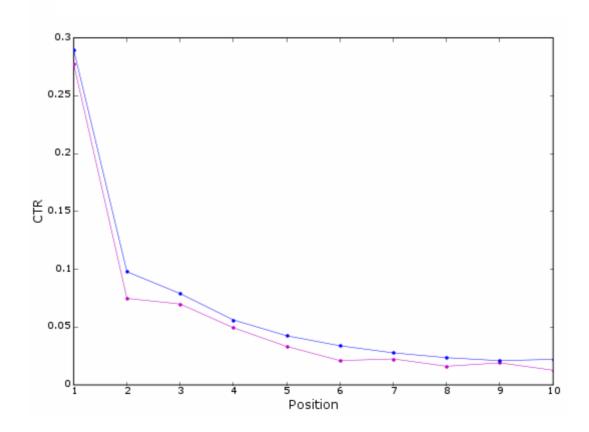
Order of presentation influences where users look
 AND where they click





Strong implicit behavior...

Users strongly believe that the search engine rank order matters





3

Macro: Understanding the behavior of the many

- We have a lot of data: many GB weekly in logs
- How to analyze it?
- How to reduce it?
 - What do you choose to forget?



A typical (long) session

```
31: Google Search [irish lotto] (4s)
     Google Result 1 www.lotto.ie/ (7s)
33:
     Google Result 1 www.lotto.ie/ (4s) (DUPE) (p=31)
34:
36:
       Nav (back/fwd) www.google.com/search (1s)
37:
     Google Result 2 www.irishlotto.net/ (2s) (p=31)
39:
       Nav (back/fwd) www.google.com/search (1s)
40:
     Google Result 3 www.irishlotto.net/main-results/2005.htm (1s) (p=31)
42:
       Nav (back/fwd) www.google.com/search (0s)
43:
     Google Result 4 www.irish-lottery.net/ (4s) (p=31)
     Google Result 4 www.irish-lottery.net/ (5s) (DUPE) (p=31)
44:
     Google Result 4 www.irish-lottery.net/ (3s) (DUPE) (p=31)
45:
48:
       Nav (back/fwd) www.google.com/search (6s)
49:
     Google Result 8 www.interlotto.com/irish/ (6s) (p=31)
51:
       Nav (back/fwd) www.google.com/search (1s)
52:
     Google Result 9 lottery.loguax.co.uk/irish-lottery.htm (21s) (p=31)
54:
       Nav casino.loquax.co.uk/ (29s)
55:
       Nav casino.loguax.co.uk/offers/173/Virgin-Casino.htm (4s)
56:
       Nav (new window) casino.loquax.co.uk/offers/173/Virgin-Casino.htm (0s)
57:
       Nav (new window) clkuk.tradedoubler.com/click (7s)
58:
       Nav (back/fwd) casino.loguax.co.uk/ (10s) (p=56)
       Nav casino.virgingames.com/game/menu.do (15s) (p=57)
59:
60:
       Nav (back/fwd) lottery.loquax.co.uk/irish-lottery.htm (0s) (p=58)
61: Google Search [irish lotto] (3s) (DUPE) (p=31)
```

Google Result 10 online.casinocity.com/lotteries/irish-lotto/ (11s)



34

Non-Google actions: "work" from the user's pov

31: Google Result 1 www.lotto.ie/ (7s)
32: Google Result 1 www.lotto.ie/ (4s) (DUPE) (p=31)
33: Google Result 2 www.irishlotto.net/ (2s) (p=31)
34: Google Result 2 www.irishlotto.net/ (2s) (p=31)
40: Google Result 3 www.irishlotto.net/main-results/2005.htm (1s) (p=31)
43: Google Result 4 www.irish-lottery.net/ (4s) (p=31)
44: Google Result 4 www.irish-lottery.net/ (5s) (DUPE) (p=31)
45: Google Result 4 www.irish-lottery.net/ (3s) (DUPE) (p=31)
49: Google Result 8 www.interlotto.com/irish/ (6s) (p=31)

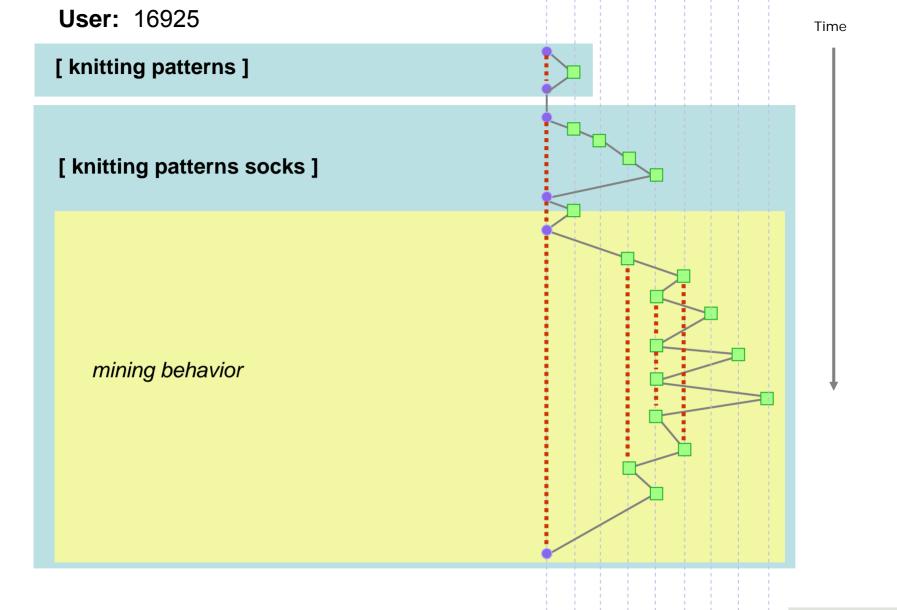
- 61: Google Search [irish lotto] (3s) (DUPE) (p=31)
- 63: Google Result 10 online.casinocity.com/lotteries/irish-lotto/ (11s)



Evidence of multitasking

```
100: Google Search [free roulette] (4s) (DUPE) (p=78)
     Google Result 7 www.getlyrical.com/general/free casino games/free online roulette.html (3s)
103: Google Result 7 www.getlyrical.com/general/free_casino_games/free_online_roulette.html (19s) (DUPE) (p=100)
     Google Result 8 www.saliu.com/Roulette.htm (56s) (p=100)
112: Google Search [shockwave] (4s)
114: Google Result 3 www.shockwave.com/sw/home/ (10s)
117: Google Result 5 sdc.shockwave.com/shockwave/download/download.cgi (16s) (p=112)
120: Google Search [free roulette] (3s) (DUPE) (p=78)
      Google Result 1 www.ildado.com/free_roulette.html (15s) (DUPE)
124: Google Search [free proffessional roulette] (2s)
126: Google Search (spell correct) [free professional roulette] (10s)
      Google Result 3 imagesculptor.com/Roulette/free-roulette-professional-system.php (5s)
129: Google Result 3 imagesculptor.com/Roulette/free-roulette-professional-system.php (8s) (DUPE) (p=126)
      Google Result 7 www.amazon.com/exec/obidos/tg/detail/-/B0007XRSQ4?v=glance (2s) (p=126)
```

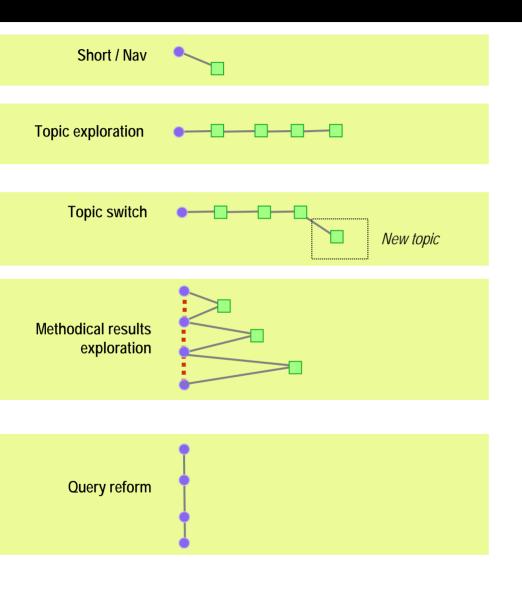


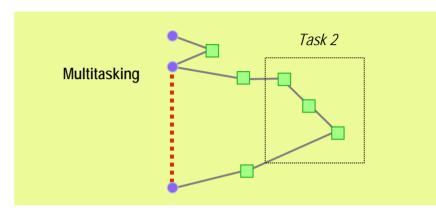


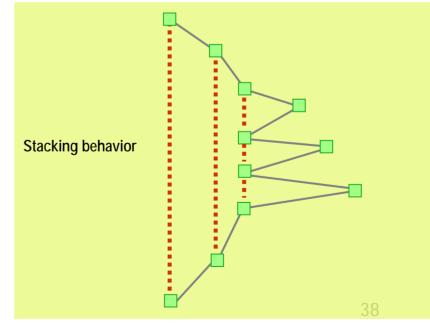


site page
SERP
same page

Kinds of behaviors we see in the data









Session patterns

Search is often a two-step process:

- 1. find or navigate to a good site ("orienteering")
- 2. browse for the answer there

[actor most oscars] vs. [oscars]

Teleporting (the other strategy)

"I wouldn't use Google for this, I would just go to..."

Possible reasons:

- don't realize that they can search for the information directly
- formulating the query seems too hard
- user trusts the source, rather than Google intermediary



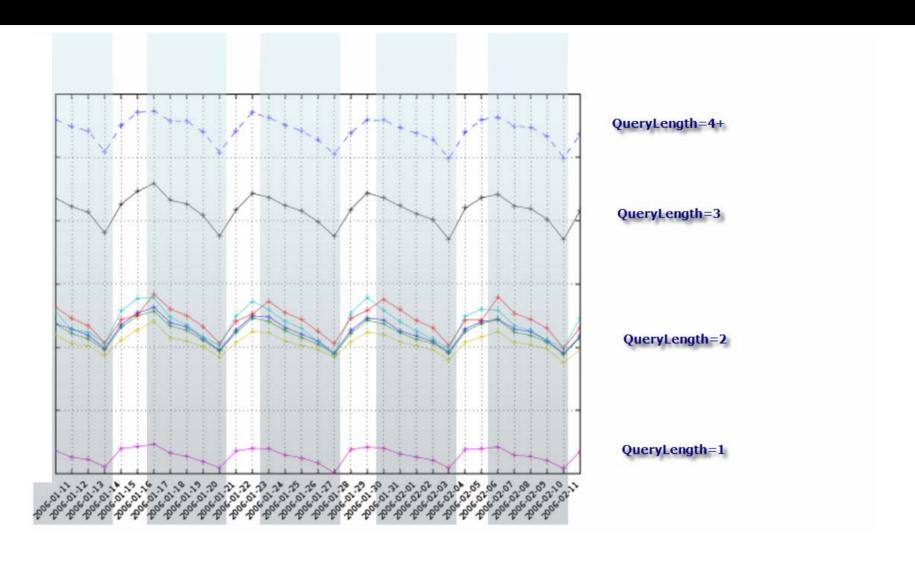
To be a strong user...

Need to have fairly deep knowledge...

- What sites are possible
- What's in a given site (what's likely to be there)
- Authority of source / site
- Index structure (time, place, person, …) → what kinds of searches?
- How to read a SERP critically



Average session duration by query length over time





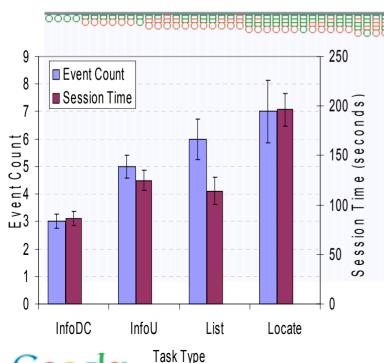
Measurable differences between different task types?

Informational/Directed/Closed

"Find a painting by Georges Seurat called "La Grande Jatte""

Informational/Locate

"Search for a man's watch that is water resistant to 100 meters and under \$100"



Main effect of task type on:

- Event count
 - (Kruskal-Wallis: $\chi^2(3)=368.3$; p<.001) and
- Session time

(Kruskal-Wallis: $\chi^2(3)=368.7$; p<.001)

Info Direct-Closed < Info Undirected <= List < Locate

Mental models

- How DO people think about what a search engine does?
 - Completely keyword search?
 - Full-text indexing?
 - Partial-text indexing?
 - Link anchors?
- What **DOES** one need to know to use search effectively?
 - Relevance?
 - Keyword term frequency?
 - Layered index?
 - Spider / crawling?

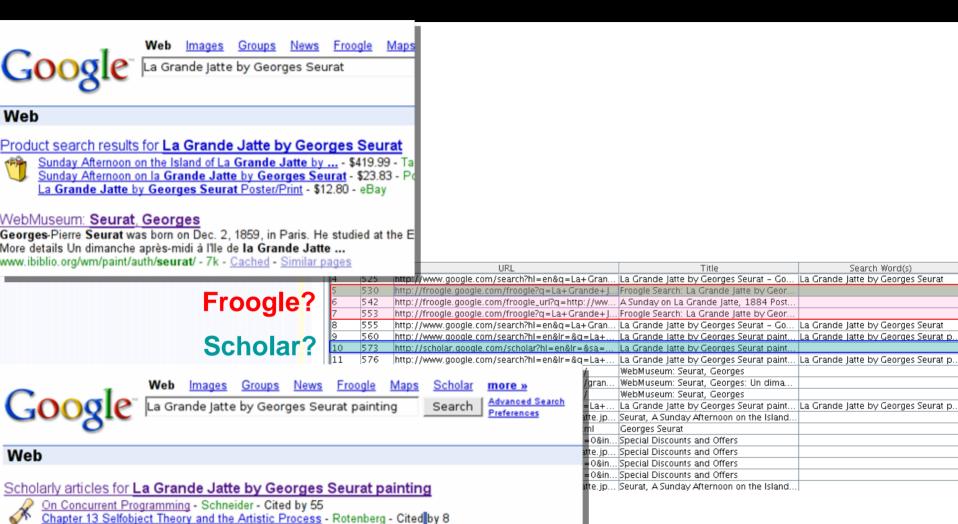


Mental models 1

- 1. Predictable behavior
- 2. What content is indexed?
- 3. How does Google look it up?
- 4. How are the results ranked?
- 5. What's in the index? (different kinds of content)



Looking for an image



for an image here...



The Art Institute of Chicago: Art Access

The artist visited La Grande Jatte many times, making drawings and more than ... With his

www.artic.edu/artaccess/AA Impressionist/pages/IMP 7.shtml - 14k - Cached - Similar pages

precise method and technique, Seurat conceived of his painting as ...

Many ways to ask about a painting... many ways to respond...

Query Terms	OneBox	First Google search result
georges seurat "la grande jatte" georges seurat la grande jatte "la grande jatte" la grand jatte george seurat, la grande jatte george seurat "la grande jatte" painting la grand jatte "la grande jatte by georges seurat" george seurat la grande jatte georges seurat painting	None None None None None None None None	The Art Institute of Chicago: Art Access The Art Institute of Chicago: Art Access Seurat, A Sunday Afternoon on the Island of La Grande Jatte The Art Institute of Chicago: Art Access WebMuseum: Seurat, Georges The Art Institute of Chicago: Art Access The Art Institute of Chicago: Art Access Sunday Afternoon on the Island of La Grande Jatte Posters by Webmuseum: Seurat, Georges Webmuseum: Seurat, Georges
la grande jatte	Image	The Art Institute of Chicago: Art Access
la grande jatte georges la grande jatte by georges seurat georges seurat painting grande jatte la grande jatte painting painting la grand jatte seurat	Product Product Product Product Product	The art institute of Chicago Webmuseum: Seurat, Georges The Art Institute of Chicago: Art Access Seurat, A Sunday after noon on the island The Art Institute of Chicago: Art Access
seurat la grande jatte <mark>pic</mark> la grande jatte by george seurat seurat la grande jatte <mark>image</mark>	Book Book Book	FlickrBlog Webmuseum: Seurat, Georges Webmuseum: Seurat, Georges
La Grande Jatte by Georges Seurat painting	Scholar	The Art Institute of Chicago: Art Access



...with many OneBoxes...



News Froogle Maps

la grande jatte by george seurat



News Froogle Images Groups La Grande Jatte by Georges Seurat painting Scholar Search

Prefere

Web

Book results for la grande jatte by george seurat



Enchanted Evenings - by Geoffrey Block - 410 pages
Painting and Sculpture in Europe 1880-1940 - by George Heard Hamilton

WebMuseum: Seurat, Georges

His famous canvas Sunday Afternoon on the Island of the Grande Jatte was the of an exhibition in 1886. By then Seurat was spending his winters ... www.ibiblio.org/wm/paint/auth/seurat/ - 7k - Cached - Similar pages

Web

Scholarly articles for La Grande Jatte by Georges Seurat painting



On Concurrent Programming - Schneider - Cited by 55 Chapter 13 Selfobject Theory and the Artistic Process - Rotenberg - Cited by 8

The Art Institute of Chicago: Art Access

The artist visited La Grande Jatte many times, making drawings and more than ... With his precise method and technique. Seurat conceived of his painting as ... www.artic.edu/artaccess/AA Impressionist/pages/IMP 7.shtml - 14k - Cached - Similar pages



Froogle Groups News

la grande jatte



Images Groups la grande jatte georges

News Froogle Maps

Scholar

Web

Image results for la grande jatte







Web

Product search results for la grande jatte georges



Georges Seurat Sunday Afternoon Island La Grande Jatte - \$2.99 - eBay "Sunday Afternoon on la Grande Jatte," Georges Seurat ... - \$1.99 - EaselWeasel.com Sunday Afternoon on the Island of La Grande Jatte by ... - \$419.99 - Target

The Art Institute of Chicago: Art Access

Georges Seurat, French, 1859-1891 ... The artist visited La Grande Jatte many times, making drawings and more than 30 oil sketches to prepare for the final ... www.artic.edu/artaccess/AA Impressionist/pages/IMP 7.shtml - 14k - Cached - Similar pages

The Art Institute of Chicago: Art Access

A Sunday on La Grande Jatte-1884, 1884-86. Oil on canvas, 207.5 x 308 cm ... The artist visited La Grande Jatte many times, making drawings and more than 30 ... www.artic.edu/artaccess/AA Impressionist/pages/IMP 7.shtml - 14k - Cached - Similar pages

47

What's the mental model of oneboxen?

It's magic:

"... I don't know how to make it come back..."

"... why does this... thing... keep being at the top?"













Mental model 2: We are all cargo cultists







Mental model 3

- 1. Predictable behavior Can I predict what will happen when I do X?
- 2. How is content indexed? Is it full-text? How are images indexed?.....
- 3. How does Google look it up? Which keywords should I pick?
- 4. How are the results ranked? What does the order mean?
- 5. What's in the index? What kinds of documents can I search?



Culture of search 1

What does it mean to Google something?

"...let me google this on Yahoo..."

How does always available search change your conversations?





Culture of search 2

Has ubiquitous search changed expectations about knowledge?

Type 1: **Eternal verities** (F = ma; Antarctica is a continent)

Type 2: Mid-term (Sacramento is the capital of California; there are 117 elements)

Type 3: **Ephemera** (my IP address is 9.1.2.142; use Muggle.google.com for your proxy)



Culture of search 3

- Key question: What do you really need to know?
 - recognition knowledge?
 - contextual knowledge?
- When the cost of discovery and access is low... does that change your expectation of others?



A new literacy

Four kinds of knowledge medical knowledge & skills needed to search: plumbing knowledge ...etc... site: ricoh.com "double quotes" minus (as exclude) domain pure engine plus (include) technique knowledge filetype:pdf intitle:"cheat sheet" ... etc ... reverse dictionary keyword frequencies contents of domains information search Wikipedia mapping strategy ... etc... knowing when to shift knowing when to stop move from wide to

narrow; preserving

state; etc...

Google⁻

Co-evolution

Search engines will continue to change

- change is constant...
 new document types, new searches, new possibilities
- that's the point of all our studies / testing
- things will continue to change rapidly

- Search engines need to match capabilities with user expectations and understandable user mental models
 - → need to continually refine understanding of user population's mental models
 - → need to detect when a particular model is in play



Shared responsibility

For search engines:

- To create a system that behaves predictably
- To understand expectations of entire breadth of users

For our users:

- To learn the basics of how search engines work
- To have a functional mental model



For the digital library community:

- To educate our users in broadly effective models of

research content organization

... and how to evolve...



END