

Welcome to



CHI 2015

CROSSINGS

SEOUL • KOREA

ACM-W Athena Lecturer

Athena is the Greek goddess of wisdom, courage, inspiration, civilization, law and justice, mathematics, strength, war strategy, the arts, crafts, and skill. The Athena Lecturer Award celebrates women researchers who have made fundamental contributions to Computer Science.

Susan Dumais

Large-Scale Behavioral Data:
Potential and Pitfalls





MORE ACM AWARD
WINNERS



ACM-W Athena Lecturer Award

LARGE-SCALE BEHAVIORAL DATA: POTENTIAL AND PITFALLS

Apr 23, 2015


Susan Dumais, Microsoft Research

Overview

- Large-scale behavioral log data
 - ▣ Web-based services enable us to capture traces of human behavior *in situ* at a scale previously unimaginable
 - ▣ Transformed how web-based systems are designed, evaluated and improved
- Opportunities
 - ▣ Examples from web search
 - ▣ Observations: Understand behavior
 - ▣ Experiments: Improve systems and services
- Challenges and limitations

20 Years Ago ... In Web and Search

- ❑ NCSA Mosaic graphical browser two years old
- ❑ CHI 1995 online presence




CHI 95 was the 1995 Conference on Human Factors in Computing Systems and was sponsored by [ACM/SIGCHI](#). CHI '95 was held in Denver, Colorado, USA (May 7 - 11, 1995). [CHI 96](#) will take place in Vancouver, Canada (April 14-18, 1996).

Do you have complaints, comments, suggestions or praise for something about CHI '95? We want to hear from you. Send email to chi95-feedback@sigchi.acm.org.

What's New?


- 95-12-06 The proceedings are back, re-designed but still not completely done.

Contents

 [Electronic Proceedings](#)

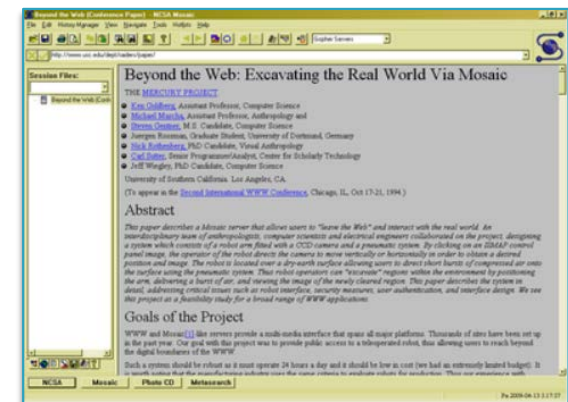
CHI '95 CDROM

The conference proceedings are being re-designed. The content of the papers and such is finished, but not all of the "surrounding" aspects are finished yet.

 [Advance Program](#)

Quick jumps into interesting parts of the Advance Program:

- [Table of Contents](#)
- [Conference-at-a-Glance](#) (requires graphical browser)
- [Technical Program Schedule](#)
- [Housing Reservation Information](#)
- [Conference Registration Information](#)



20 Years Ago ... In Web and Search

- ❑ NCSA Mosaic graphical browser two years old
- ❑ CHI 1995 online presence
- ❑ Size of the web
 - # web sites: 2.7k
- ❑ Size of Lycos search engine
 - # web pages in index: 54k
- ❑ Behavioral logs
 - # queries/day: 1.5k
 - Most search and logging client-side



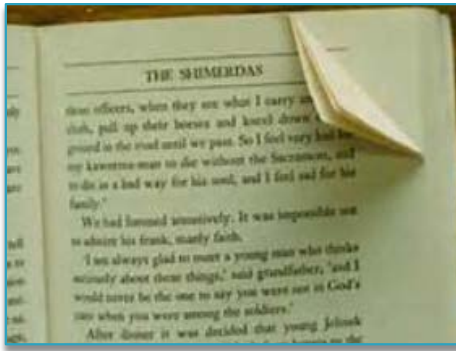
• [Top 5% Sites](#) •

Today ... Search is Everywhere

- A billion web sites
- Trillions of pages indexed by search engines
- Billions of web searches and clicks per day
- Search is a core fabric of everyday life
 - ▣ Diversity of tasks and searchers
 - ▣ Pervasive (web, desktop, enterprise, apps, etc.)
- Understanding and supporting searchers more important now than ever before

What Are Behavioral Logs?

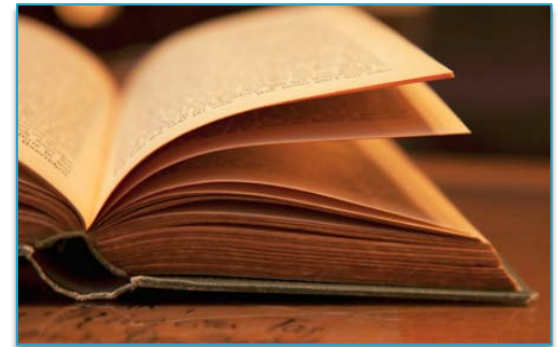
- Traces of human behavior
 - ▣ ... seen through the lenses of whatever sensors we have



incarnate—a rare pleasure in our age of etheralization, when all that is solid is melting into zeroes and ones.

In a Screen Age, the eye is glutted and the sense of touch starved. The electronic book robs us of the erotic of paper. Sure, an audio clip could emulate the sound of turning pages, just as a screen could impersonate a specific copy of a book—J. Edgar Hoover's *Lolita*, say, replete with obscene marginalia (I'm making this up)—but never its feel. Smart as it is, electronic paper can't learn, by which I mean it can't wrinkle at the touch of wet fingers turning pages in the bathtub; can't remember the stained ring of that glass of red wine you imprudently used to hold your place; can't speak volumes, from its margins and endpapers, about everyone who has ever totted a thought in it. Implicit in the possession of a book is the history of a book's present readings—that is to say, its life.

yes



Chapter 2 Relativity II

$$KE = E_k = \int_{u=0}^u \mathbf{F} \cdot d\mathbf{u} = \int_{u=0}^u \frac{d\mathbf{p}}{dt} \cdot d\mathbf{u} = \int_{u=0}^u \frac{d\mathbf{p}}{dt} \cdot \frac{d\mathbf{u}}{dt} dt = \int_{u=0}^u \frac{d\mathbf{p}}{dt} \cdot \mathbf{u} dt$$

Using a suitable low-velocity approximation of the integral in this equation is not difficult but requires a bit of physics. It is left as an exercise (Problem 2-2) to show that

$$\frac{d\mathbf{p}}{dt} \cdot \mathbf{u} = \frac{d}{dt} \left[\frac{1}{2} m \mathbf{u} \cdot \mathbf{u} \right] = m \mathbf{u} \cdot \frac{d\mathbf{u}}{dt}$$

Substituting this into the integral in Equation 2-8, we obtain

$$KE = \int_{u=0}^u m \mathbf{u} \cdot \frac{d\mathbf{u}}{dt} dt = \int_{u=0}^u m \mathbf{u} \cdot d\mathbf{u} = \frac{1}{2} m \mathbf{u} \cdot \mathbf{u} = \frac{1}{2} m u^2$$

Equation 2-9 defines the relativistic kinetic energy. Notice that, as we wanted earlier, E_k is not $\frac{1}{2} m u^2$ or even $\frac{1}{2} m v^2$. This is strikingly evident in Figure 2-3. However, agreement with our second condition on the relativistic total energy E , Equation 2-6, does approach $m_0 c^2$ when $u \ll c$. We can check this assertion by noting that for $u \ll c$, expanding γ by the binomial theorem yields

$$\gamma = \left(1 - \frac{u^2}{c^2} \right)^{-1/2} \approx 1 + \frac{1}{2} \frac{u^2}{c^2} + \frac{3}{8} \frac{u^4}{c^4} + \dots$$

The expression for kinetic energy in Equation 2-9 consists of two terms. One term, $m_0 c^2$, depends on the speed of the particle through the factor γ , and the other term, $\frac{1}{2} m_0 u^2$, is independent of the speed. The quantity $m_0 c^2$ is called the rest energy of the particle, and the quantity associated with the second term is the relativistic kinetic energy. It is often defined as the sum of the kinetic energy and the rest energy:

$$E = \gamma m_0 c^2 = m_0 c^2 + KE$$

Total energy $E = \gamma m_0 c^2 = m_0 c^2 + KE$

Thus, the work done by a net force increases the energy of the system from the rest energy $m_0 c^2$ to $\gamma m_0 c^2$ because the total energy from $m_0 c^2$.

For a particle at rest relative to an observer, $\mathbf{u} = 0$, Equation 2-10 becomes $E = m_0 c^2$. What a coincidence! Equation 2-10 can be written as

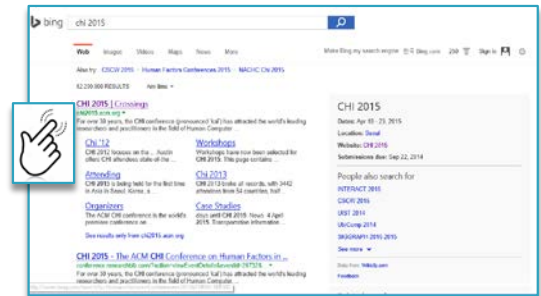
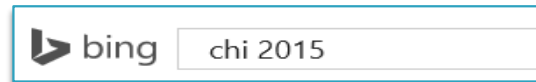
$$E = \gamma m_0 c^2 = m_0 c^2 + KE = \left\{ \begin{array}{l} \text{rest energy} \\ + \\ \text{kinetic energy} \end{array} \right\}$$

Before the advent of relativity theory, it was thought that mass was a conserved quantity; consequently, it would always be the same before and after an interaction.



What Are Behavioral Logs?

- Traces of human behavior
 - ▣ ... seen through the lenses of whatever sensors we have
 - ▣ Web search: queries, results, clicks, dwell time, etc.



- Actual, real-world (*in situ*) behavior
 - ▣ Not ...
 - Recalled behavior
 - Subjective impressions of behavior
 - Controlled experimental task

How Are Behavioral Logs Used?

Edit wear and read wear. [Hill, Hollan, et al., CHI 1992]



InterTwine, Fourney et al.,
UIST 2014

Figure showing a screenshot of a Google search for "gimp black and white with col...". The search results include a link to "GIMP - Selective Colorization" with a description: "Giving credit where credit is due: I did not come up with GIMP from a reader comment I saw in a 'hand-coloring'". Below the link are two images labeled "before" and "after" showing a butterfly. The search also shows "GIMP - Converting Color Images to B&W" with a link to a tutorial. A sidebar on the right shows "History" and "Interests".

Figure showing a screenshot of a Bing search for "chi 2015". The search results include a list of suggestions: "chi 2015 >", "chi 2015 conference", "chi 2015 workshops", "chi 2015 submission", "chi 2015 twitter", "chi 2015 rebuttal", "chi 2015 subcommittee", and "chi 2015 student design competition". The search history shows "CHI 2015 | Crossings" with the URL "http://chi2015.acm.org/". The search also shows "Browse search history for 'chi 2015'". The Bing logo and "MS Beta" are visible in the top left corner.

Figure showing a screenshot of a Bing History page. The page displays a search bar for "Search your history" and a list of search results. The results include "CHI 2015 | Crossings" and "Meeting Planner - Find best time acro...". The page also shows "YOUR SEARCH: meeting planner time zones" and "Taiwan earthquake 2015". The Bing logo and "History" are visible in the top left corner.

Kinds of Behavioral Data

Lab Studies

In lab, controlled tasks, with detailed instrumentation and interaction

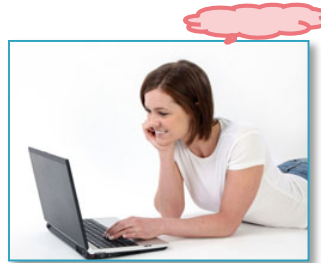


- 10-100s of people (and tasks)
- Known tasks, carefully controlled
- Detailed information: video, gaze-tracking, think-aloud protocols
- Can evaluate experimental systems

Kinds of Behavioral Data

Panel Studies

*In the wild, real-world tasks,
ability to probe for detail*



- 100-1000s of people (and tasks)
- In-the-wild
- Special client instrumentation
- Can probe about specific tasks, successes/failures

Kinds of Behavioral Data

Log Studies

In the wild, no explicit feedback but lots of implicit feedback



- ❑ Millions of people (& tasks)
- ❑ In-the-wild
- ❑ Diversity and dynamics
- ❑ Abundance of data, but it's noisy and unlabeled

Kinds of Behavioral Data

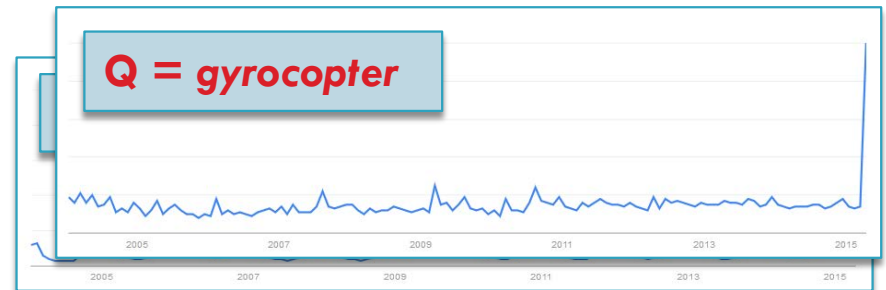
	Observational	Experimental
Lab Studies <i>Controlled tasks, in laboratory, with detailed instrumentation</i>	In-lab behavior observations	In-lab controlled tasks, comparisons of systems
Panel Studies <i>In the wild, real-world tasks, ability to probe for detail</i>	Ethnography, case studies, panels (e.g., Nielsen)	Clinical trials and field tests
Log Studies <i>In the wild, no explicit feedback but lots of implicit feedback</i>	Logs from a single system	A/B testing of alternative systems or algorithms

Goal: Build an abstract picture of behavior

Goal: Decide if one approach is better than another

Benefits of Behavioral Logs

- Real-world
 - ▣ Portrait of real behavior
- Large-scale
 - ▣ Millions of people and tasks
 - ▣ Rare behaviors are common
 - ▣ Small differences can be measured
 - ▣ Tremendous diversity of behaviors and information needs (the “long tail”)
- Real-time
 - ▣ Feedback is immediate



Behavioral Logs and Web Search

□ How do you go from 2.4 words to anything sensible?

■ Content

- Match (query, page content)

■ Link structure

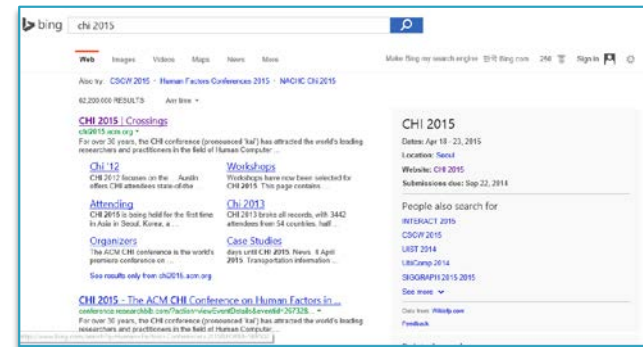
- Used to set non-uniform priors on pages

■ User behavior

- Anchor text
- Query-clicks
- Query reformulations

■ Contextual metadata

- Who, what, where, when, ...



Driven by ...
behavioral data

□ Aggregate behavioral data used to improve search algorithms and experiences

Surprises In (Early) Web Search Logs

- Early web search log analysis
 - ▣ Silverstein et al. 1998, Broder 2002
- Web search \neq library search
 - ▣ Queries are very short, 2.4 words
 - ▣ Lots of people search for sex
 - ▣ “Navigating” is common, 30-40%
 - Getting to web sites vs. finding out about things
 - ▣ Queries are not independent, e.g., tasks
 - ▣ Amazing diversity of information needs (“long tail”)

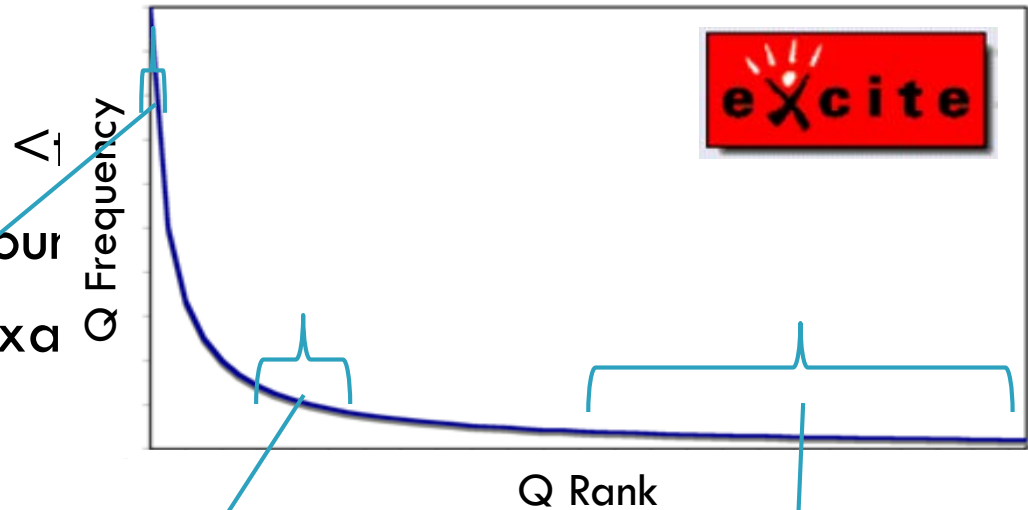


Queries Are Not Equally Likely

Excite 1999 data

- ~2.5 mil queries
- Head: top 250 accounts
- Tail: ~950k occur exactly once

Zipf Distribution



Top 10 Q

- sex
- hotmail
- yahoo
- games
- chat
- mp3
- horoscope
- weather
- pokemon
- ebay

Navigational queries, one-word queries

Query Freq = 10

- antivirus AND download
- sony playstation cheat codes
- email clipart
- no bake cookies
- interior design schools

Multi-word queries, specific URLs

Query Freq = 1

- acm98
- winsock 1.1 y2k compliant
- how do you say father and son in korean
- email address for paul allen the seattle seahawks owner

Complex queries, rare info needs, misspellings, URLs

Queries Vary Over Time and Task

□ Time

- Periodicities

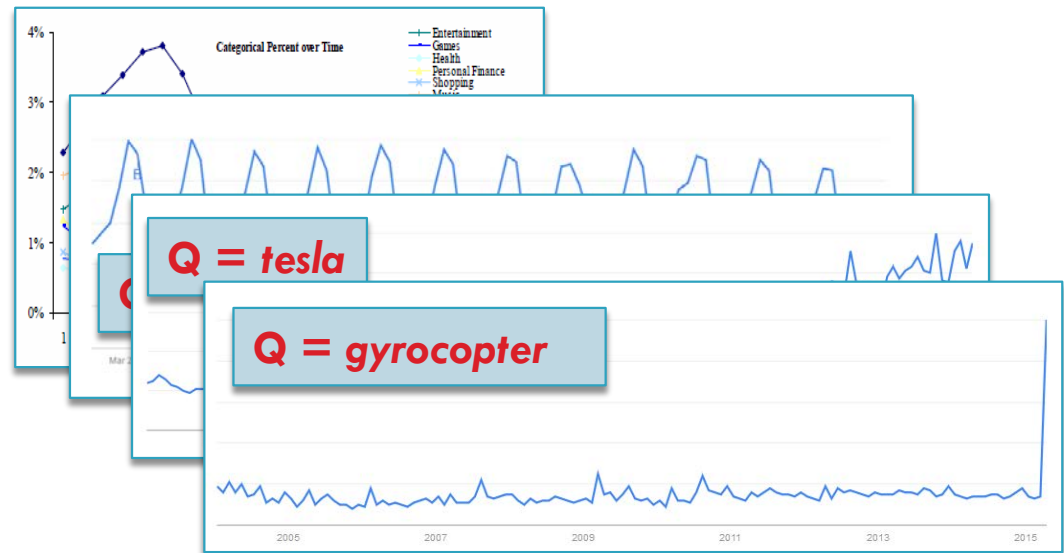
- Trends

- Events

□ Task/Individual

- Sessions

- Longer history



(Q=ACM Awards | Susan vs. Garth)



Query	Time	UserID
chi 2015	10:41 am 4/15/15	142039
social science	10:44 am 4/15/15	142039
computational social science	10:56 am 4/15/15	142039
chi 2015	11:21 am 4/15/15	659327
intercontinental hotel seoul	11:59 am 4/15/15	659327
restaurants seattle	12:01 pm 4/15/15	318222
pikes market restaurants	12:17 pm 4/15/15	318222
stuart shulman	12:18 pm 4/15/15	142039
daytrips in seattle, wa	1:30 pm 4/15/15	554320
chi 2015	2:30 pm 4/15/15	435451
chi 2015 program	2:32 pm 4/15/15	435451
chi registration times	2:42 pm 4/15/15	435451
computational social science	4:56 pm 4/15/15	142039
chi 2015	5:02 pm 4/15/15	312055
xxx clubs in seattle	10:14 pm 4/15/15	142039
sex videos	1:49 am 4/16/15	142039

Query	Time	UserID
chi 2015	10:41 am 4/15/15	142039
social science	10:44 am 4/15/15	142039
computational social science	10:56 am 4/15/15	142039
chi 2015	11:21 am 4/15/15	659327
intercontinental hotel seoul	11:59 am 4/15/15	659327
restaurants seattle	12:01 pm 4/15/15	318222
pikes market restaurants	12:17 pm 4/15/15	318222
stuart shulman	12:18 pm 4/15/15	142039
daytrips in seattle, wa	1:30 pm 4/15/15	554320
chi 2015	2:30 pm 4/15/15	435451
chi 2015 program	2:32 pm 4/15/15	435451
chi registration times	2:42 pm 4/15/15	435451
computational social science	4:56 pm 4/15/15	142039
chi 2015	5:02 pm 4/15/15	312055
xxx clubs in seattle	10:14 pm 4/15/15	142039
sex videos	1:49 am 4/16/15	142039

Query typology
E.g., “navigational queries”

Query	Time	UserID
chi 2015	10:41 am 4/15/15	142039
social science	10:44 am 4/15/15	142039
computational social science	10:56 am 4/15/15	142039
chi 2015	11:21 am 4/15/15	659327
intercontinental hotel seoul	11:59 am 4/15/15	659327
restaurants seattle	12:01 pm 4/15/15	318222
pikes market restaurants	12:17 pm 4/15/15	318222
stuart shulman	12:18 pm 4/15/15	142039
daytrips in seattle, wa	1:30 pm 4/15/15	554320
chi 2015	2:30 pm 4/15/15	435451
chi 2015 program	2:32 pm 4/15/15	435451
chi registration times	2:42 pm 4/15/15	435451
computational social science	4:56 pm 4/15/15	142039
chi 2015	5:02 pm 4/15/15	312055
xxx clubs in seattle	10:14 pm 4/15/15	142039
sex videos	1:49 am 4/16/15	142039

Query typology
E.g., “navigational queries”

Query behavior
E.g. “common Q”

Query	Time	UserID
chi 2015	10:41 am 4/15/15	142039
social science	10:44 am 4/15/15	142039
computational social science	10:56 am 4/15/15	142039
chi 2015	11:21 am 4/15/15	659327
hilton san francisco	11:59 am 4/15/15	659327
restaurants seattle	12:01 pm 4/15/15	318222
pikes market restaurants	12:17 pm 4/15/15	318222
stuart shulman	12:18 pm 4/15/15	142039
daytrips in seattle, wa	1:30 pm 4/15/15	554320
chi 2015	2:30 pm 4/15/15	435451
chi 2015 program	2:32 pm 4/15/15	435451
chi registration times	2:42 pm 4/15/15	435451
computational social science	4:56 pm 4/15/15	142039
chi 2015	5:02 pm 4/15/15	312055
xxx clubs in seattle	10:14 pm 4/15/15	142039
sex videos	1:49 am 4/16/15	142039

Query typology
E.g., “navigational queries”

Query behavior
E.g. “common Q”

Long-term trends
E.g. “repeat Q or topic”

Query	Time	UserID
chi 2015	10:41 am 4/15/15	142039
social science	10:44 am 4/15/15	142039
computational social science	10:56 am 4/15/15	142039
chi 2015	11:21 am 4/15/15	659327
hilton san francisco	11:59 am 4/15/15	659327
restaurants seattle	12:01 pm 4/15/15	318222
pikes market restaurants	12:17 pm 4/15/15	318222
stuart shulman	12:18 pm 4/15/15	142039
daytrips in seattle, wa	1:30 pm 4/15/15	554320
chi 2015	2:30 pm 4/15/15	435451
chi 2015 program	2:32 pm 4/15/15	435451
chi registration times	2:42 pm 4/15/15	435451
computational social science	4:56 pm 4/15/15	142039
chi 2015	5:02 pm 4/15/15	312055
xxx clubs in seattle	10:14 pm 4/15/15	142039
sex videos	1:49 am 4/16/15	142039

Query typology
E.g., “navigational queries”

Query behavior
E.g. “common Q”

Long-term trends
E.g. “repeat Q or topic”

Short-term task
E.g. “get organized for chi”

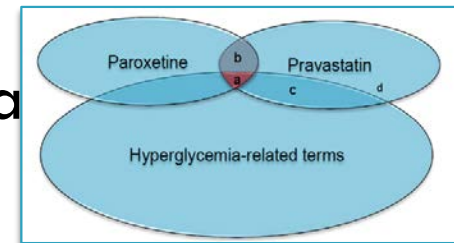
Query	Time	UserID
chi 2015	10:41 am 4/15/15	142039
social science	10:44 am 4/15/15	142039
computational social science	10:56 am 4/15/15	142039
chi 2015	11:21 am 4/15/15	659327
intercontinental hotel seoul	11:59 am 4/15/15	659327
restaurants seattle	12:01 pm 4/15/15	318222
pikes market restaurants	12:17 pm 4/15/15	318222
stuart shulman	12:18 pm 4/15/15	142039
daytrips in seattle, wa	1:30 pm 4/15/15	554320
chi 2015	2:30 pm 4/15/15	435451
chi 2015 program	2:32 pm 4/15/15	435451
chi registration times	2:42 pm 4/15/15	435451
computational social science	4:56 pm 4/15/15	142039
chi 2015	5:02 pm 4/15/15	312055
xxx clubs in seattle	10:14 pm 4/15/15	142039
sex videos	1:49 am 4/16/15	142039

Uses of Analysis

- **Q typology**
 - ▣ E.g., ranking
- **Q frequency**
 - ▣ E.g., caching
- **Repeat Q**
 - ▣ E.g., history
- **Sessions**
 - ▣ E.g., ranking
- **Test set development**
- **Complementary research**

Beyond Improving Web Search

- Search for health information common and important
 - ▣ 80% of U.S. adults use web search to find medical info
 - ▣ 1 in 250 people query about the top-100 prescription drugs
- Mining health search data to identify adverse drug effects and drug interactions
 - ▣ Today detection based on reports from patients
 - ▣ Case study of Paroxetine and Pravastatin
 - 2011 report: Paroxetine + Pravastatin => Hyperglycemia
 - Pre-2011 search logs: Increased use of terms related to hyperglycemia (e.g., thirsty, appetite increase, frequent urination, high blood sugar)
 - ▣ Search logs can increase speed and scale of detection



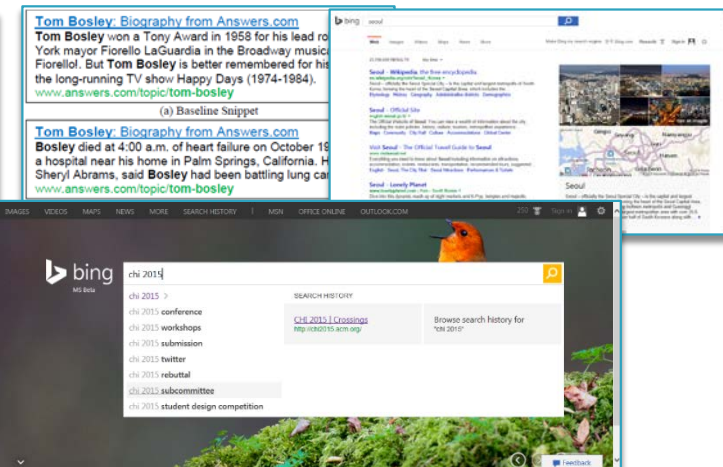
Value of Observational Logs

- Examine search behavior in natural settings
 - ▣ Description: *What do people currently do?*
 - ▣ Prediction: *What will people do in similar situations?*
- Provide insights about how people search at scale
 - ▣ Identify important problems to study
 - ▣ Enable more realistic simulations and evaluations
- Motivate ideas for how to improve search system
 - ▣ E.g. ranking algorithms, presentation, interaction
- Provide a potentially powerful lens for scientific inquiry in medical and social domains

From Observations to Experiments

- Observations generate insights about behavior and ideas for improvements
- Experiments are the way to systematically improve
 - ▣ Controlled experiments to compare system variants
 - ▣ Used to study all aspects of web search systems

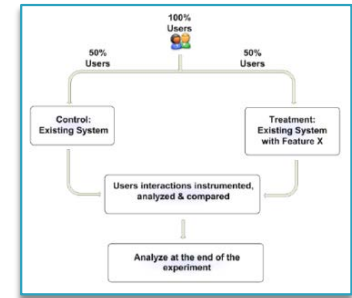
- System latency
- Ranking algorithms
- Snippet generation
- Spelling and query suggestions
- Presentation



- Data- vs. HIPPO-driven design

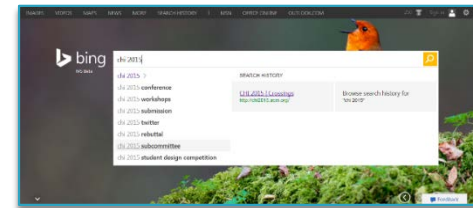
Experiments At Web Scale

- Basic questions
 - ▣ What do you want to evaluate?
 - ▣ What metric(s) do you care about?
- Within- vs. between-subject designs
 - ▣ Within: Interleaving (for ranking changes); otherwise temporal split between experimental and control conditions
 - ▣ Between: More widely useful, but higher variance
- Some things easier to study than others
 - ▣ Algorithmic vs. Interface vs. Social Systems
- Counterfactuals, Power, and Ramping-Up important



Value of Behavioral Logs

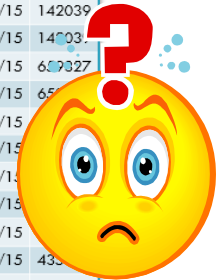
- Provide (often surprising) insights about how people interact with existing search systems
 - ▣ Focus efforts on supporting actual (vs. presumed) activities
 - E.g., Diversity of tasks, searchers, contexts of use, etc.
 - ▣ Suggest experiments about important or unexpected behaviors
 - ▣ Used to develop predictive models and evaluation sets
 - ▣ Support new search experiences
- Improve system performance
 - ▣ Caching, ranking algorithms, presentation, interaction, etc.
- Changed how systems are designed, evaluated, improved



What Logs Alone Cannot Tell Us

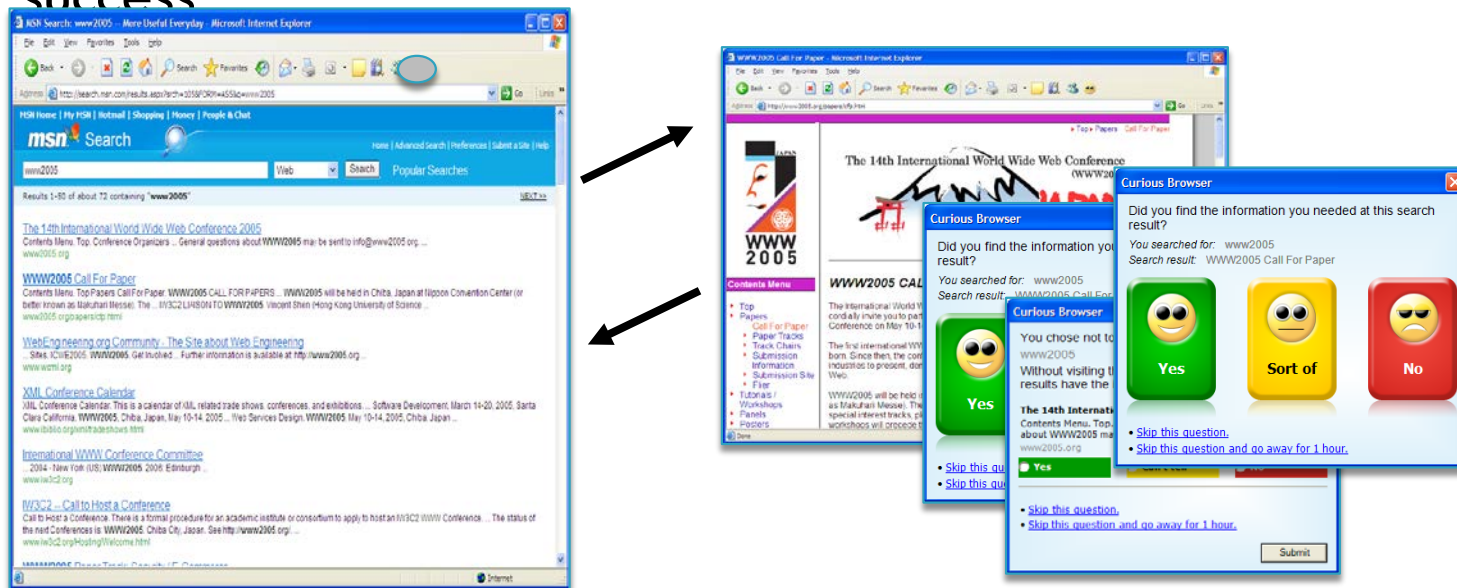
- ❑ Limited annotations
 - ▣ People's intent
 - ▣ People's success
 - ▣ People's experience
 - ▣ People's attention
- ❑ Behavior can mean many things
- ❑ Limited to existing systems and interactions
- ❑ Logs convey “what” people are doing, but not “why”
- ❑ Complement with other techniques to provide a more complete picture (e.g., lab, panel studies, modeling)

Query	Time	UserID
chi 2015	10:41 am 4/15/15	142039
social science	10:44 am 4/15/15	142039
computational social science	10:56 am 4/15/15	142039
chi 2015	11:21 am 4/15/15	607927
intercontinental hotel seoul	11:59 am 4/15/15	657927
restaurants seattle	12:01 pm 4/15/15	142039
pikes market restaurants	12:17 pm 4/15/15	142039
stuart shulman	12:18 pm 4/15/15	142039
daytrips in seattle, wa	1:30 pm 4/15/15	142039
chi 2015	2:30 pm 4/15/15	142039
chi 2015 program	2:32 pm 4/15/15	435451
chi registration times	2:42 pm 4/15/15	435451
computational social science	4:56 pm 4/15/15	142039
chi 2015	5:02 pm 4/15/15	312055
xxx clubs in seattle	10:14 pm 4/15/15	142039
sex videos	1:49 am 4/16/15	142039



Example: Relevance

- Curious Browser – linking implicit and explicit
 - ▣ Capture many *implicit actions* (e.g., query, click, dwell time, scroll)
 - ▣ Probe for *explicit judgments* of page relevance and session success



Example: Relevance

- Curious Browser – linking implicit and explicit
 - ▣ Capture many *implicit actions* (e.g., query, click, dwell time, scroll)
 - ▣ Probe for *explicit judgments* of page relevance and session success
- Learned models to predict explicit judgments (page relevance, session success) from implicit indicators
 - ▣ Relevance of web page to query
 - Click => 45% accuracy in predicting relevance
 - Click + Dwell time + Session end => 75% accuracy
 - ▣ Session success
 - Very highly correlated with clicks rated at “relevant”
... but misses effort, expectations, delight
- Use learned models to predict relevance, success

Example: Abandonment

- So, no clicks is a bad sign, right?
... it depends
- Increasingly search engines provide “answers” vs. 10-blue links
- Retrospective survey
- *In situ* abandonment survey (implicit-explicit)
- Predictive model for good vs. bad abandonment

1 US Dollar equals
0.94 Euros
1 EUR equals 1.07 USD - MSN Money

Weather in Seattle, Washington
msn.com/weather Data from WDT

Now Thu Fri Sat Sun Mon Next 5 days

44°
Mostly Sunny

°F / °C 62° / 41° 59° / 43° 54° / 42° 55° / 42° 56° / 43°

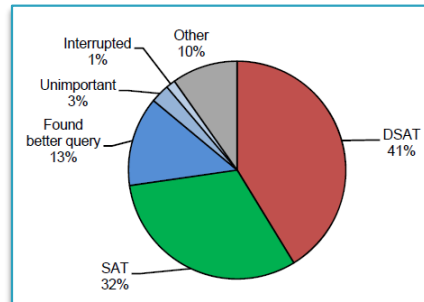
★ Stay updated on this location's weather

Perfect Baked Potatoes - What's Cooking America
whatscookingamerica.net/Q-A/PotatoBaking.htm ▾

Perfect Baked Potato Recipe: Recipe Type: Potatoes, Vegetables Yields: serves many

Prep time: 10 min Bake time: 60 min Ingredients: Baking Potatoes

[Potato Recipes](#) · [Hints & Tips](#) · [Coarse Or Sea Salt](#) · [Cooking Articles](#) · [Olive Oil](#)



Query Abandonment Survey

You abandoned your search query:
maui weather

Why did you not click on the search results?

I found what I was looking for on the search page

...in a direct answer (stock quote, weather, map, definition, spelling correction, etc.)

...in the summary of a search result

...somewhere else

I was dissatisfied with the results

I got interrupted or had more important things to do

Other

Done

Ignore now Ignore for 1hr

Example: Re-Finding

- Observational log analysis
 - 33% of queries are repeat queries
 - 39% of clicks are repeat clicks
- Many are navigational queries
 - E.g., *chi 2015* -> chi2015.acm.org
- “Personal” navigational queries
 - Different intents across individuals, but same intent for an individual
 - E.g., ACM Awards (for S.Dumais) -> awards.acm.org
 - E.g., ACM Awards (for G.Brooks) -> acmcountry.com
 - Simple accurate personalization algorithm
- A/B experiments show benefits

		Repeat Click	New Click
Repeat Query	33%	29%	4%
New Query	67%	10%	57%
		39%	61%

Example: Sessions and Tasks

- Queries do not occur in isolation
 - ▣ 60% of search sessions contain multiple queries
 - ▣ 50% of search time spent in sessions of 30+ mins
- Model this query context
 - ▣ Short-term session activities (queries, clicks)
 - ▣ Long-term preferences and interests
- A/B experiments show improvements
- But ... broader support for accomplishing tasks still lacking



Uniqueness of Behavioral Logs

Lab Studies

Controlled tasks, in laboratory, with detailed instrumentation

Panel Studies

In the wild, real-world tasks, ability to probe for detail

Log Studies

In the wild, no explicit feedback but lots of implicit feedback

Behavioral search logs uniquely

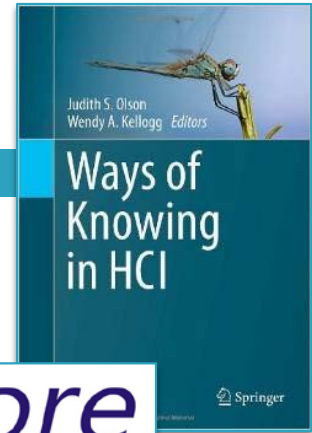
- ❑ Real-world: Represent natural task distribution
- ❑ Large-scale: Highlight diversity of tasks, queries and searcher interaction
- ❑ Real-time: Interaction immediately reflects things in the world

Logs from a single system

A/B testing of alternative systems or algorithms

Summary

- Large-scale behavioral log data
 - ▣ Provide traces of human behavior *in situ* at a scale and fidelity previously unimaginable
- Changed how web-based systems are designed, evaluated and improved
 - ▣ Observations: enable us to characterize behavior
 - ▣ Experiments: used to improve web search
 - ▣ More important now than ever before
- CHI community should lead in shaping best practices in behavioral log studies



□ Thank you!

□ More info at:

<http://research.microsoft.com/~sdumais>

References

- White et al. (2014). Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clinical Pharmacology & Therapeutics*.
- White et al. (2013). Web-scale pharmacovigilance: Listening to signals from the crowd. *Journal of the American Medical Informatics Association*.
- Kohavi et al. (2009). Controlled experiments on the web: Survey and Practical guide. *Data Mining and Knowledge Discovery*.
- Dumais et al. (2014). Understanding user behavior through log data analysis. In Oldon and Kellogg (Eds), *Ways of Knowing in HCI*.
- Fox et al. (2005). Evaluating implicit measures to improve web search. *ACM:TOIS*.
- Diriye et al. (2012). Leaving so soon? Understanding and predicting web search abandonment rationales. *CIKM 2012*
- Teevan et al. (2007). Information re-retrieval: Repeat queries in Yahoo's logs. *SIGIR 2007*.
- Tyler & Teevan (2010). Large-scale query log analysis of re-findings. *WSDM 2010*.
- Bennett et al., (2012). Modeling the impact of short- and long-term behavior on search personalization. *SIGIR 2012*.