

Online and Offline Evaluation of Search Engine Quality

Evangelos Kanoulas



UNIVERSITY OF AMSTERDAM

Different approaches to evaluation

- User-studies
- Collection-based evaluation
- In-situ evaluation
 - A/B Testing
 - Interleaving

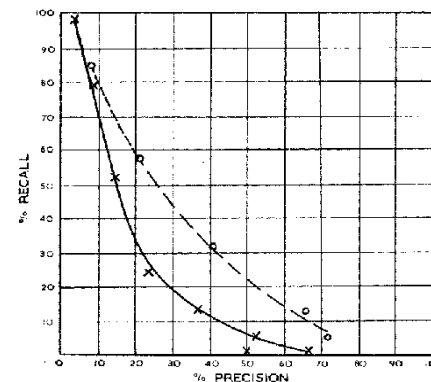
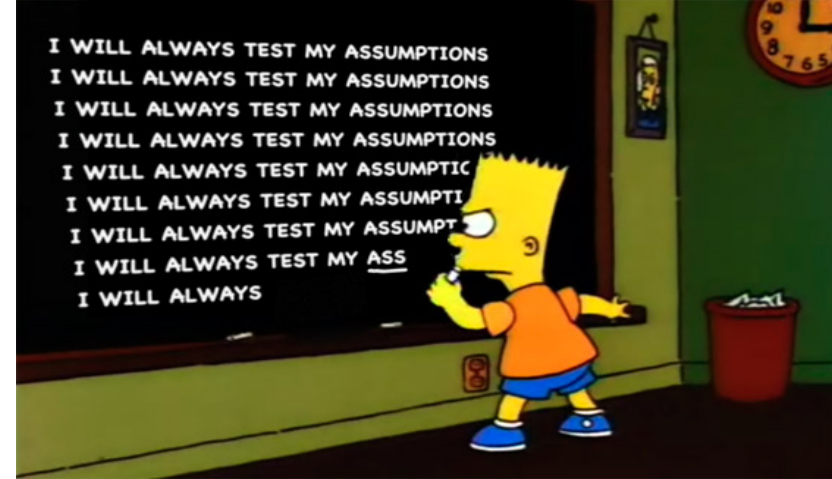


FIGURE 4.814P INDEX LANGUAGE III, 5, 8 SEARCH E
200 DOCUMENTS
(Index 2-language III, 1, 2 Broken line)



Cranfield
UNIVERSITY

Outline

PART I

1. Collection-based Evaluation
2. Comparative Evaluation

PART II

3. Online User Behavior
4. A/B Testing
5. Interleaving
6. Comparative Studies

3. Online User Behavior

Offline vs. Online Assumptions

- Basic assumptions:
 - Offline:
 - **assessors** can tell you what is **relevant**
 - Online:
 - **online user behavior** can tell you what is **relevant**

Online User Behavior

- **Key assumption:** observable user behavior reflects relevance
- Implicit in this: Users behave (somewhat) rationally
 - Real users have a goal when they use an IR system
 - They aren't just bored, typing and clicking pseudo-randomly
 - They consistently work towards that goal
 - An non-relevant result doesn't draw most users away from their goal
 - They aren't trying to confuse you
 - Most users are not trying to provide malicious data to the system

Online User Behavior

- This assumption gives us “high fidelity”
 - Real users replace the judges;
 - No ambiguity in information need;
 - Users actually want results;
 - Measure performance on real queries
- But introduces a major challenge
 - We can't train the users
 - How do we know when they are happy? Real user behavior requires careful design and evaluation

Different User Signal

- Clicks
- Mouse movement
- Browser action
 - bookmark, save, print
- Time
 - dwell time, time on SERP
- Explicit judgment
 - likes, favourites..
- Other page elements
 - share, ...
- Long term effects
 - sessions per user, abandonment, ...
- Reformulations

Search Engine Result Page (SERP)

The screenshot shows a Google search for "PhD advice". The search bar at the top contains the text "PhD advice" and a microphone icon. Below the search bar, there are tabs for "Web", "Images", "Videos", "News", "Shopping", "More", and "Search tools". The "Web" tab is selected. The search results show "About 111,000,000 results (0.46 seconds)". The first result is "Philip Guo - Advice for new Ph.D. students" from pgbovine.net, dated Nov 24, 2013. The second result is "PhD Advice - Find a PhD" from www.findaphd.com, dated Feb 12, 2014. The third result is "6 Essential Study Tips for the PhD Student | Top Universities" from www.topuniversities.com, dated Feb 12, 2014. The fourth result is "Surviving a PhD – 10 Top Tips... | The Thesis Whisperer" from thesiswhisperer.com, dated Jul 16, 2012. The fifth result is "Graduate School Advice: 10 Things To Know Before Starting ..." from www.nextscientist.com. The sixth result is "15 Tips For PhD Students In Their First Week - Next Scientist" from www.nextscientist.com. The seventh result is "Some Modest Advice for Graduate Students | Stearns Lab" from stearnslab.yale.edu. The eighth result is "PhD Talk: 20 Tips for Surviving your PhD" from phdtalk.blogspot.com. The ninth result is "How to stay sane through a PhD: get survival tips from fellow ..." from www.theguardian.com. The tenth result is "Finishing your PhD thesis: 15 top tips from those in the know ..." from www.theguardian.com.

Google PhD advice

Web Images Videos News Shopping More Search tools

About 111,000,000 results (0.46 seconds)

Philip Guo - Advice for new Ph.D. students
pgbovine.net/early-stage-PhD-advice.htm
Nov 24, 2013 - I know this sounds presumptuous, but if you just started a Ph.D. program, especially in science or engineering, bookmark this page and read it ...

PhD Advice - Find a PhD
www.findaphd.com/advice/
Welcome to the FindAPhD advice section. Whether you're looking for a PhD, or you're a current PhD student, this section has something for you. Check out the ...

6 Essential Study Tips for the PhD Student | Top Universities
www.topuniversities.com/blog/6-essential-study-tips-phd-student
Feb 12, 2014 - What makes a PhD student successful, productive and happy? Check out these six essential study tips for the PhD student.

Surviving a PhD – 10 Top Tips... | The Thesis Whisperer
thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/
Jul 16, 2012 - Alex is also on Twitter where he tweets about sustainability, academia, PhD advice and life. I hope you will head on over there and check out ...

Graduate School Advice: 10 Things To Know Before Starting ...
www.nextscientist.com/graduate-school-advice-series-starting-phd/
From a senior PhD student to a starting PhD student, this is the graduate school advice nobody will tell you but that you need to succeed and get your PhD title.

15 Tips For PhD Students In Their First Week - Next Scientist
www.nextscientist.com/15-tips-phd-students-start/
During the first week of a PhD you feel lost. Follow these tips for PhD students that are at the beginning of their PhDs and get a head start.

Some Modest Advice for Graduate Students | Stearns Lab
stearnslab.yale.edu/some-modest-advice-graduate-students
Secondly, your PhD work will shape your future. It is your choice of a field in which to carry out a life's work. It is also important to the dynamic of science that your ...

PhD Talk: 20 Tips for Surviving your PhD
phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html
Sep 19, 2013 - PhD studies are the highest level of education, and the road can be frustrating and exhausting at times, but the final result (your dissertation) is ...

How to stay sane through a PhD: get survival tips from fellow ...
www.theguardian.com › Higher Education Network › Academics
Mar 20, 2014 - You might have chosen to take a look at this blog because you are currently feeling overwhelmed by your PhD, or perhaps you just know of ...

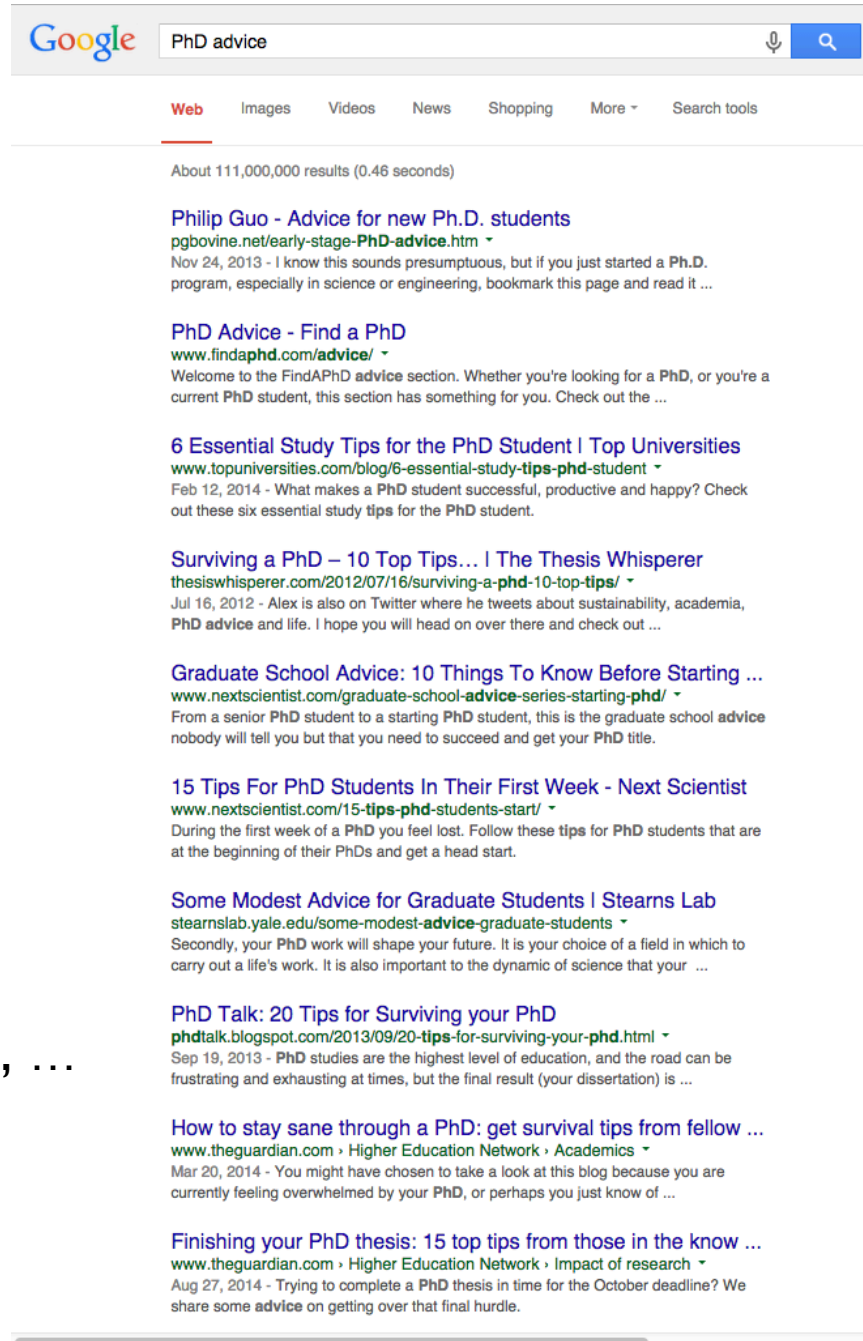
Finishing your PhD thesis: 15 top tips from those in the know ...
www.theguardian.com › Higher Education Network › Impact of research
Aug 27, 2014 - Trying to complete a PhD thesis in time for the October deadline? We share some advice on getting over that final hurdle.

User Logs

AOL-user-ct-collection — less — 116x53					
710766	wwwpeoplesearch.comwww.reviewplace.seardh	2006-05-30 22:10:13			
710766	wwwpeoplesearch.comwww.reviewplace.seardh	2006-05-30 22:10:33			
711391	can not sleep with snoring husband	2006-03-01 01:24:00			
711391	cannot sleep with snoring husband	2006-03-01 01:24:07	9		http://www.wjla.com
711391	cannot sleep with snoring husband	2006-03-01 01:24:07	9		http://www.wjla.com
711391	cannot sleep with snoring husband	2006-03-01 01:33:06	1		http://www.epinions.com
711391	jackie zeaman nude	2006-03-01 15:26:27			
711391	jackie zeman nude	2006-03-01 15:26:38			
711391	strange cosmos	2006-03-01 16:07:15	1		http://www.strangecosmos.com
711391	mansfield first assembly	2006-03-01 16:09:20	1		http://www.mansfieldfirstassembly.org
711391	mansfield first assembly	2006-03-01 16:09:20	3		http://netministries.org
711391	reverend harry myers	2006-03-01 16:10:07			
711391	reverend harry myers	2006-03-01 16:10:30			
711391	national enquirer	2006-03-01 17:13:14	1		http://www.nationalenquirer.com
711391	how to kill mockingbirds	2006-03-01 17:18:11			
711391	how to kill mockingbirds	2006-03-01 17:18:33			
711391	how to kill annoying birds in your yards	2006-03-01 17:18:58			
711391	how to kill annoying birds in your yards	2006-03-01 17:19:53	2		http://www.sortprice.com
711391	how to rid your yard of noisy annoying birds	2006-03-01 17:23:08	3		http://shopping.msn.com
711391	how to rid your yard of noisy annoying birds	2006-03-01 17:23:08	10		http://www.bergen.org
711391	how to rid your yard of noisy annoying birds	2006-03-01 17:24:35	15		http://www.saferbrand.com
711391	how do i get mocking birds out of my yard	2006-03-01 17:27:17			
711391	how do i get mockingbirds out of my yard	2006-03-01 17:27:36	9		http://www.asri.org
711391	how do i get mockingbirds out of my yard	2006-03-01 17:30:14			
711391	how to get rid of noisy loud birds	2006-03-01 17:30:52	3		http://www.bird-x.com
711391	how to get rid of noisy loud birds	2006-03-01 17:30:52	1		http://forums2.gardenweb.com
711391	how to get rid of noisy loud birds	2006-03-01 17:30:52	10		http://www.birding.com
711391	mansfield first assembly	2006-03-01 18:31:36	3		http://netministries.org
711391	beth moore	2006-03-01 19:42:41	1		http://www.lproof.org
711391	judy baker ministries	2006-03-01 19:49:03	2		http://www.embracinggrace.com
711391	god will fulfill your hearts desires	2006-03-01 19:59:06	10		http://www.pureintimacy.org
711391	online friendships can be very special	2006-03-01 23:09:37			
711391	online friendships can be very special	2006-03-01 23:09:57			
711391	online friendships	2006-03-01 23:10:24			
711391	cypress fairbanks isd	2006-03-02 07:56:53	1		http://www.cfisd.net
711391	people are not always how they seem over the internet	2006-03-02 08:31:51			
711391	friends online can be different in person	2006-03-02 08:32:42			
711391	friends online can be different in person	2006-03-02 08:33:04	13		http://www.salon.com
711391	boston butts	2006-03-02 09:47:36			
711391	community christian church houston tx	2006-03-02 16:07:53			
711391	gay churches in houston tx	2006-03-02 16:08:23			
711391	community gospel church in houston tx	2006-03-02 16:08:45	2		http://www.communitygospel.org
711391	houston tx is one hot place	2006-03-02 18:04:44			
711391	houston tx is one hot place to live	2006-03-02 18:04:55	9		http://travel.yahoo.com
711391	houston tx is one hot place to live	2006-03-02 18:16:05	1		http://www.houston-texas-online.com
711391	texas hill country and sights around san antonio tx	2006-03-02 18:19:00	5		http://www.answers.c
om					

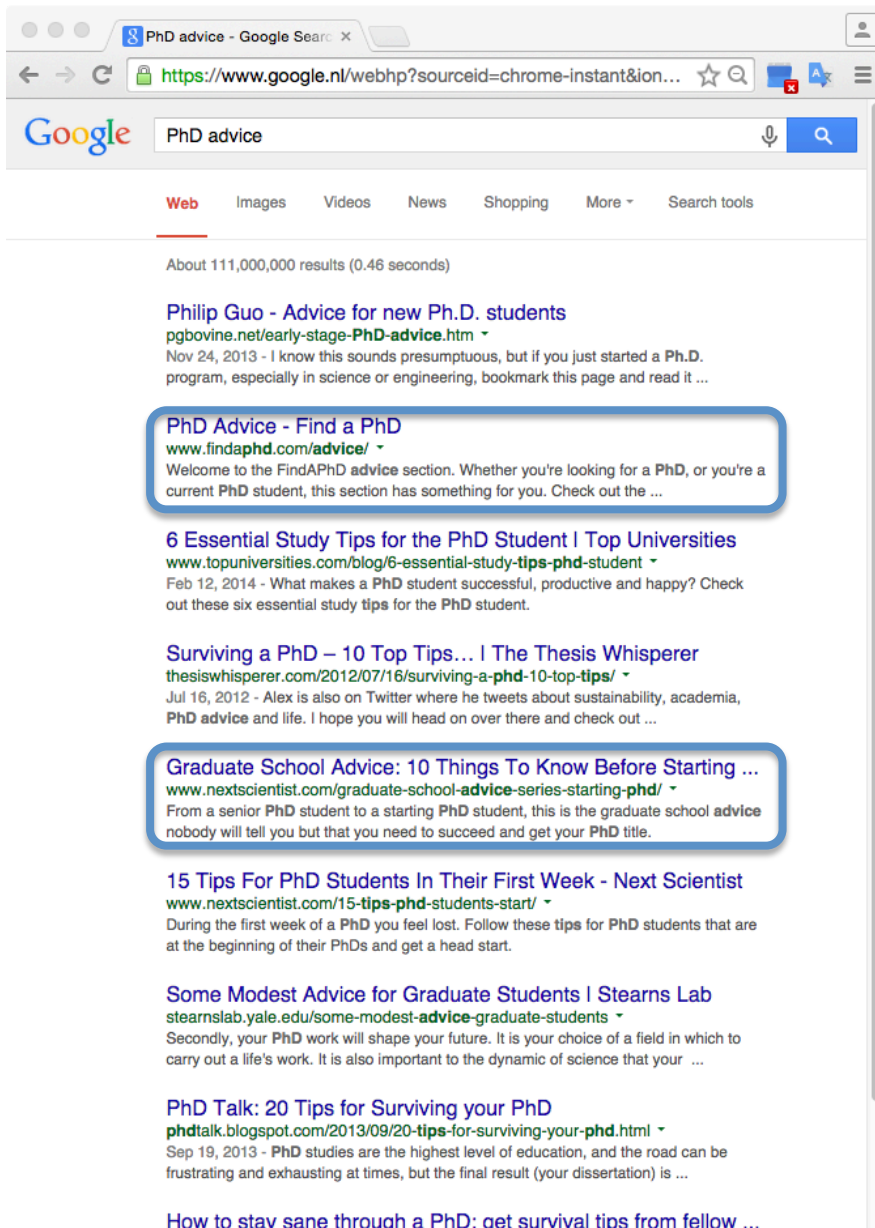
Different User Signal

Search Engine Result Page (SERP)



- Clicks
- Mouse movement
- Browser action
 - bookmark, save, print
- Time
 - dwell time, time on SERP
- Explicit judgment
 - likes, favourites..
- Other page elements
 - share, ...
- Long term effects
 - sessions per user, abandonment, ...
- Reformulations

Interpreting Clicks



- How good are clicks?
 - Are these two clicked pages equally “good”?
- How bad are non-clicks?
 - Not relevant
 - Not examined
 - The snippet gave the answer

Interpreting Clicks

Beyond Clicks: Query Reformulation as a Predictor of Search Satisfaction

Ahmed Hassan
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
hassanam@microsoft.com

Xiaolin Shi, Nick Craswell, Bill Ramsey
Microsoft Bing
One Microsoft Way
Redmond, WA 98052, USA
xishi,nickcr,brams@microsoft.com

- The user performed the following search on July 1st, 2012.

bing woman dies in a fatal accident in greenfield, minnesota

3.160.000 RESULTS

Narrow by language ▾

Narrow by region ▾

[Fatal car crashes and road traffic accidents in Greenfield ...](#)

[www.cit](#)
US acci **Woman Killed In Greenfield Crash**
acciden June 30, 2012 9:03 PM

[Star News | Otsego woman, 34, dies in Greenfield crash](#)

[erstarnews.com/2012/07/01/otsego-woman-34-dies-in-greenfield-crash](#) ▾
1-7-2012 · A 34-year-old Otsego woman was killed in an auto accident Saturday, June 30, in Greenfield, according to the Hennepin County Sheriff's Department.

[Man dies in Greenfield Township motorcycle accident ...](#)

[www.goerie.com/.../man-dies-in-greenfield-township-motorcycle-accident](#) ▾
26-5-2015 · ... A 69-year-old man was killed and his passenger was seriously injured when their ... **Man dies in Greenfield Township motorcycle accident.** Staff ...

[21-year-old Annandale man dies in head-on crash on Hwy. ...](#)

[www.delanoheraldjournal.com/...man-dies...hwy-55-in-greenfield-tuesday](#) ▾
A 21-year-old Annandale man died from injuries received in a head-on accident on Highway 55 in Greenfield Jan. 12 just before 4 p.m. ... MN and the surrounding area

[Man killed in Greenfield car accident - YouTube](#)

[www.youtube.com/watch?v=3EfXpg_ssuA](#) ▾
By WWLP-22News · 35 sec · 156 views · Added 12-1-2014

21-year-old Annandale man dies in head-on crash on Hwy. 55 in Greenfield Tuesday

Posted on January 13, 2010 by Ryan Gueningsman DHJ Managing Editor

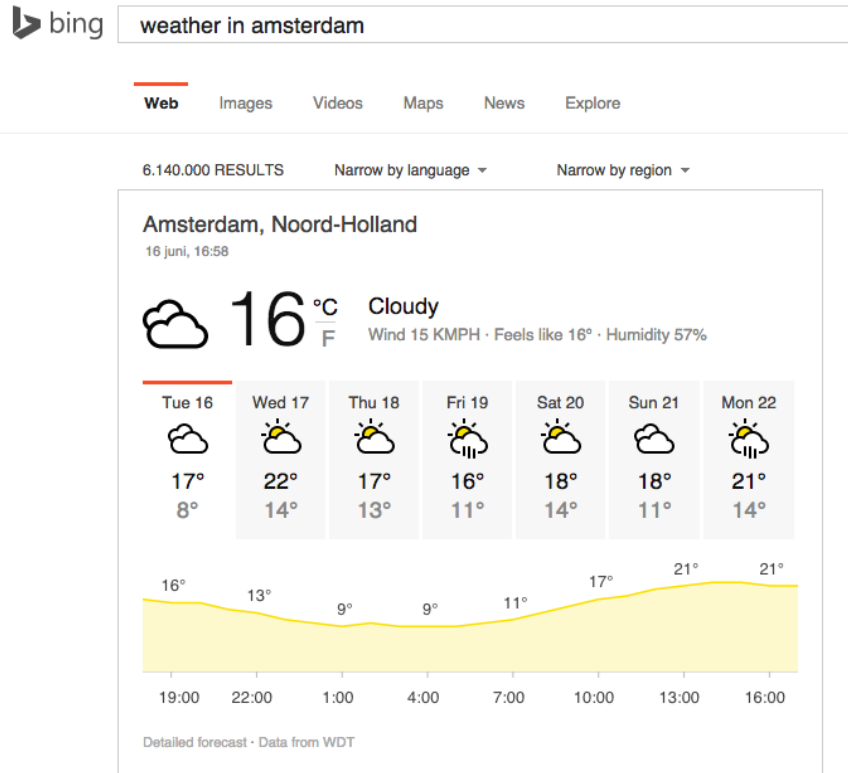
- Clicks do not always mean satisfaction.

Interpreting Clicks

Beyond Clicks: Query Reformulation as a Predictor of Search Satisfaction

Ahmed Hassan
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
hassanam@microsoft.com

Xiaolin Shi, Nick Craswell, Bill Ramsey
Microsoft Bing
One Microsoft Way
Redmond, WA 98052, USA
xishi,nickcr,brams@microsoft.com



- Lack of clicks does not always mean dissatisfaction.

[Amsterdam, Netherlands Weather - 10 Day Weather ...](#)

www.weather.com/weather/today/l/Amsterdam Netherlands NLXX0002

Rain or shine? Be prepared with the most accurate 10 day forecast for **Amsterdam**, Netherlands, with highs, lows, chance of precipitation and more from **weather.com**

[Amsterdam, Netherlands Forecast | Weather Underground](#)

www.wunderground.com/weather-forecast/NL/Amsterdam.html ▾

Weather Underground provides local & long range **Weather** Forecast, **weather** reports, maps & tropical **weather** conditions for locations worldwide.

[BBC Weather - Amsterdam](#)

www.bbc.com/weather/2759794 ▾

Detailed **weather** for **Amsterdam** with a 5 to 10 day forecast, giving a look further ahead.

Interpreting Clicks

- Clicks are **biased** and **noisy**, but **useful**
 - Clicks are noisy
 - they don't always mean what you hope
 - absence of clicks is not always negative

Interpreting Clicks

Evaluating the Accuracy of Implicit Feedback from
Clicks and Query Reformulations in Web Search

THORSTEN JOACHIMS
Dept. of Computer Science, Cornell University
and
LAURA GRANKA
Google Inc.
and
BING PAN
School of Business and Economics, College of Charleston
and
HELENE HEMBROOKE
Dept. of Information Science, Cornell University
and
FILIP RADLINSKI
Dept. of Computer Science, Cornell University
and
GERI GAY
Dept. of Information Science, Cornell University

Accurately Interpreting Clickthrough Data as Implicit Feedback

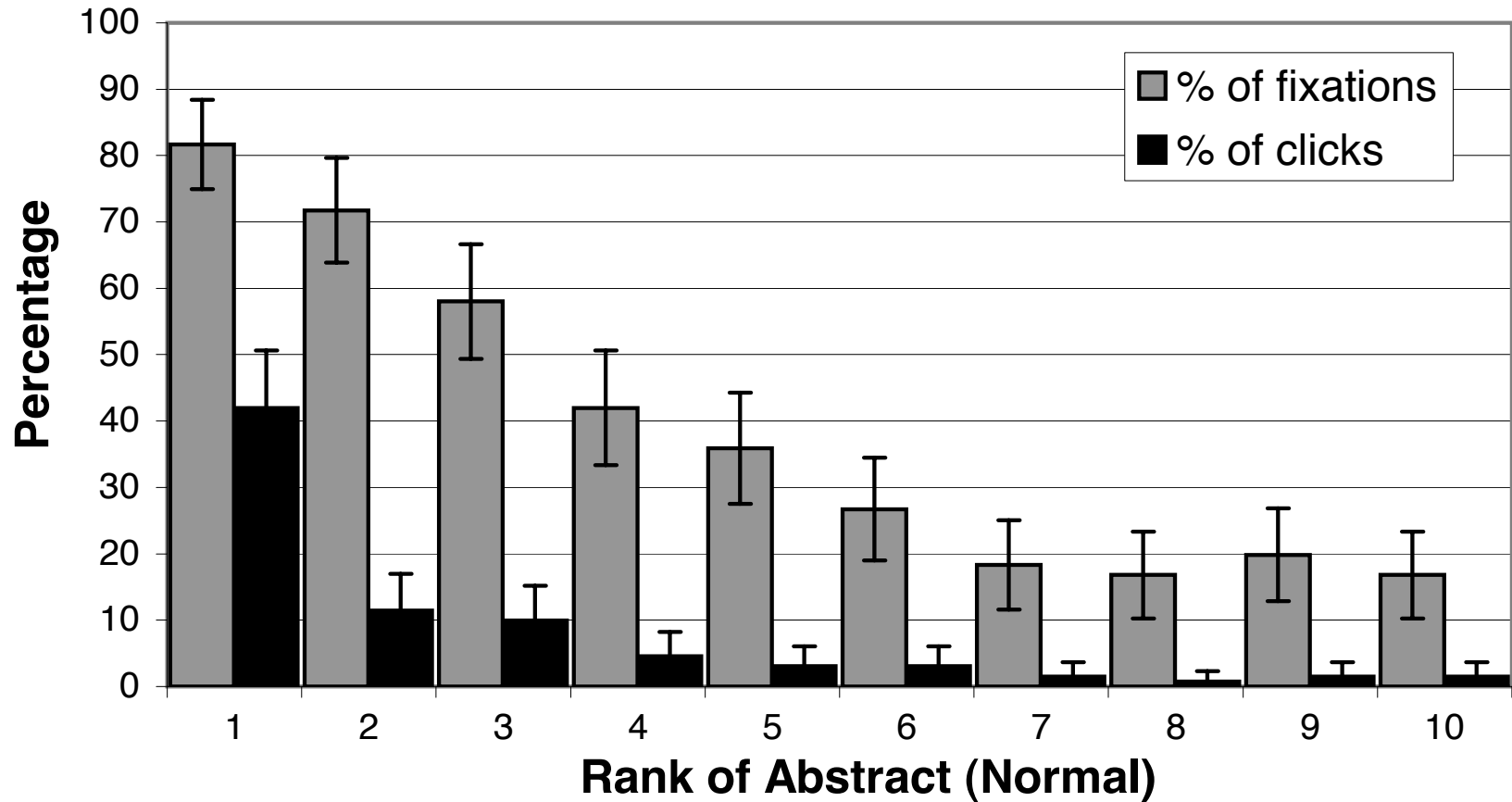
Thorsten Joachims
Dept. of Computer Science
Cornell University
Ithaca, NY, USA
tj@cs.cornell.edu

Laura Granka
Dept. of Communication
Stanford University
Palo Alto, CA, USA
granka@stanford.edu

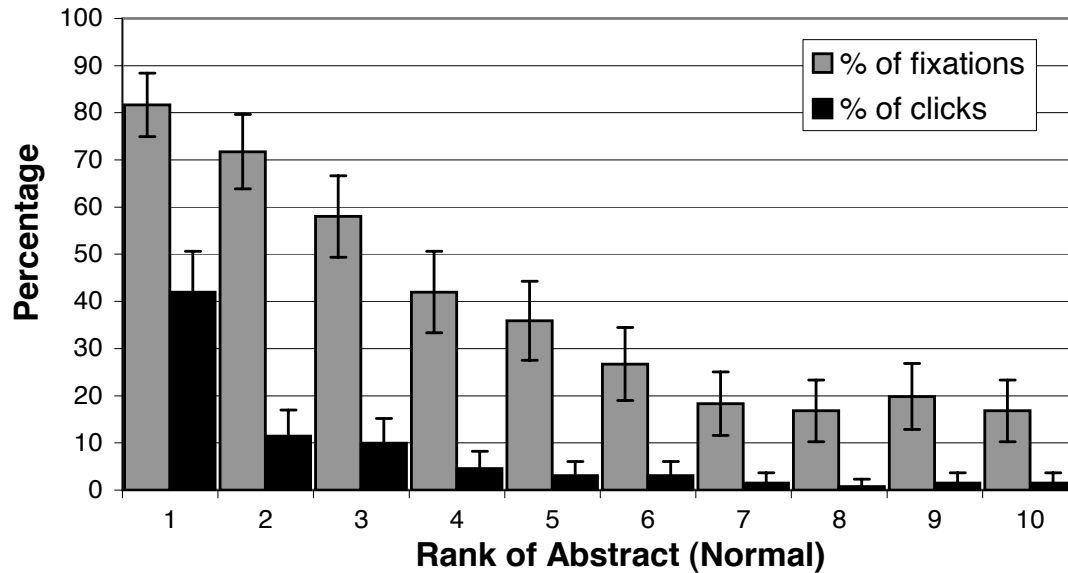
Bing Pan,
Helene Hembrooke &
Geri Gay
Information Science
Cornell University
Ithaca, NY, USA
{bp58,hah4,gkg1}@cornell.edu

- Lab study of web search
- 16 subjects, 5 navigational and 5 informational search tasks each
- Behavior recorded using eye-tracking

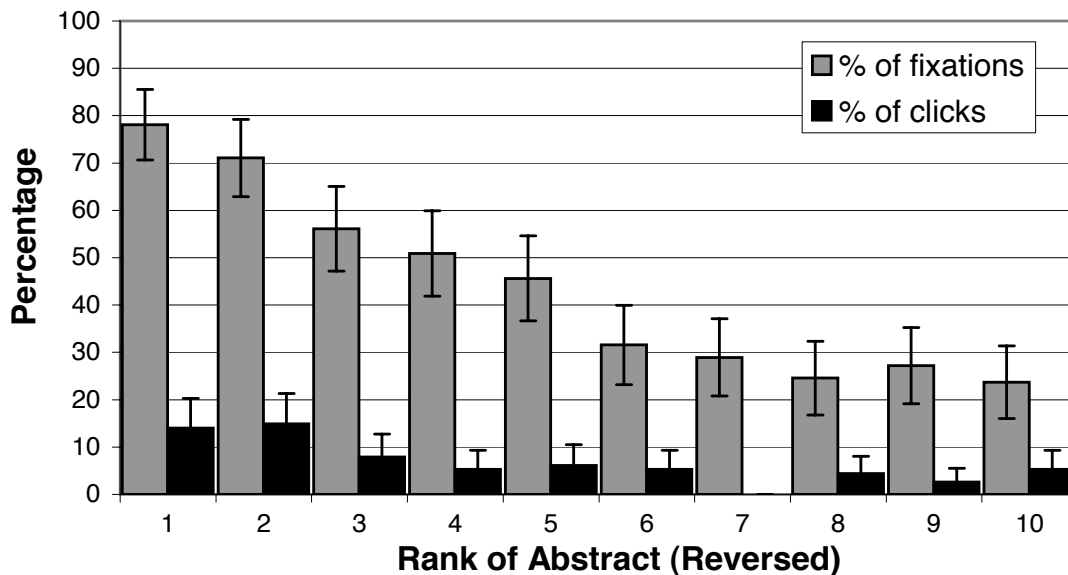
What do Users View / Click?



Are Clicks Affected by Relevance?



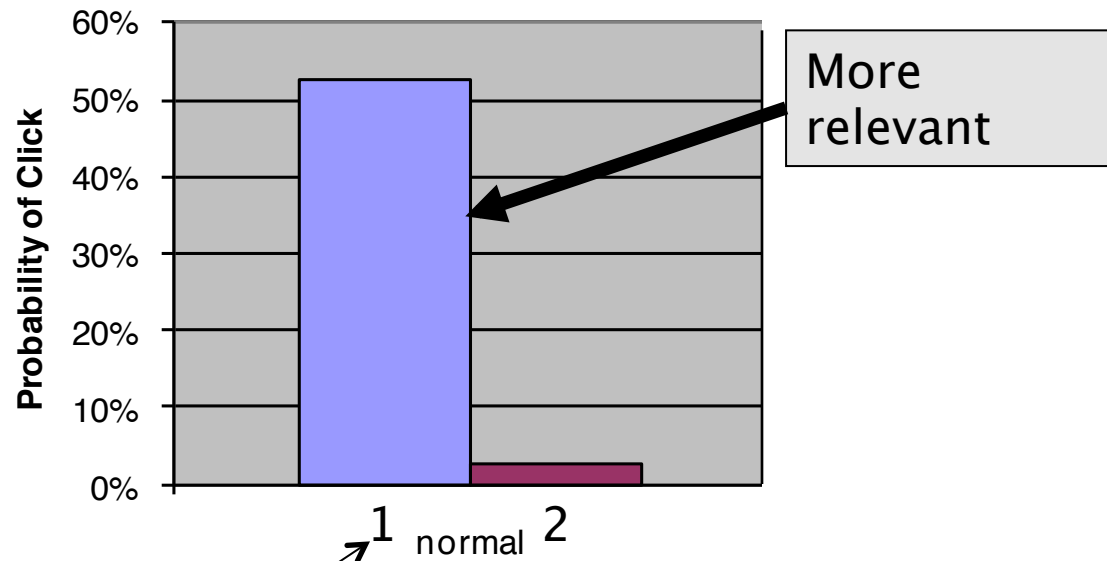
Average rank of
click:
2.66



Average rank of
click:
4.03*

Position Bias

Hypothesis: ~~Order of presentation influences where users look, but not where they click!~~



Normal:

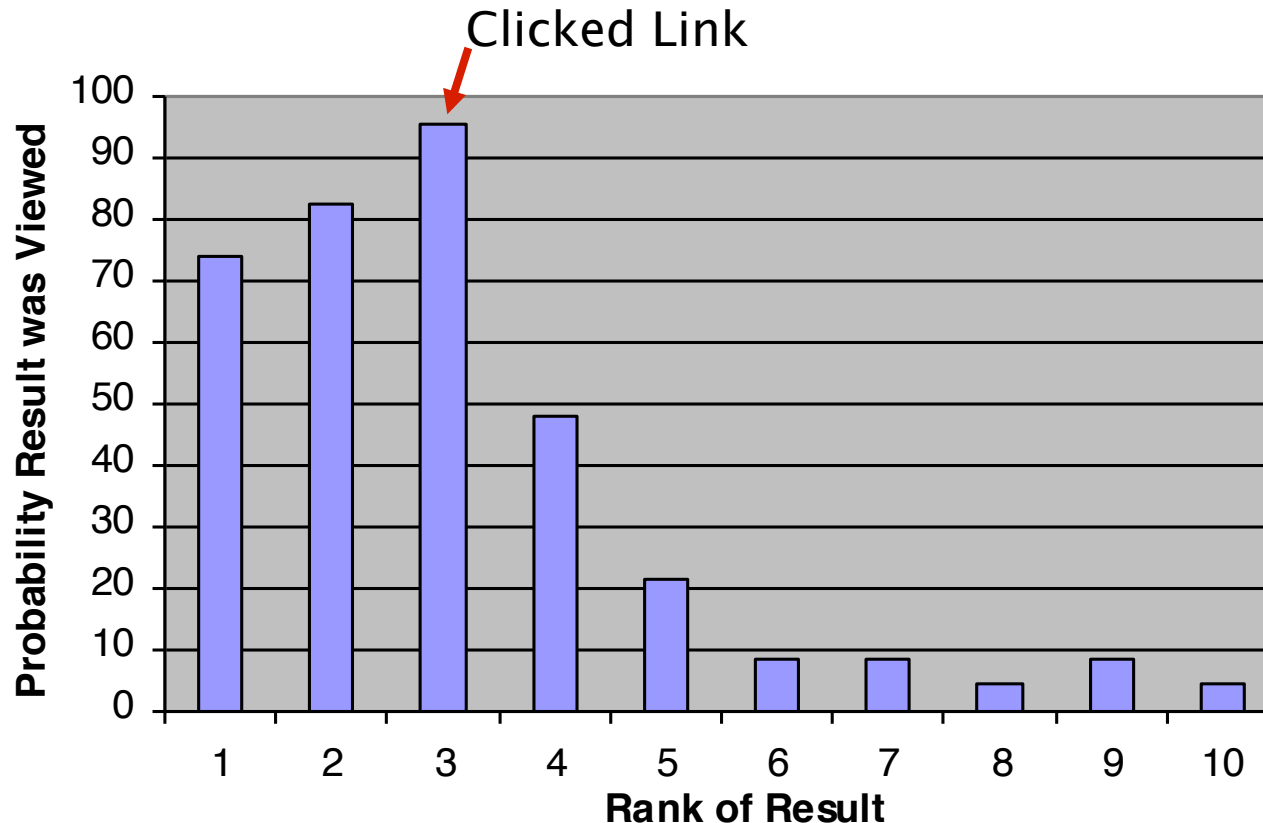
Google's order
of results

Users appear to have trust in Google's ability
to rank the most relevant result first.

Swapped:

Order of top 2
results swapped

Which Results are Viewed **Before** Click?



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

Is Click = Relevant?

- Can we simply interpret clicks as relevance
 - This would provide relevance labels, and a **collection-based evaluation** could be run
- A variety of biases make this difficult:
 - **Position Bias:**
Users are more inclined to examine and click on higher-ranked results
 - **Contextual Bias:**
Whether users click on a result depends on other nearby results
 - **Attention Bias:**
Users click more on results which draw attention to themselves

Interpreting Clicks

- Clicks are **biased** and **noisy**, but **useful**
 - Clicks are noisy
 - they don't always mean what you hope
 - absence of clicks is not always negative
 - Clicks are biased
 - users won't click on things you didn't show them
 - user are likely to click on things that appear **high** in the ranking
 - **presentation** matters
 - documents, snippets, images, colors, font size, grouped with other documents
 - **surrounding results** matter

Interpreting Clicks

- Clicks are **biased** and **noisy**, but **useful**
 - Clicks are noisy
 - they don't always mean what you hope
 - absence of clicks is not always negative
 - Clicks are biased
 - users won't click on things you didn't show them
 - user are likely to click on things that appear **high** in the ranking
 - **presentation** matters
 - documents, snippets, images, colors, font size, grouped with other documents
 - **surrounding results** matter

However: In the long run, clicks do point in the **right** direction

Different User Signal

- Clicks
- Mouse movement
- Browser action
 - bookmark, save, print
- Time
 - dwell time, time on SERP
- Explicit judgment
 - likes, favourites..
- Other page elements
 - share, ...
- Long term effects
 - sessions per user, abandonment, ...
- Reformulations

Search Engine Result Page (SERP)

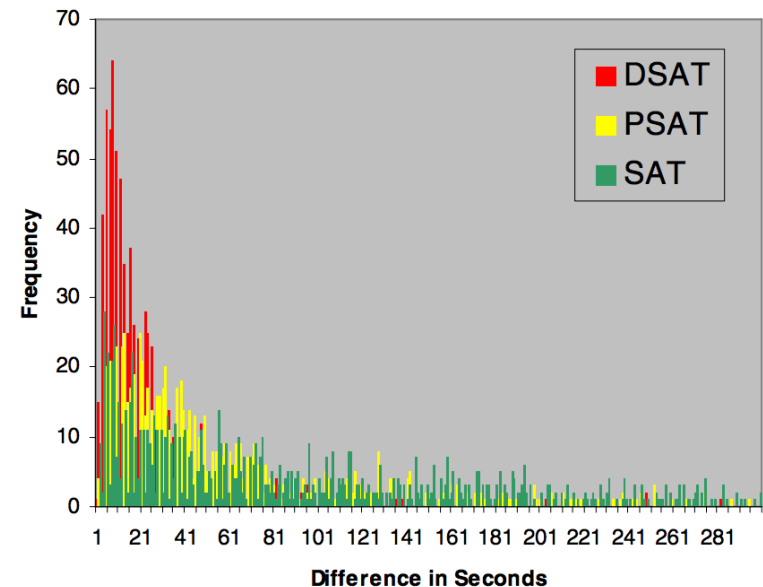
The screenshot shows a Google search for 'PhD advice'. The search bar at the top contains the text 'PhD advice' and a microphone icon. Below the search bar, there are tabs for 'Web', 'Images', 'Videos', 'News', 'Shopping', 'More', and 'Search tools'. The 'Web' tab is selected. The results show 'About 111,000,000 results (0.46 seconds)'. The first result is 'Philip Guo - Advice for new Ph.D. students' from 'pgbovine.net/early-stage-PhD-advice.htm', dated Nov 24, 2013. The second result is 'PhD Advice - Find a PhD' from 'www.findaphd.com/advice/', dated Feb 12, 2014. The third result is '6 Essential Study Tips for the PhD Student | Top Universities' from 'www.topuniversities.com/blog/6-essential-study-tips-phd-student', dated Feb 12, 2014. The fourth result is 'Surviving a PhD – 10 Top Tips... | The Thesis Whisperer' from 'thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/', dated Jul 16, 2012. The fifth result is 'Graduate School Advice: 10 Things To Know Before Starting ...' from 'www.nextscientist.com/graduate-school-advice-series-starting-phd/'. The sixth result is '15 Tips For PhD Students In Their First Week - Next Scientist' from 'www.nextscientist.com/15-tips-phd-students-start/'. The seventh result is 'Some Modest Advice for Graduate Students | Stearns Lab' from 'stearnslab.yale.edu/some-modest-advice-graduate-students'. The eighth result is 'PhD Talk: 20 Tips for Surviving your PhD' from 'phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html'. The ninth result is 'How to stay sane through a PhD: get survival tips from fellow ...' from 'www.theguardian.com'. The tenth result is 'Finishing your PhD thesis: 15 top tips from those in the know ...' from 'www.theguardian.com'.

Beyond Clicks

Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White

{stevef, kuldeepk, markmyd, sdumais, tomwh}@microsoft.com

- A large number of implicit feedback tested
- Two most important in predicting SAT clicks
 - **Dwell Time**
 - Time spent on a clicked page.
 - SAT click > 30 secs
 - **Exit Type**
 - The way in which the user exited the result – kill browser window, new query, navigate using history, favorites or URL entry or time out.

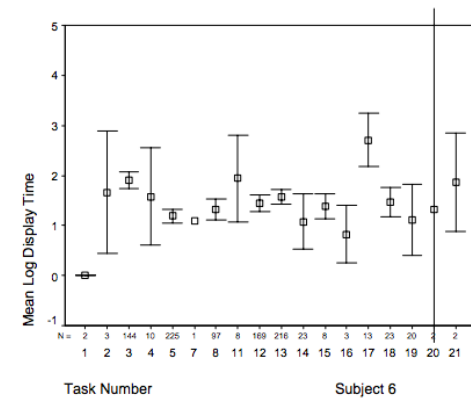
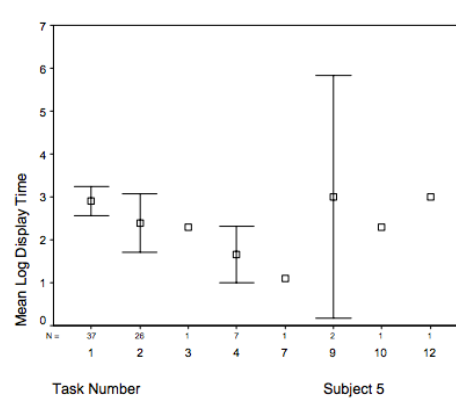
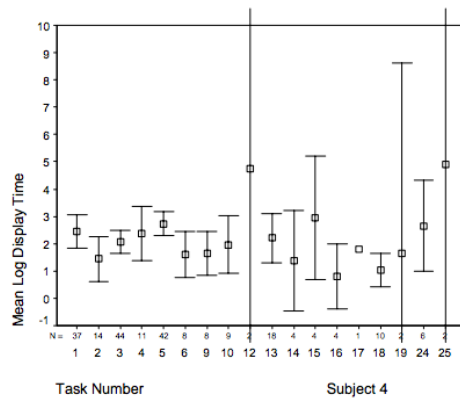
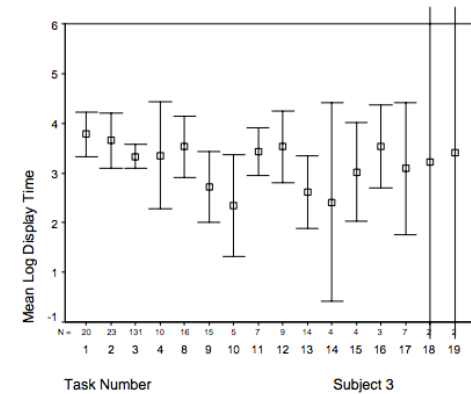
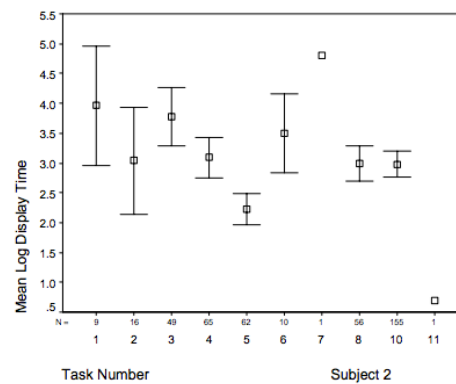
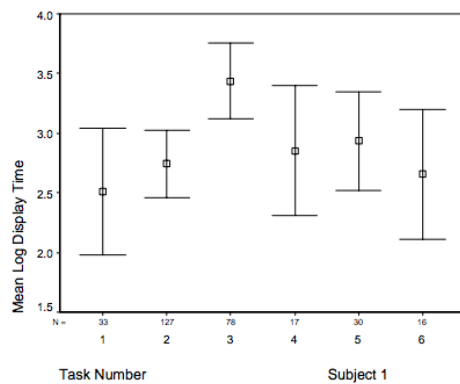


Beyond Clicks: Dwell Time

Display Time as Implicit Feedback: Understanding Task Effects

Diane Kelly
SILS
University of North Carolina
Chapel Hill, NC 27599-3360 USA
kelly@ils.unc.edu

Nicholas J. Belkin
SCILS
Rutgers University
New Brunswick, NJ 08901 USA
nick@belkin.rutgers.edu



Beyond Clicks: Dwell Time

Modeling Dwell Time to Predict Click-level Satisfaction

Youngho Kim^{1*}, Ahmed Hassan², Ryen W. White², and Imed Zitouni²

¹ University of Massachusetts, 140 Governors Drive, Amherst, MA 01003, USA

² Microsoft, One Microsoft Way, Redmond, WA 98052, USA

yhkim@cs.umass.edu, {hassanam, ryenw, izitouni}@microsoft.com

- Model Dwell Time by a Gamma distribution:

$$t \sim \Gamma(k, \theta)$$

- Maximum Likelihood Estimation

- given SAT and DSAT clicks
- for each click segment
- $P(t \mid \text{SAT}, \text{att})$ and $P(t \mid \text{DSAT}, \text{att})$

- Query–click attributes to generate click segments

- Query topic attributes
- Query type attributes
- Page topic attributes
- Reading level attributes

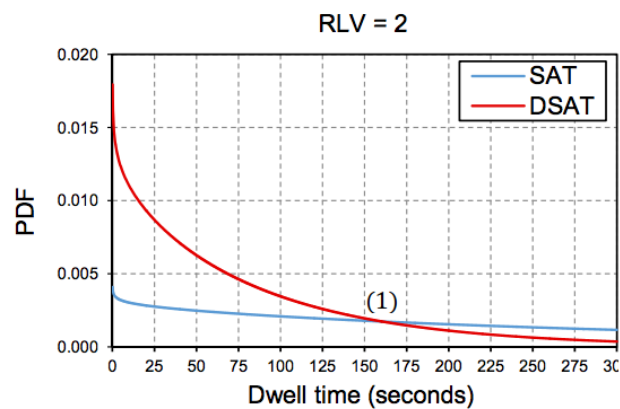


Figure 3: SAT and DSAT PDFs of RLV = 2 (easy).

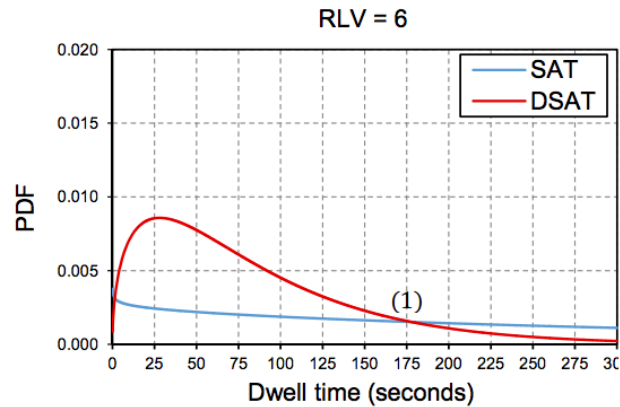


Figure 4: SAT and DSAT PDFs of RLV = 6 (moderate).

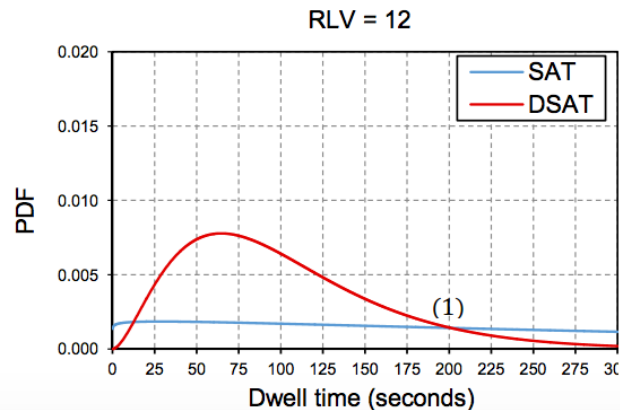


Figure 5: SAT and DSAT PDFs of RLV = 12 (difficult).

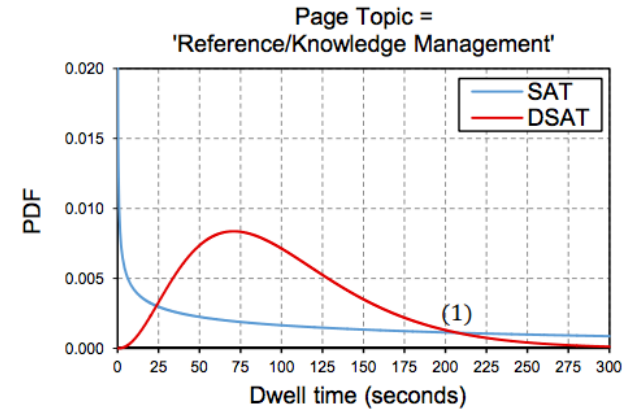


Figure 6: SAT and DSAT PDFs of a sample attribute.

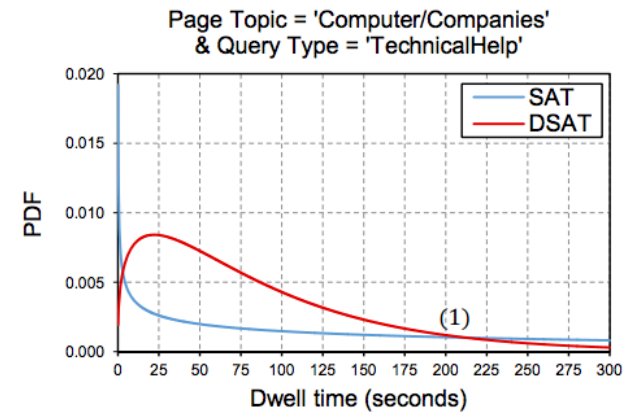


Figure 7: SAT and DSAT PDFs of a “technical” segment.

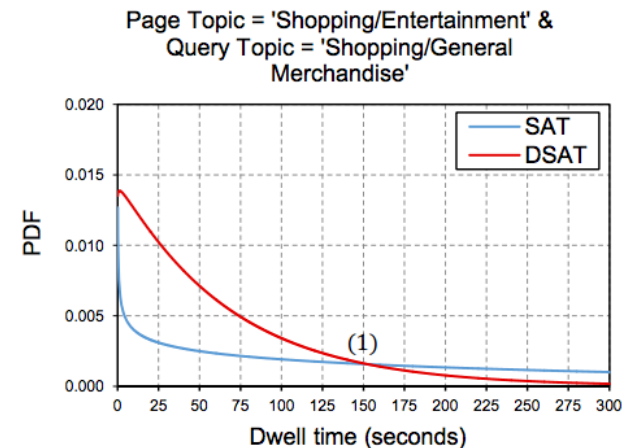


Figure 8: SAT and DSAT PDFs of a “shopping” segment.

Different User Signal

- Clicks
- Mouse movement
- Browser action
 - bookmark, save, print
- Time
 - dwell time, time on SERP
- Explicit judgment
 - likes, favourites..
- Other page elements
 - share, ...
- Long term effects
 - sessions per user, abandonment, ...
- Reformulations

Search Engine Result Page (SERP)

The screenshot shows a Google search for 'PhD advice'. The search bar at the top contains the text 'PhD advice' and a search icon. Below the search bar, there are tabs for 'Web', 'Images', 'Videos', 'News', 'Shopping', 'More', and 'Search tools'. The 'Web' tab is selected. The results show 'About 111,000,000 results (0.46 seconds)'. The first result is 'Philip Guo - Advice for new Ph.D. students' from 'pgbovine.net/early-stage-PhD-advice.htm', dated Nov 24, 2013. The second result is 'PhD Advice - Find a PhD' from 'www.findaphd.com/advice/', dated Feb 12, 2014. The third result is '6 Essential Study Tips for the PhD Student | Top Universities' from 'www.topuniversities.com/blog/6-essential-study-tips-phd-student', dated Feb 12, 2014. The fourth result is 'Surviving a PhD – 10 Top Tips... | The Thesis Whisperer' from 'thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/', dated Jul 16, 2012. The fifth result is 'Graduate School Advice: 10 Things To Know Before Starting ...' from 'www.nextscientist.com/graduate-school-advice-series-starting-phd/'. The sixth result is '15 Tips For PhD Students In Their First Week - Next Scientist' from 'www.nextscientist.com/15-tips-phd-students-start/'. The seventh result is 'Some Modest Advice for Graduate Students | Stearns Lab' from 'stearnslab.yale.edu/some-modest-advice-graduate-students'. The eighth result is 'PhD Talk: 20 Tips for Surviving your PhD' from 'phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html'. The ninth result is 'How to stay sane through a PhD: get survival tips from fellow ...' from 'www.theguardian.com'. The tenth result is 'Finishing your PhD thesis: 15 top tips from those in the know ...' from 'www.theguardian.com'.

Beyond Dwell Time: Post-click Behavior

- Dwell time not sufficient
- Interactions on landing pages
 - Cursor movements and scrolling
 - Reading vs. Scanning

Beyond Dwell Time: Estimating Document Relevance from Cursor Movements and other Post-click Searcher Behavior

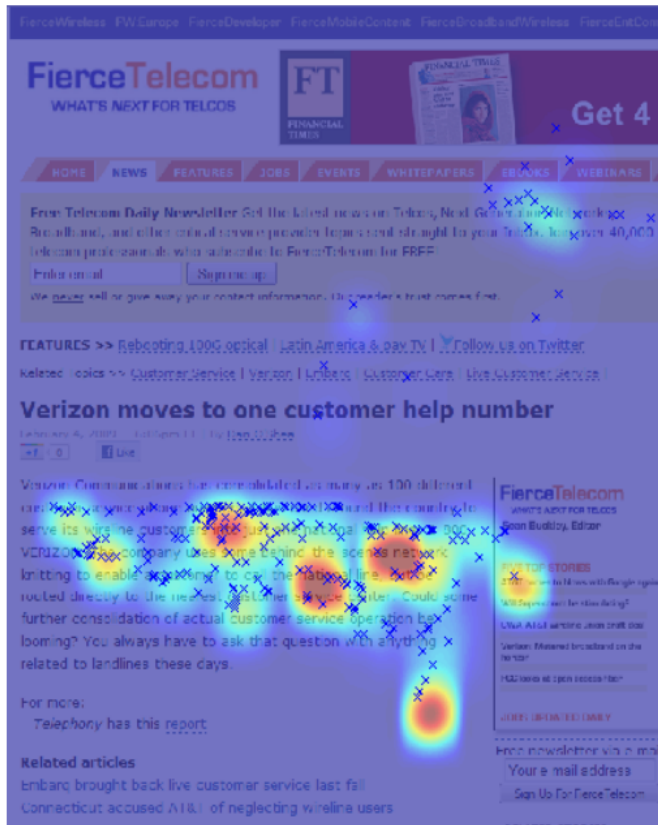
Qi Guo
Mathematics & Computer Science Department
Emory University
qguo3@emory.edu

Eugene Agichtein
Mathematics & Computer Science Department
Emory University
eugene@mathcs.emory.edu

Beyond Dwell Time: Post-click Behavior

Example 1:

Find the phone number of the Verizon Wireless helpline for Massachusetts



(a) relevant (dwell time: 30s)



(b) non-relevant (dwell time: 30s)

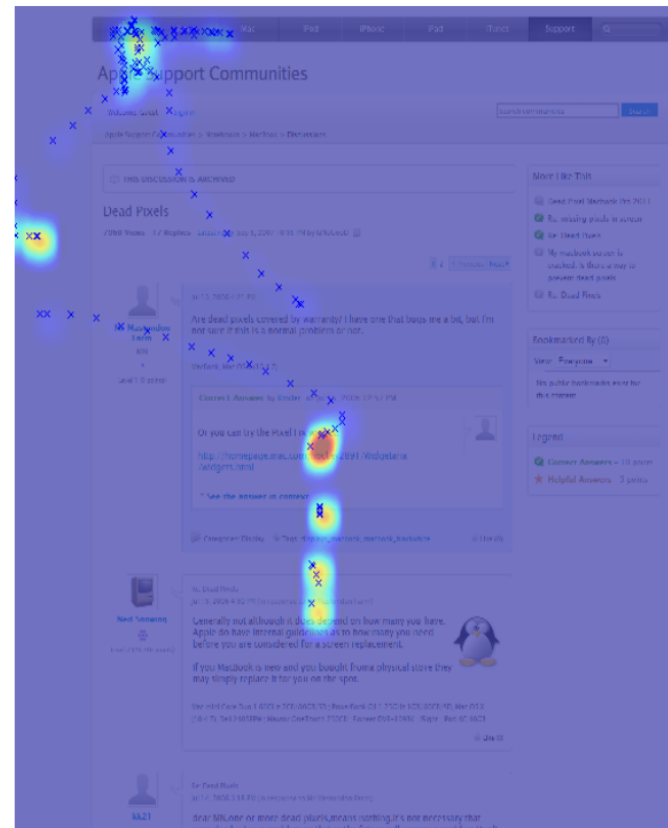
Beyond Dwell Time: Post-click Behavior

Example 2:

How many pixels must be dead on a MacBook before Apple will replace the laptop?



(a) relevant (dwell time: 70s)



(b) non-relevant (dwell time: 80s)

Beyond Dwell Time: Post-click Behavior

- Patterns of post-click interactions:
 1. Periods of horizontal reading
 2. Focused attention
 3. Left-prevalence
 4. “Scanning” followed by “reading”
 5. “Reading” followed by “scanning”
 6. “Skipping” – quick scrolling

Beyond Dwell Time: Post-click Behavior

- Features:

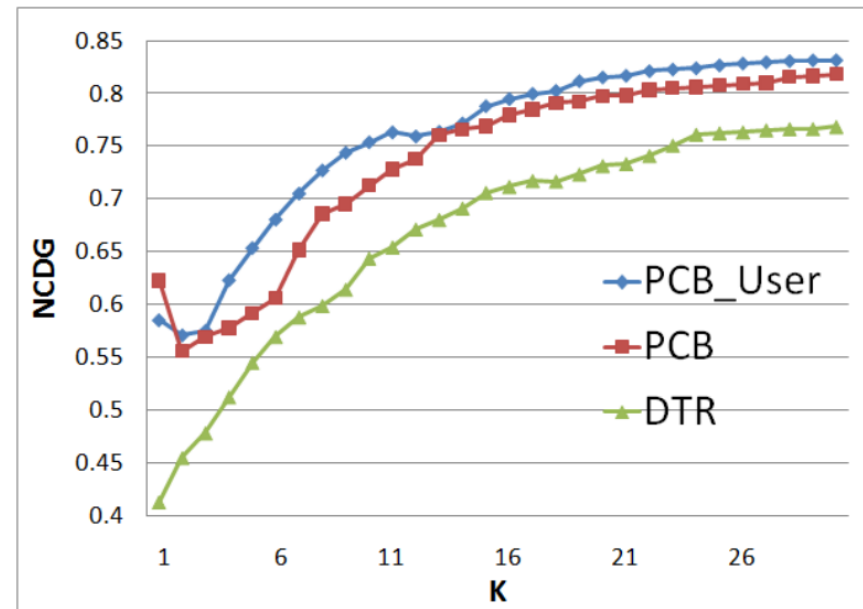
Dwell (1)	<i>dwell</i> : time of the page view in seconds	0.167**
Rank (1)	<i>rank</i> : the rank of the document or the rank of the origin (i.e., the landing page) of the search trail that the document is on if its rank is not available	-0.073
Cursor (14)	<i>cursorcnt</i> : num. of cursor movements	0.164**
	<i>cursorfreq</i> : cursorcnt/dwell	-0.082*
	<i>dist</i> : total overall distance the cursor traveled in pixels	-0.137**
	<i>xdist</i> : total distance the cursor traveled horizontally in pixels	0.101**
	<i>ydist</i> : total distance the cursor traveled horizontally in pixels	0.172**
	<i>speed</i> : dist/dwell	-0.101**
	<i>xspeed</i> : xdist/dwell	-0.143**
	<i>yspeed</i> : ydist/dwell	-0.124**
	<i>xmin</i> : minimal x coordinate	0.112**
	<i>ymin</i> : minimal y coordinate	0.093*
	<i>xmax</i> : maximal x coordinate	0.067
	<i>ymax</i> : maximal y coordinate	0.243**
	<i>xrange</i> : xmax-xmin	-0.006
	<i>yrange</i> : ymax-ymin	0.172**

Scroll (5)	<i>scrlcnt</i> : num. of vertical scrolls	-0.008
	<i>scrlfreq</i> : scrlcnt/dwell	-0.206**
	<i>scrlldist</i> : total vertical scroll distance	-0.092*
	<i>scrlspeed</i> : scrlldist/dwell	-0.212**
	<i>scrlmax</i> : maximum scroll top	-0.026
AOI (3)	<i>dwell_aoi</i> : total time the cursor spent in the pre-defined Area of Interest (AOI)	0.227**
	<i>cursorcnt_aoi</i> : cursor count in AOI	0.189**
	<i>cursorfreq_aoi</i> : cursorcnt/dwell	-0.195**
Task (6)	<i>avg_dwell</i> : average dwell time of preceding page views in the task	0.081*
	<i>querycnt</i> : num. of preceding queries	-0.138**
	<i>serpcnt</i> : num. of preceding search engine result page (SERP) views	-0.142**
	<i>clkcnt</i> : num. of preceding clicks	-0.171**
	<i>ctr</i> : clkcnt/serpcnt	0.085*
	<i>tasktime</i> : total time elapsed in seconds since the task started	-0.046

Beyond Dwell Time: Post-click Behavior

- Results:
 - Correlation w/ relevance
 - Re-ranking

<i>Single Feature Group</i>	<i>RR</i>	<i>BRT</i>
<i>PCB</i>	0.399*+	0.411*+
<i>cursor</i>	0.326*+	0.389*+
<i>scroll</i>	0.277+	0.268*+
<i>aoi</i>	0.261*+	0.177*
<i>task</i>	0.201*	0.146*
<i>dwell</i>	0.184*	0.136
<i>rank</i>	0.04	0.136
<i>DTR</i>	0.211	0.231



Beyond Clicks: Cursor movements on SERP

- Cursor – gaze relationship

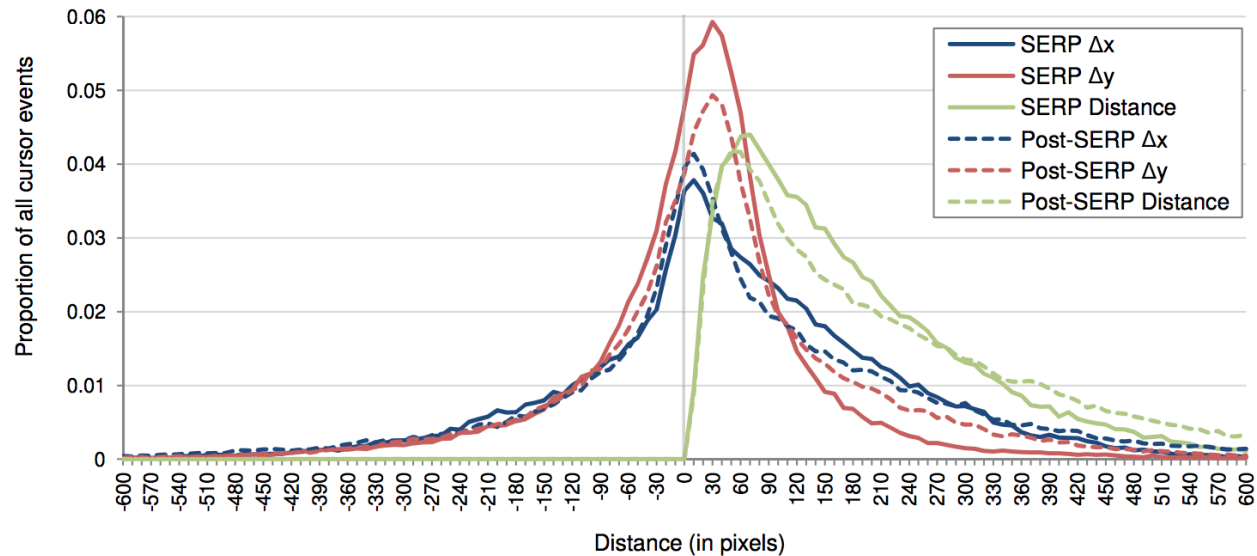


Figure 1. Δx , Δy , and Euclidean distance plotted in a frequency distribution for SERP and post-SERP pages. Solid lines represent these distances gathered on the SERP, while dashed lines represented distances gathered on post-SERP pages (landing pages).

No Clicks, No Problem: Using Cursor Movements to Understand and Improve Search

Jeff Huang
Information School
University of Washington
chi@jeffhuang.com

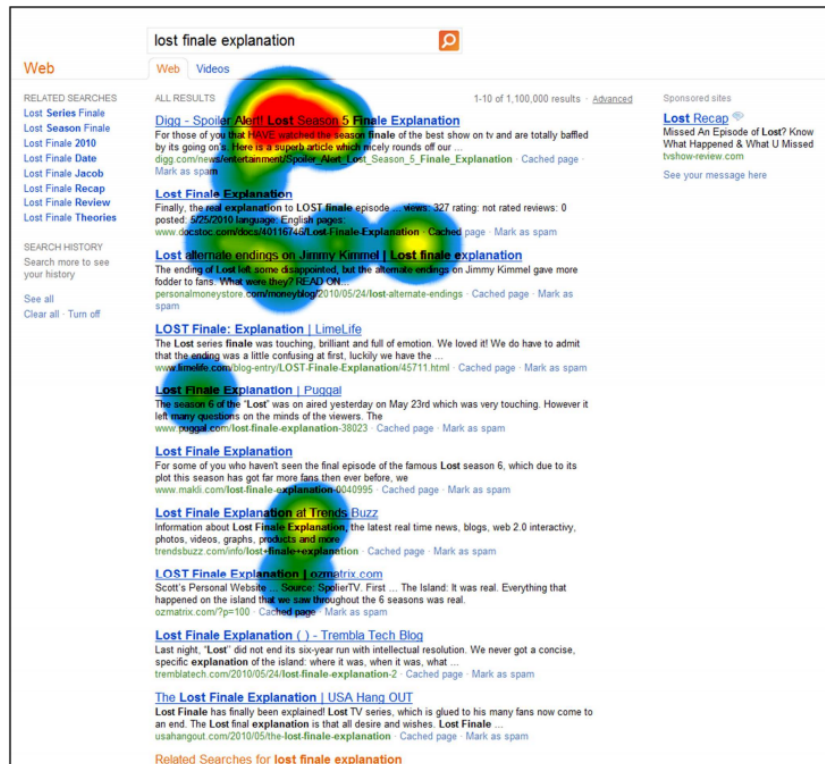
Ryen W. White
Microsoft Research
Redmond, WA 98052
ryenw@microsoft.com

Susan Dumais
Microsoft Research
Redmond, WA 98052
sdumais@microsoft.com

Beyond Clicks: Cursor movements on SERP

• Cursor movement vs. clicks

Click positions



Cursor movement positions



Figure 2. Heatmaps of all click positions (left) and recorded cursor positions (right) for the query *[lost finale explanation]*. Heavy interaction occurs in red/orange/yellow areas, moderate interaction in green areas, light interaction in blue areas.

Beyond Clicks: Cursor movements on SERP

- Average time **hovering** result titles

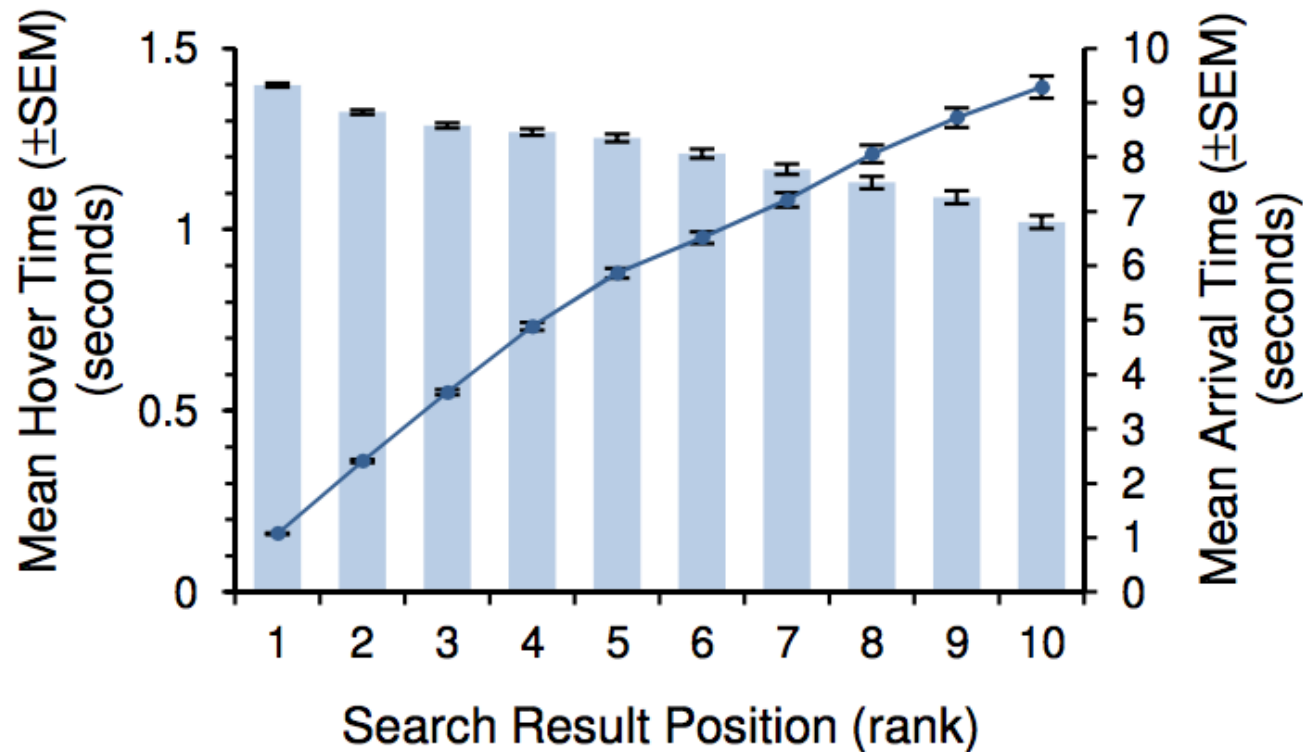


Figure 3. Mean title hover duration (bars) and mean time for cursor to arrive at each result (circles).

Beyond Clicks: Cursor movements on SERP

- Results hovered before clicked, etc.

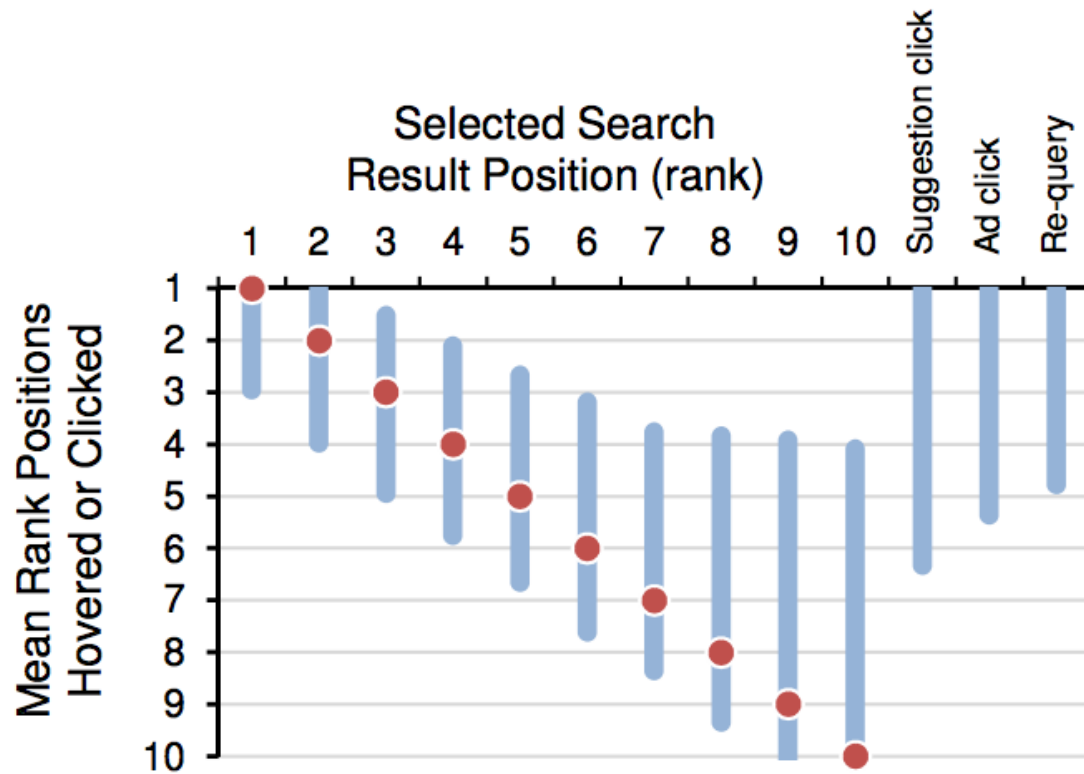


Figure 4. Mean number of search results hovered over before users clicked on a result (above and below that result). Result clicks are red circles, result hovers are blue lines.

Beyond Clicks: Cursor movements on SERP

- Unclicked hover vs. clicks

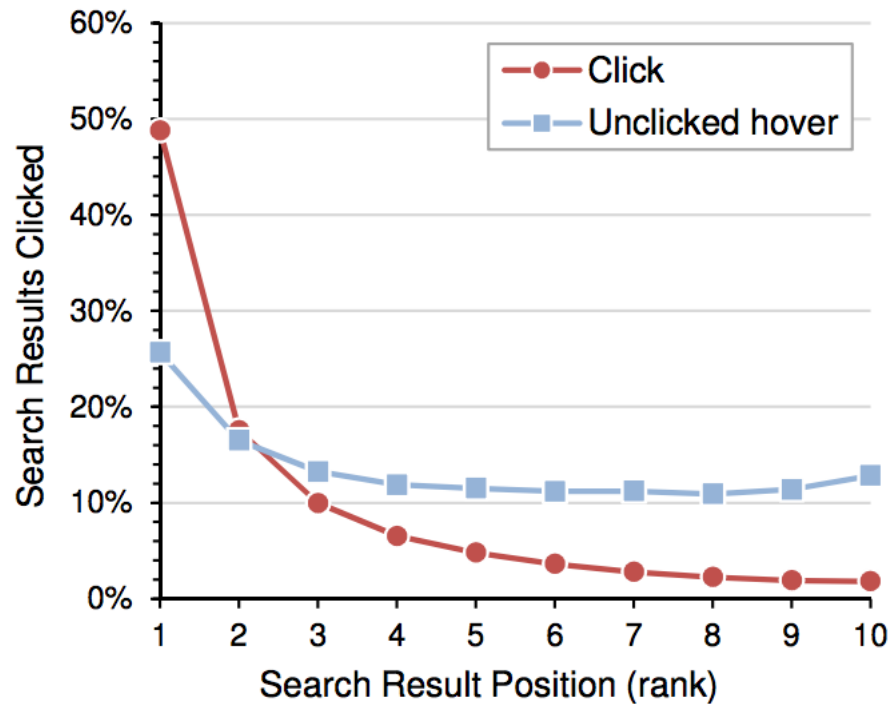


Figure 6. Proportion of search results that are eventually clicked after an unclicked hover, plotted against the click distribution from Figure 5.

Beyond Clicks: Cursor movements on SERP

- Correlations with relevance

Table 3. Correlations between click and hover features and relevance judgments for queries with and without clicks.

Result clicks or no clicks	Feature source	Correlation with human relevance judgments
Clicks (N=1194)	Clickthrough rate (c)	0.42
	Hover rate (h)	0.46
	Unclicked hovers (u)	-0.26
	Max hover time (d)	-0.15
	Combined ¹	0.49
No clicks (N=96)	Hover rate	0.23
	Unclicked hovers	0.06
	Max hover time	0.17
	Combined ²	0.28

¹ $y = 2.25 - 0.1c + 1.38h - 0.08u - 0.12d$; ² $y = 0.36 + 0.80h + 0.22u + 0.30d$

Different User Signal

- Clicks
- Mouse movement
- Browser action
 - bookmark, save, print
- Time
 - dwell time, time on SERP
- Explicit judgment
 - likes, favourites..
- Other page elements
 - share, ...
- Long term effects
 - sessions per user, abandonment, ...
- Reformulations

Search Engine Result Page (SERP)

The screenshot shows a Google search for 'PhD advice'. The search bar at the top contains the text 'PhD advice' and a microphone icon. Below the search bar, there are tabs for 'Web', 'Images', 'Videos', 'News', 'Shopping', 'More', and 'Search tools'. The 'Web' tab is selected. The results show 'About 111,000,000 results (0.46 seconds)'. The first result is 'Philip Guo - Advice for new Ph.D. students' from 'pgbovine.net/early-stage-PhD-advice.htm', dated Nov 24, 2013. The second result is 'PhD Advice - Find a PhD' from 'www.findaphd.com/advice/', dated Feb 12, 2014. The third result is '6 Essential Study Tips for the PhD Student | Top Universities' from 'www.topuniversities.com/blog/6-essential-study-tips-phd-student', dated Feb 12, 2014. The fourth result is 'Surviving a PhD – 10 Top Tips... | The Thesis Whisperer' from 'thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/', dated Jul 16, 2012. The fifth result is 'Graduate School Advice: 10 Things To Know Before Starting ...' from 'www.nextscientist.com/graduate-school-advice-series-starting-phd/'. The sixth result is '15 Tips For PhD Students In Their First Week - Next Scientist' from 'www.nextscientist.com/15-tips-phd-students-start/'. The seventh result is 'Some Modest Advice for Graduate Students | Stearns Lab' from 'stearnslab.yale.edu/some-modest-advice-graduate-students'. The eighth result is 'PhD Talk: 20 Tips for Surviving your PhD' from 'phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html', dated Sep 19, 2013. The ninth result is 'How to stay sane through a PhD: get survival tips from fellow ...' from 'www.theguardian.com', dated Mar 20, 2014. The tenth result is 'Finishing your PhD thesis: 15 top tips from those in the know ...' from 'www.theguardian.com', dated Aug 27, 2014.

Beyond Clicks

Beyond Clicks: Query Reformulation as a Predictor of Search Satisfaction

Ahmed Hassan
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
hassanam@microsoft.com

Xiaolin Shi, Nick Craswell, Bill Ramsey
Microsoft Bing
One Microsoft Way
Redmond, WA 98052, USA
xishi,nickcr,brams@microsoft.com

- The user performed the following search on July 1st, 2012.

bing woman dies in a fatal accident in greenfield, minnesota greenfield, mn accident

3.160.000 RESULTS Narrow by language Narrow by region

21-year-old Annandale man dies in head-on crash on Hwy. 55 in Greenfield Tuesday

Posted on January 13, 2010 by Ryan Gueningsman DHJ Managing Editor

Fatal car crashes and road traffic accidents in Greenfield ...
www.city-data.com/accidents/acc-Greenfield-Minnesota.html
US accidents; Accidents in Greenfield, MN; Fatal car crashes and road traffic accidents in Greenfield, Minnesota.

Star News | Otsego woman, 34, dies in Greenfield crash
erstarnews.com/2012/07/01/otsego-woman-34-dies-in-greenfield-crash
1-7-2012 · A 34-year-old Otsego woman was killed in an auto accident Saturday, June 30, in Greenfield, according to the Hennepin County Sheriff's Department.

Man dies in Greenfield Township motorcycle accident ...
www.goerie.com/.../man-dies-in-greenfield-township-motorcycle-accident
26-5-2015 · ... A 69-year-old man was killed and his passenger was seriously injured when their ... Man dies in Greenfield Township motorcycle accident. Staff ...

21-year-old Annandale man dies in head-on crash on Hwy. ...
www.delanoheraldjournal.com
A 21-year-old Annandale man was killed in a head-on crash on Highway 55 in Greenfield on Saturday, June 30, 2012 at 9:03 PM.

Woman Killed In Greenfield Crash

Man killed in Greenfield car accident - YouTube
www.youtube.com/watch?v=3EfXpg_ssua
By WWLP-22News · 35 sec · 156 views · Added 12-1-2014
Slippery conditions on a Greenfield street likely contributed to the death of 31 year old

- Clicks do not always mean satisfaction.

Beyond Clicks: Query Reformulation

- Given a query Q1, SERP, and Q2, predict SERP level satisfaction
- Ground truth model
 - CTR and CTR-30
- Experimental model
 - Query similarity
 - Q1 and Q2 overlap if one term in common
 - Time between queries
 - Quick (less than or equal to 5 minutes) vs. Non-quick reformulation

Beyond Clicks: Query Reformulation

- Reformulation vs. CTR

Table 1. Relative CTR for different subsets of pairs, using word overlap and a 5 minute time threshold

	<i>overall</i>	<i>non-overlap</i>	<i>overlap</i>
<i>overall</i>	0%	11%	-21%
<i>non-quick</i>	25%	24%	29%
<i>quick</i>	-29%	-17%	-39%

Table 2. Relative CTR-30 for different subsets of pairs, using word overlap and a 5 minute time threshold

	<i>overall</i>	<i>non-overlap</i>	<i>overlap</i>
<i>overall</i>	0%	6%	-12%
<i>non-quick</i>	6%	-1%	40%
<i>quick</i>	-7%	20%	-30%

Beyond Clicks: Query Reformulation

- Classification:
 - Clicks
 - SAT Clicks
 - Reformulation (several features)
 - Similarity & Time
 - Reformulation + Clicks
 - If reformulation then DSAT
 - If not reformulation then use clicks
 - Reformulation + Clicks (classifier)

Beyond Clicks: Query Reformulation

Table 5. Query Success Prediction Performance

		<i>Accuracy</i>	<i>SAT Precision</i>	<i>SAT Recall</i>	<i>DSAT Precision</i>	<i>DSAT Recall</i>	<i>SAT F1</i>	<i>DSAT F1</i>
1	Clicks Only	38.86%	35.88%	64.21%	46.77%	21.51%	46.04%	29.47%
2	Sat Clicks Only ($\tau=10s$)	51.20%	42.19%	54.34%	61.10%	49.05%	47.50%	54.42%
3	Sat Click Only ($\tau=30s$)	56.07%	46.22%	49.87%	63.75%	60.31%	47.97%	61.98%
4	Sat Click Only ($\tau=50s$)	60.61%	51.80%	43.55%	65.18%	72.28%	47.32%	68.54%
5	Reformulation Only	79.17%	64.79%	97.16%	97.58%	68.41%	77.74%	80.43%
6	Reformulation + Clicks (2 stages)	73.17%	68.70%	62.37%	75.78%	80.56%	65.38%	78.10%
7	Reformulation + SAT Click ($\tau=10s$) (2 stages)	73.22%	73.85%	52.76%	72.97%	87.22%	61.55%	79.46%
8	Reformulation + SAT Click ($\tau=30s$) (2 stages)	73.01%	76.51%	48.42%	71.80%	89.83%	59.31%	79.81%
9	Reformulation + SAT Click ($\tau=50s$) (2 stages)	71.99%	78.92%	42.37%	70.06%	92.26%	55.14%	79.64%
10	Reformulation + Clicks (Classifier)	84.23%	77.74%	81.19%	88.53%	86.04%	79.43%	87.27%

No Clicks

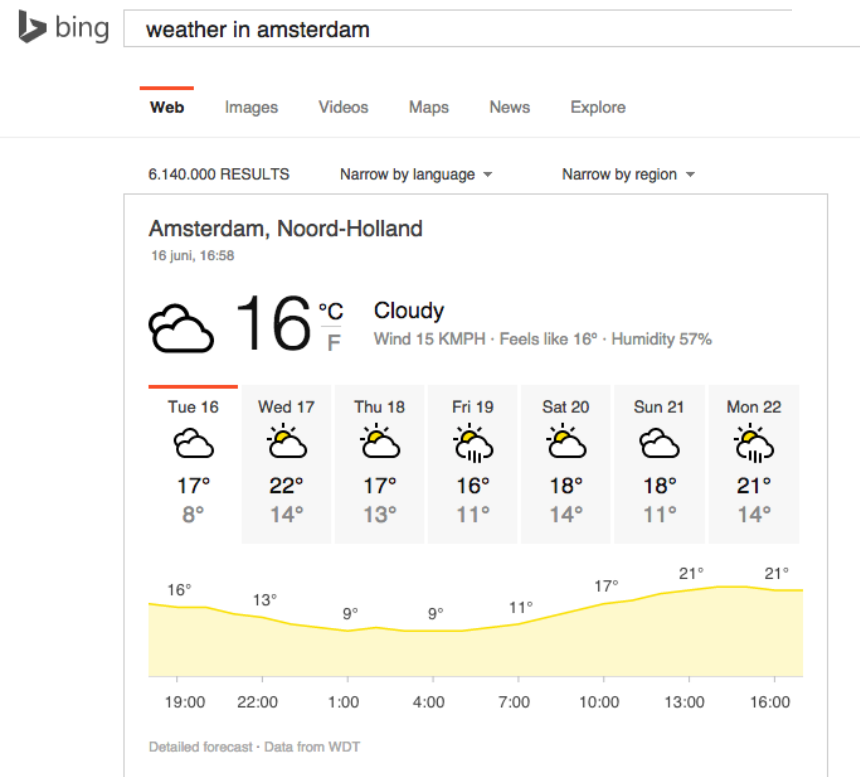
Leaving So Soon? Understanding and Predicting Web Search Abandonment Rationales

Abdigani Diriye¹, Ryen W. White², Georg Buscher², and Susan T. Dumais²

¹University College London Interaction Centre, University College London, UK, WC1E 6BT

²Microsoft Corporation, One Microsoft Way, Redmond, WA, USA 98052

a.diriye@ucl.ac.uk, {ryenw, georgbu, sdumais}@microsoft.com



[Amsterdam, Netherlands Weather - 10 Day Weather ...](#)

[www.weather.com/weather/today//Amsterdam](#) Netherlands NLXX0002

Rain or shine? Be prepared with the most accurate 10 day forecast for **Amsterdam**, Netherlands, with highs, lows, chance of precipitation and more from **weather.com**

[Amsterdam, Netherlands Forecast | Weather Underground](#)

[www.wunderground.com/weather-forecast/NL/Amsterdam.html](#)

Weather Underground provides local & long range **Weather** Forecast, **weather** reports, maps & tropical **weather** conditions for locations worldwide.

[BBC Weather - Amsterdam](#)

[www.bbc.com/weather/2759794](#)

Detailed **weather** for **Amsterdam** with a 5 to 10 day forecast, giving a look further ahead.

[Amsterdam, Netherlands Weather Forecast and Conditions ...](#)

[www.weather.com/weather/today//Amsterdam](#) Netherlands NLXX0002

Amsterdam, Netherlands **weather** forecast and **weather** conditions. Today's and

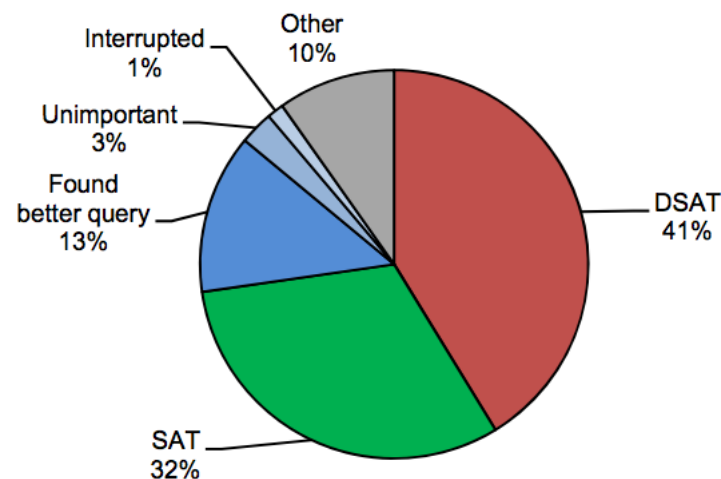


Figure 1. Reasons for SERP abandonment.

No Clicks

Good Aban

rnet Search

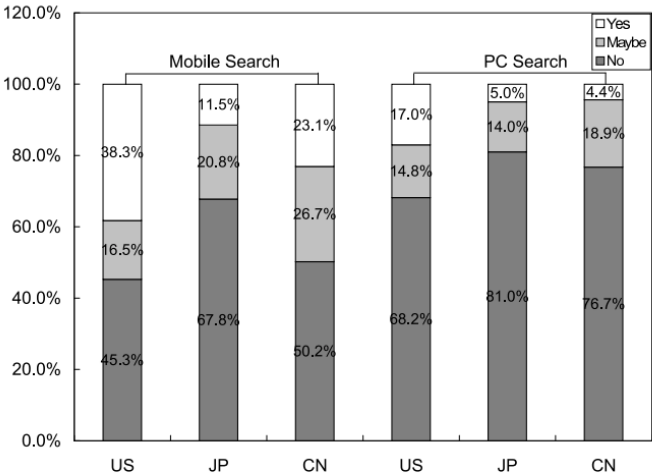


Figure 1: Percentages of queries classified as “Yes”, “Maybe”, “No” with respect to the *potential good abandonment* definition in six abandoned query samples.

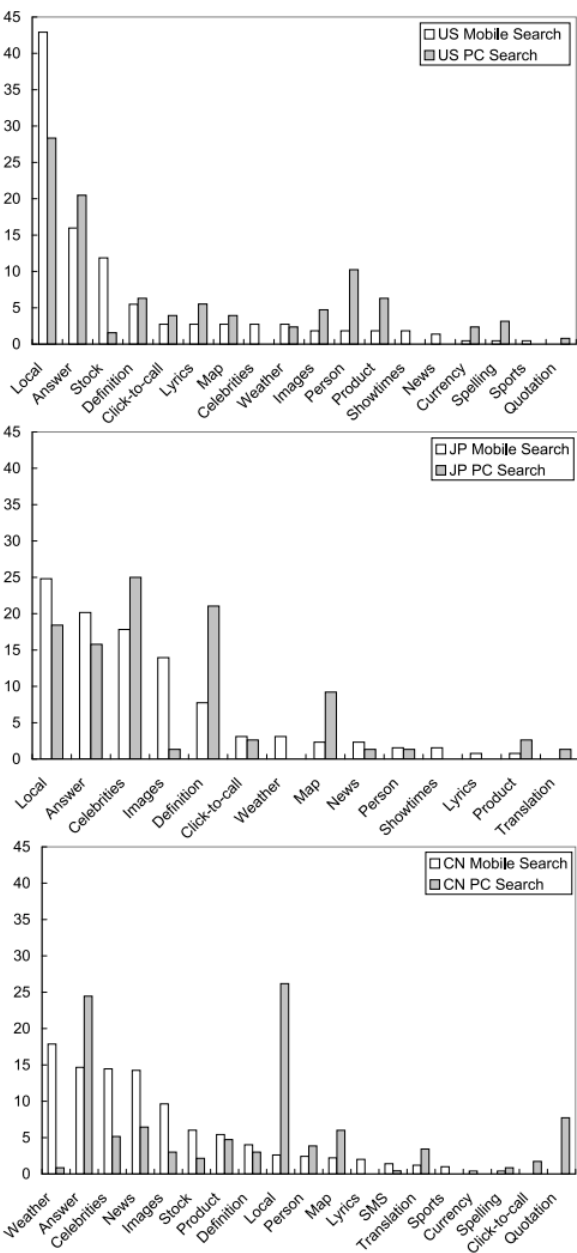


Figure 3: Category distribution (in percentage) of potential good abandonment queries in mobile and PC searches in three countries. The categories are sorted by their prevalence in mobile search for each locale.

Cursor Movement: Good vs Bad Abandonment

No Clicks, No Problem: Using Cursor Movements to Understand and Improve Search

Jeff Huang
Information School
University of Washington
chi@jeffhuang.com

Ryen W. White
Microsoft Research
Redmond, WA 98052
ryenw@microsoft.com

Susan Dumais
Microsoft Research
Redmond, WA 98052
sdumais@microsoft.com

- Cursor trail length
 - Total distance traveled by the cursor on the SERP
- Movement time
 - Total time of movement on the SERP
- Cursor speed

Table 4. Features of cursor trails for queries associated with likely good and bad abandonment.

Feature	Abandonment Type			
	<i>Good</i>		<i>Bad</i>	
	<u>M</u>	<u>SEM</u>	<u>M</u>	<u>SEM</u>
Cursor trail length (px)	1084	98	1521	71
Movement time (secs)	10.3	0.9	12.8	0.6
Cursor speed (px/sec)	104	9	125	5
Number of queries	184		675	

Mouse movement subsequences

Discovering Common Motifs in Cursor Movement Data
for Improving Web Search

Dmitry Lagun
Emory University
dlagun@mathcs.emory.edu

Mikhail Ageev *
Moscow State University
mageev@yandex.ru

Qi Guo *
Microsoft
qigu@microsoft.com

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

**Different Users, Different Opinions: Predicting Search
Satisfaction with Mouse Movement Information**

Yiqun Liu[†], Ye Chen[†], Jinhui Tang[‡], Jiashen Sun^{*}, Min Zhang[†], Shaoping Ma[†], Xuan Zhu^{*}
[†]Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science &
Technology, Tsinghua University, Beijing, China

[‡]School of Computer Science & Engineering, Nanjing University of Science and Technology

^{*}Samsung R&D Institute China - Beijing
yiqunliu@tsinghua.edu.cn

- Instead of engineering complex features,
discover common subsequencies (motifs)
- **Motif** is a frequently occurring sequence
of cursor movements

Cursor Data: Challenges

- Different users examine web pages with different speed
 - Flexible distance metric:
Dynamic Time Warping
- Similar movements can appear in different parts of a web page
 - Location invariance:
normalize subsequence position

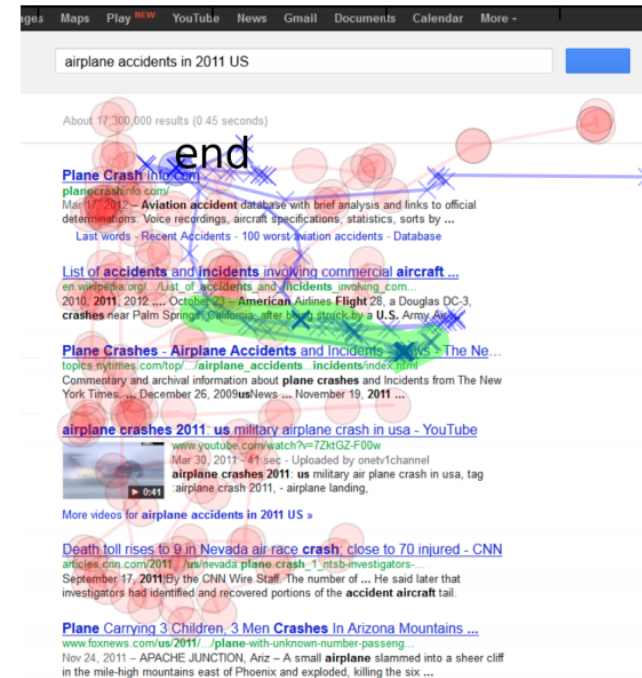
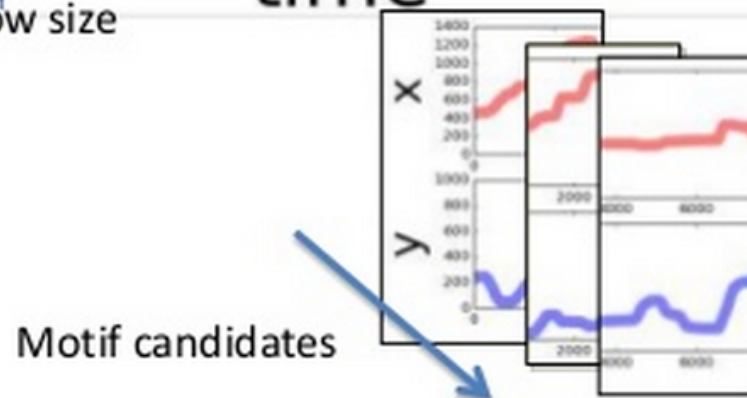
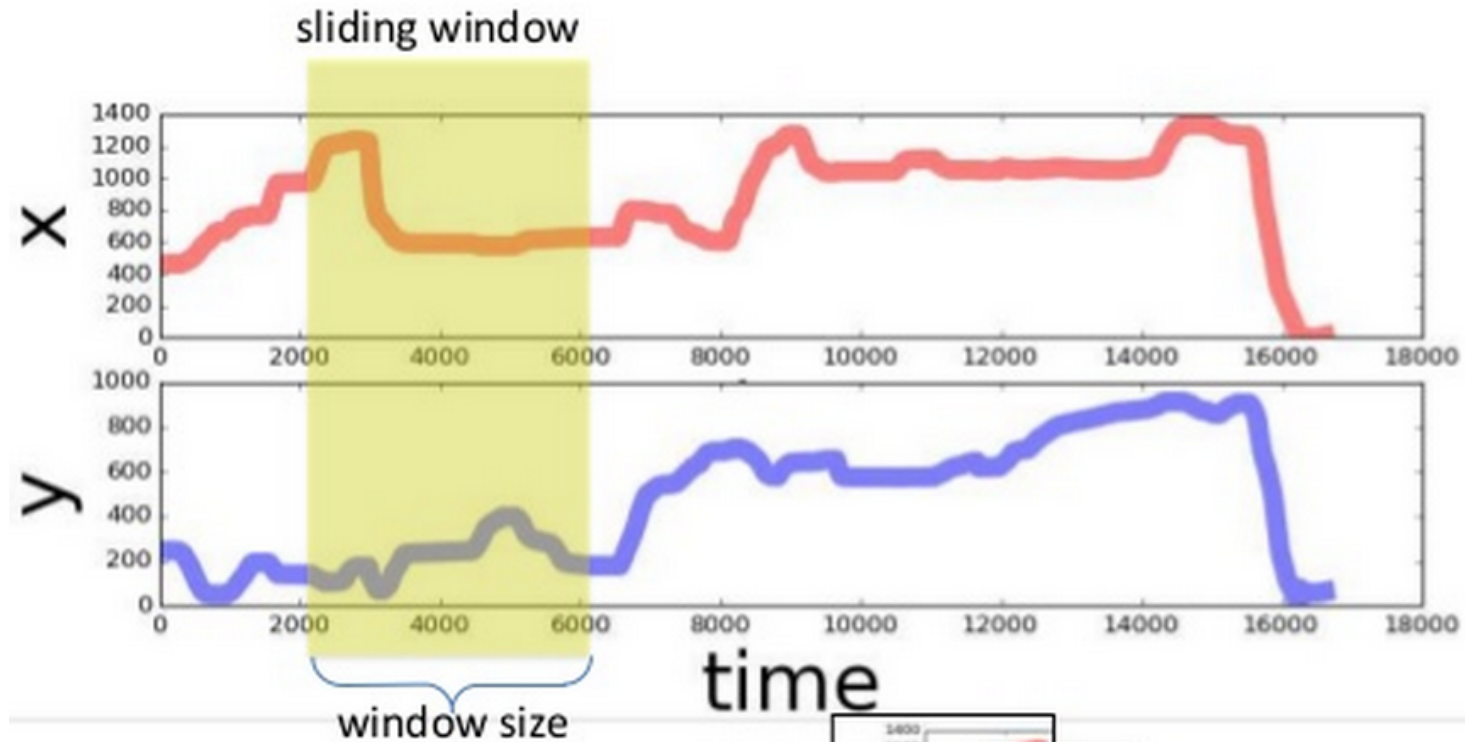


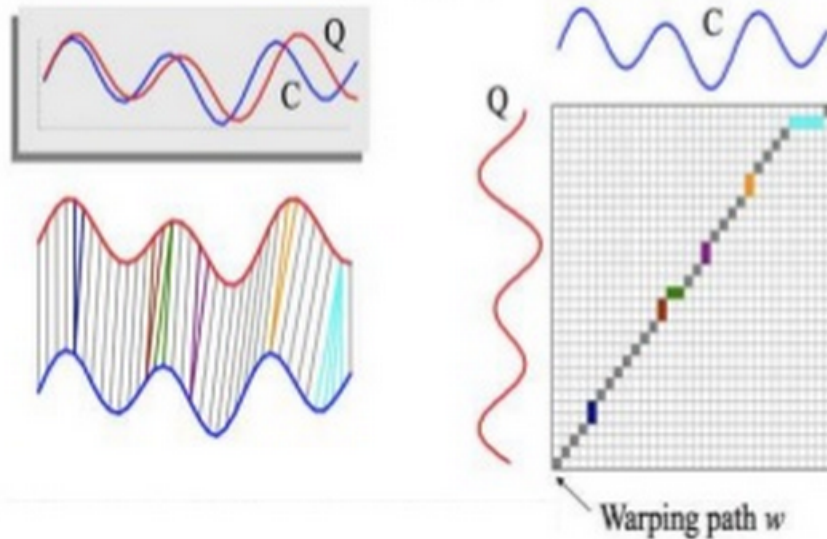
Figure 1: An example automatically discovered motif from mouse cursor data (shaded in green), corresponding to the common “follow” searcher behavior, where gaze (red circles) briefly follows the mouse cursor (blue crosses). The “end” label indicates the result click.

Motifs: Candidate Generation



Motifs: Distance Measure

- Which time series are similar?
- Popular choices:
 - Euclidian Distance
 - Dynamic Time Warping



Common motifs on SERP

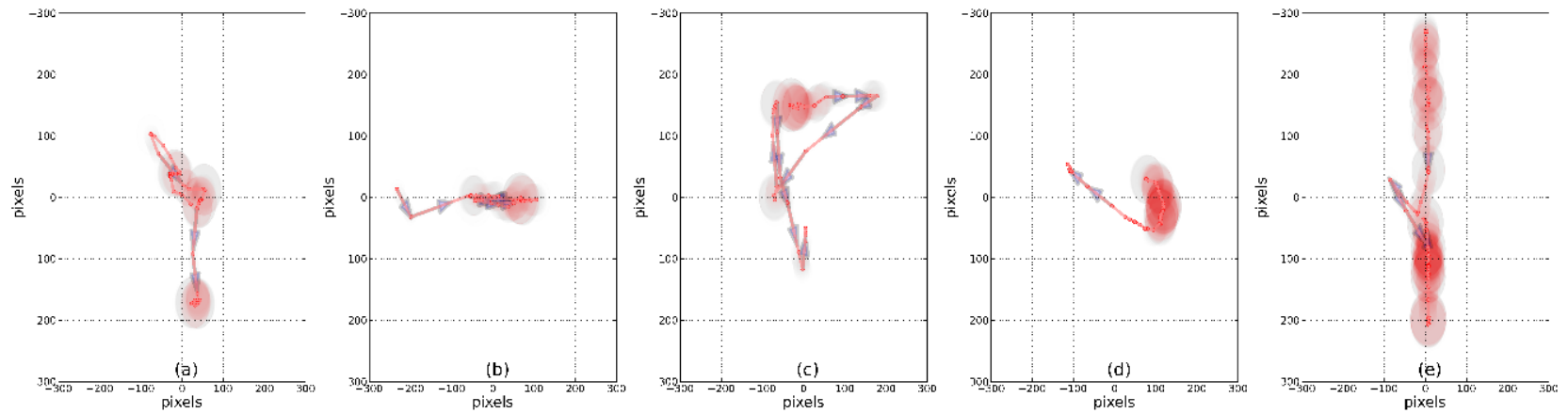


Figure 5: Top frequent motifs discovered from mouse traces recorded on search result pages (SERPs).

Common motifs on non-SERP

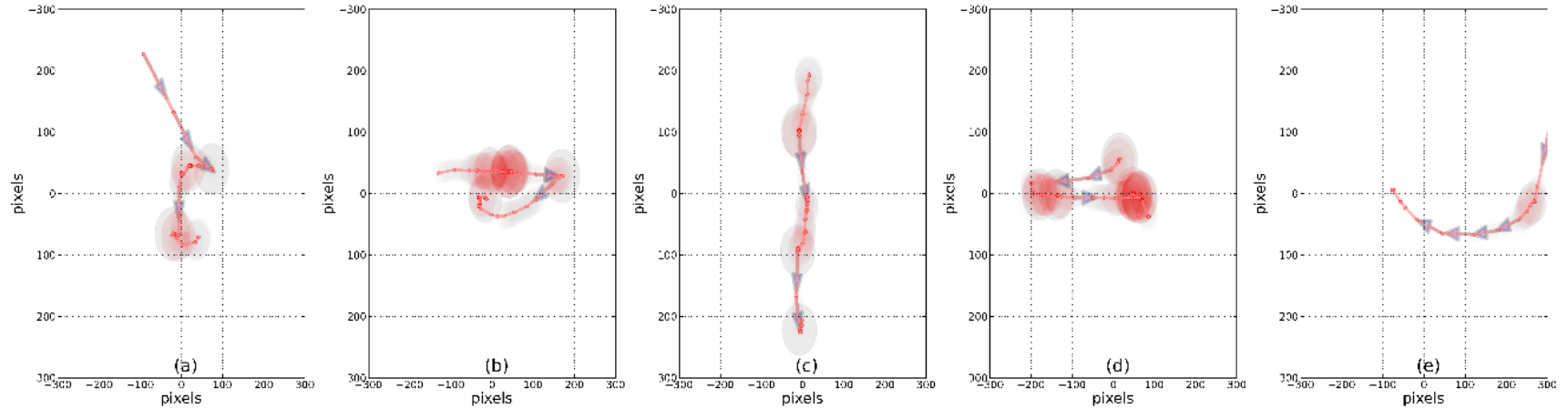


Figure 6: Top frequent motifs discovered from mouse traces recorded on landing pages (non-SERPs).

Mouse movement trails



关于斯特拉马乔尼 1
谁更具体育精神? 不知道他们谁拿到了e19组号 少帅斯特拉马乔尼出生于1976年1月9日, 76年是
个球星辈出的年代, 巴达尼, 托蒂, 舍甫琴科, 罗纳尔多, 范尼, 克卢伊维特, 内斯塔, 维埃拉...
豆瓣 - www.douban.com/2367/topic/28467125/ - 2012-3-27

安德雷·斯特拉马乔尼- 国米防线 6
安德雷·斯特拉马乔尼 (Andrea Stramaccioni) 的档案、数据、新闻、图片等资料。
国米防线球迷俱乐部 - www.inter1908.net/...html - 2013-7-9

国米少帅斯特拉马乔尼去人街头大秀时尚风 国米, 斯特拉马乔尼, 主 7
国米少帅斯特拉马乔尼夫人逛街时遭偷拍, 尽管穿着并不暴露, 但对偷镜新装时仍显得风情万种。
中国青年网 - news.youth.cn/...1214_711338.htm - 2012-12-14

【图】斯特拉马乔尼 国米主帅斯特拉马乔尼 斯特拉马乔尼 姆巴耶 安 8
永清体育门户网站新闻斯特拉马乔尼专区专为您提供详尽的斯特拉马乔尼、国米主帅斯特拉马乔尼、斯
特拉马乔尼 姆巴耶、卡维拉、斯特拉马乔尼、乔尼斯特奥姆特曼、拉马、波拉拉马、...
tyqptzx.com/29531_... - 2013-5-25

斯特拉马乔尼的公共主页- 人人网 renren.com 斯特拉马乔尼 运 12
斯特拉马乔尼, 这里是斯特拉马乔尼 在人人网的公共主页, 来人人网支持斯特拉马乔尼吧! 国际
米兰主帅, 著名教练曾率队夺得国际米兰青年队夺得首届下一代欧冠冠军! ——运...
人人网 - page.renren.com/5041322 - 2012-10-28

斯特拉马乔尼说 百度文库 13
斯特拉马乔尼说:“我了解贾马尔·坦白说, 我相信他(对我的去留)已经有所决定了, 他是个聪明人, 而且
极具才干。”尽管在外界看来, 国际米兰本周中客场不敌帕尔马有可能将...
百度文库 - wenku.baidu.com/view/502812f... - 2012-5-5-

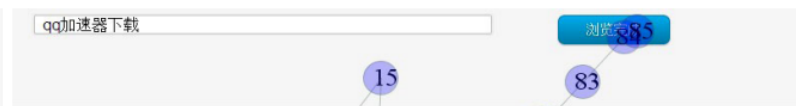
补锅匠拉涅利下课青年队主帅斯特拉马乔尼接任- 豆丁网 15
“补锅匠”拉涅利下课 国米青年队主帅斯特拉马乔尼接任 年仅35岁的斯特拉马乔尼会是国米的
救星吗? (资料图片) 拉涅利没有“补锅”的本领, 却欠缺“大厨”的能力。混...
豆丁网 - www.docin.com/372061268 - 2012-3-28 -

斯特拉马乔尼, 能与曼彻斯特共事我感到很荣幸 体育 腾讯网 18
2012年8月10日... 斯特拉马乔尼谈到国米 (微博 数据) 主席贾马尔·安德烈·斯特拉马乔尼谈到了
他同俱乐部管理层以及和管理层的关系的, 并再次谈到了转会市场。 “我对于同...
腾讯体育 - sports.qq.com/a/20120810/000032.htm - 2012-8-10

斯特拉马乔尼个人资料 斯特拉马乔尼照片 [图] 21
斯特拉马乔尼是一位意大利足球教练, 24岁开始执教国米, 2012年当地时间3月27日, 国际米兰
官方宣布拉涅利下课, 斯特拉马乔尼正式走马上任。 斯特拉马乔尼 中文名: 安德雷...
达州新闻网 - www.dazhou.net/1904/02161578.html - 2012-4-2

斯特拉马乔尼 标签 网易出品 26
2013年9月4日 [继续阅读] 分享到 来源: 网易体育 | 有0人参与 国米少帅斯特拉马乔尼对蓝黑将
士的表现予以肯定, 认为对尤文的比赛球员已经接受了考验, 而提早进入状态对球队出来...
网易体育 - sports.163.com/.6c4e5453c1f1 - 2013-9-4

(a) Example of a satisfied(SAT) search session



QQ等级加速器 QQ下载 82

当前页面：首页 > TAGS > QQ等级加速器 3 个结果已找到 13 类别：QQ软件 授权：免费
更新：恒信QQ等级加速器，通过QQ管家和超级QQ漏洞，采取多线程，无限制，一...

腾牛网 - www.qqtn.com/tags/QQ等级加速器.html - 2013-11-78

4 3

QQ会员网加速QQ会员网加速小助手 V2.0.45.82 官方最新版下

软件简介：Soft4all.com QQ会员网加速小助手 是腾讯公司刚刚发布的 QQ会员网加速器，
该加速器拥有独家智能识别技术，自动匹配最佳服务器，一键加速，从此让QQ游戏...
当下软件园 - www.xiaoyuan.com/nfo/24418.html - 2012-61

QQ网游加速QQ网游加速小助手官方 Beta1 正式版 - 绿色软件

网游加速小助手拥有独家智能识别技术，自动匹配最佳服务器，一键加速，快来下载体验吧。QQ
会员网加速小助手功能详解：1. 启动加速器 2. 立即加速 3. 启动游戏 4. 对比加速...
jz5u绿色下载站 - www.jz5u.com/v_nfo/34419.html - 2012-10-21

8 73

qq加速器官 5.2 下载网游加速小助手 5.2

版本号：2.0.45.82 版本号：2.0.45.82 QQ会员网免费使用网络加速 版本号：2.0.45.82 2年
QQ会员可免费使用网络加速 一键加速，无需自己设置 智能识别 智能选择最佳服务器...
QQ会员 - youxi.vip - www.youxi.vip/jq/2013-11-10

30 28 19 72

QQ等级加速器下载 2013-11-8 等级加速方法 pc6下载站

加速QQ等级的软件，其原理是QQ电脑管家先安装并打开QQTray后关闭QQ账号并登陆QQ30分
钟即可加速一天。适用于工作中不能聊QQ，又想提高QQ等级的朋友 点击右上角下...

pc6下载站 - www.pc6.com/v_SoftView_62335.html - 2012-2-6

QQ游戏加速器下载 软件下载国际速讯网络加速【官方网】

授权形式：文件大小：3.3M 软件语言：简体中文 软件平台：软件类别：下载次数：10 98712
评论等级：更新时间：未知地址：软件大小：未知地址：相关下载 相关下载
网络速讯 - www.exunchi.com/01/down_858.html - 2005-11-30

30 36 68

QQ等级加速器免费下载

QQ等级加速器最新下载 再相信你我也得说，再使用免费QQ挂机类软件时，一定地先申请QQ密码保护。
关于如何申请QQ密码保护，您可以在本站的免费QQ专栏中找到相关文章。QQ...
无忧PTT - www.51ppt.com.cn/v_250015/10/327.html - 2013-11-30

42 66

qq加速器下载

QQ等级加速器 2012 11-22 QQ 软件区 - 天空软件园下载次数：开发商：<http://www.k888.net/jmro>
ng800/Q 软件语言：简体中文 QQ等级加速器 可说是QQ电脑管家 4.4.816.202的最...
www.chinafasten.com/_file/qqsqz/index.htm - 2012-2-24

46

qq加速器下载 qq加速器 1.1 63 qq飞车加速器下载

打开qqctray后关闭QQ账号并登陆QQ30分钟即可加速一天。适用于工作中不能聊QQ又想提高QQ等
级的朋友，您可以在本站的免费QQ会员网加速小助手是腾讯公司刚刚发布的QQ加速器下载
www.tuanchina.com/giare/ma/land/529.html - 2013-10-5

63

qq飞车加速器 游戏区下载官方下载 加速器

腾讯游戏加速器相比其他网络游戏加速器下载 1.1.9.50 1. 腾讯官方开发，对QQ飞车加速效果
更好！ 2. 与游戏客户端完美整合，无需另行下载！ 3. 对游戏客户端自动根据现...
凯洛特 - down.klss.cn/game/26729.htm - 2013-10-19

(b) Example of a dissatisfied(DSAT) search session

Figure 1: Examples of Users' Mouse Movement Trails on SERPs

Predictive motifs

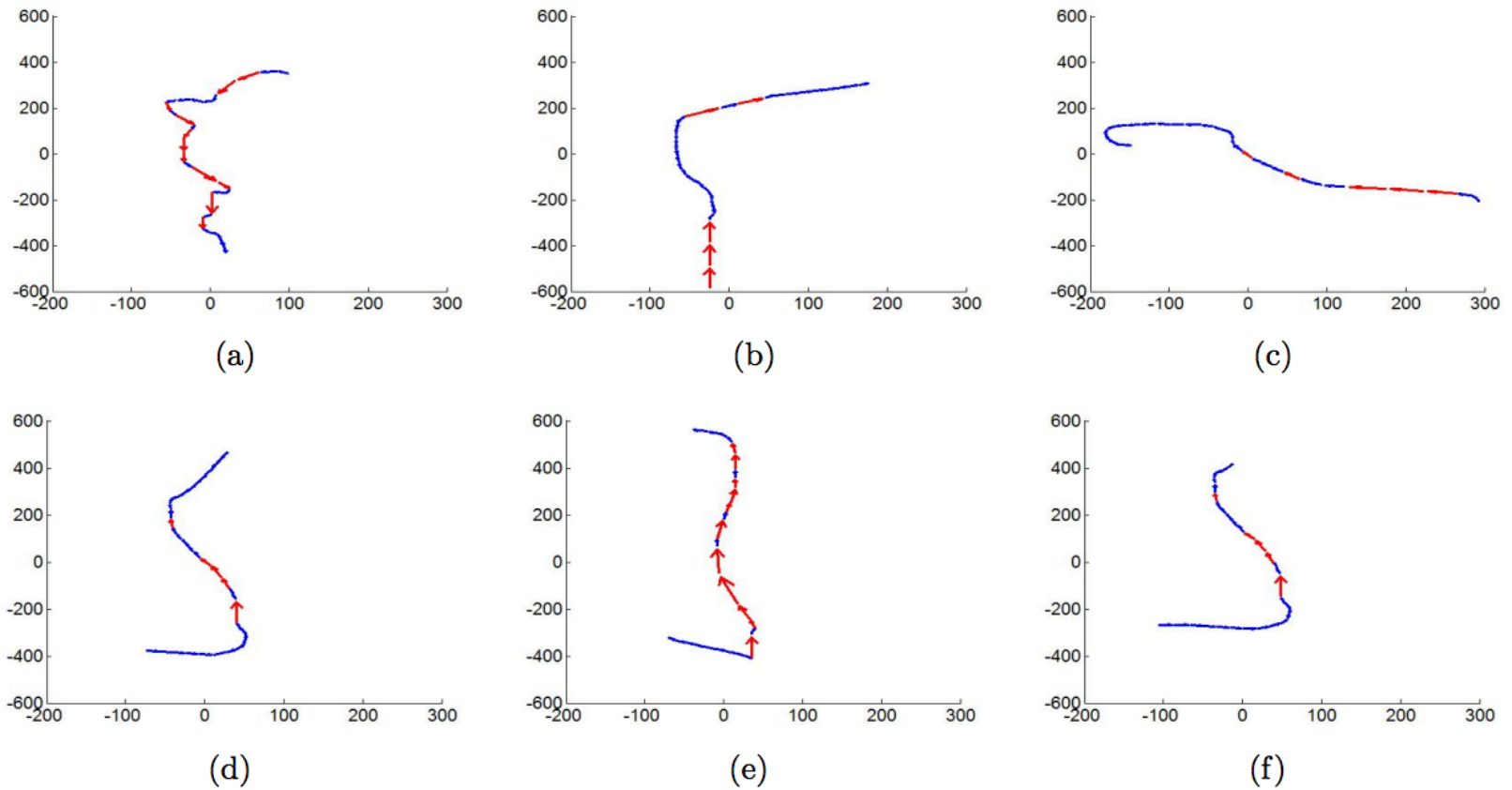
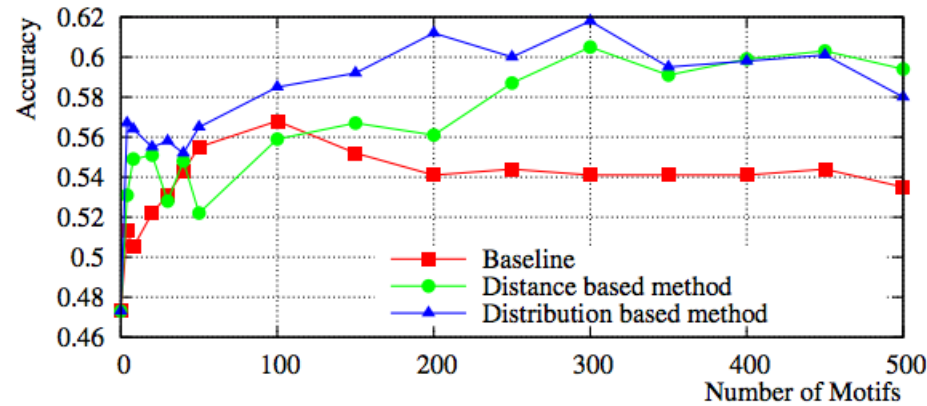
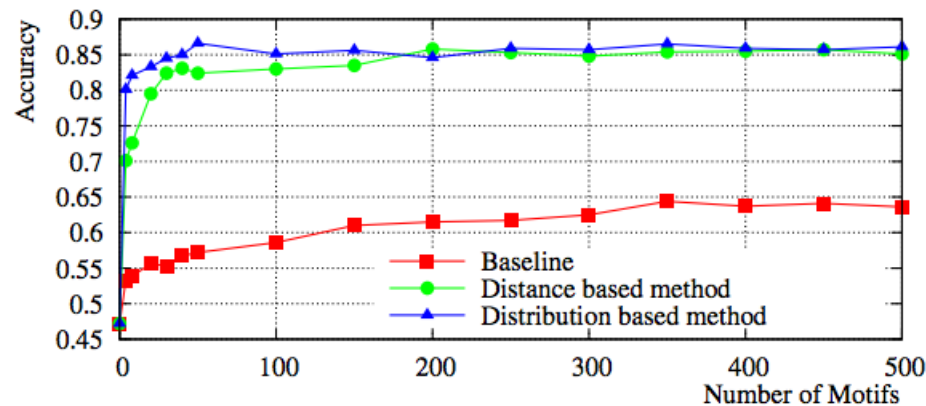


Figure 5: Predictive motifs discovered from *SAT_DATA* (a-c) and *DSAT_DATA* (d-f)

Motifs: Predicting Satisfaction



(a) Users' Annotations



(b) Assessors' Annotations

Figure 6: Prediction Performance with Different Motif Selection Strategies

Beyond SERP: Satisfaction & Engagement

- Common measures
 - Avg unique queries per session [S]
 - Avg session length per user [S]
 - Avg query success rate per user [S]
 - Avg query CTR [S]
 - Average query interval per user [S]
 - Avg daily sessions per user [E]
 - Absense Time [E]

User Engagement Analysis

Evaluating and Predicting User Engagement Change with Degraded Search Relevance

Yang Song
Microsoft Research
One Microsoft Way
Redmond, WA

yangsong@microsoft.com

Xiaolin Shi
Microsoft Bing
One Microsoft Way
Redmond, WA

xishi@microsoft.com

Xin Fu^{*}
LinkedIn Corporation
2029 Stierlin Court
Mountain View, CA

xin.fu.2007@gmail.com

On Correlation of Absence Time and Search Effectiveness

Sunandan Chakraborty^{*}
New York University
New York, USA
sunandan@cs.nyu.edu

Filip Radlinski
Microsoft
Cambridge, UK

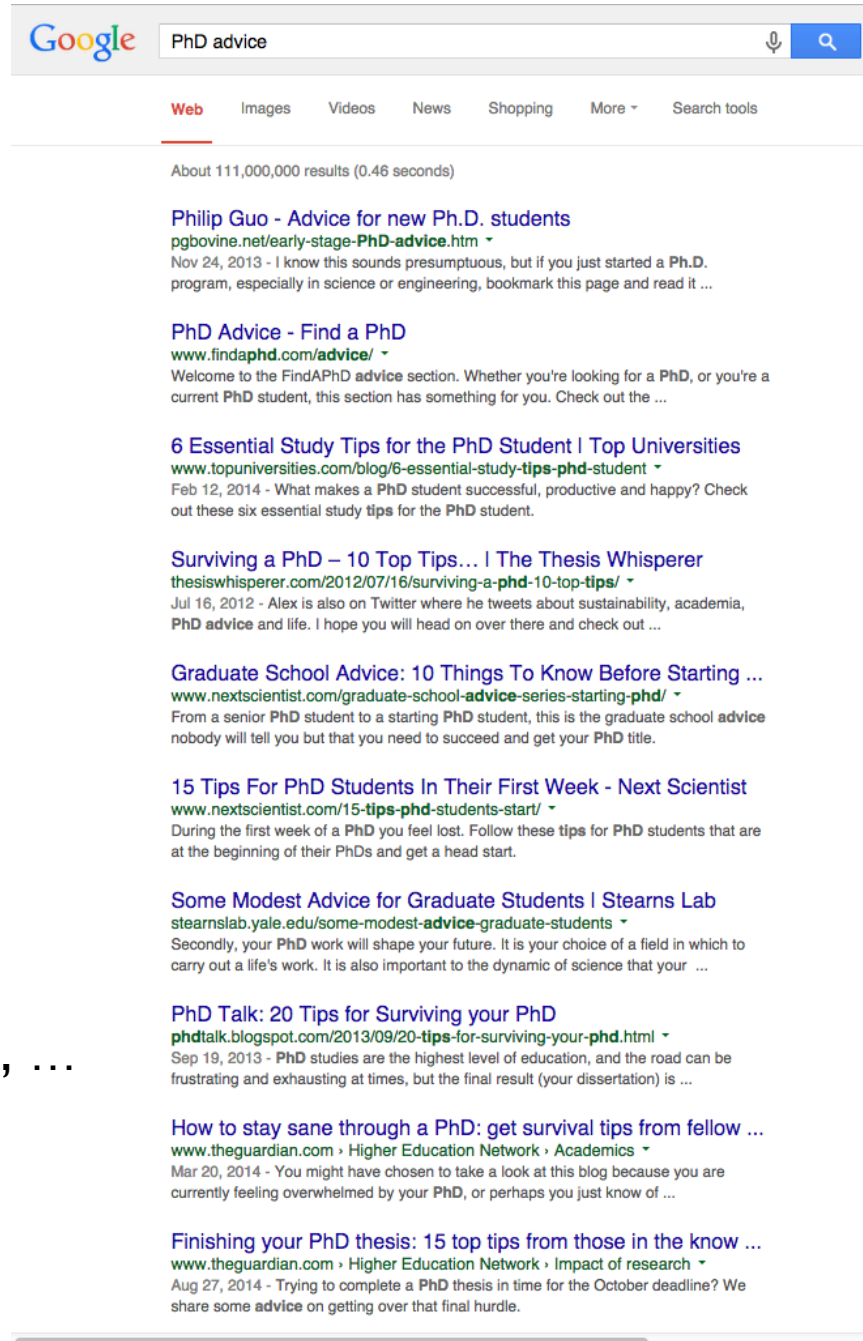
Milad Shokouhi
Microsoft
Cambridge, UK

Paul Baecke
Microsoft
London, UK

{filiprad, milads, pbaecke}@microsoft.com

Different User Signal

Search Engine Result Page (SERP)



- Clicks
- Mouse movement
- Browser action
 - bookmark, save, print
- Time
 - dwell time, time on SERP
- Explicit judgment
 - likes, favourites..
- Other page elements
 - share, ...
- Long term effects
 - sessions per user, abandonment, ...
- Reformulations

Online Evaluation Designs

1. Document Level or Ranking Level?

Document Level	Ranking Level
<p>I want to know about the <i>documents</i></p> <p>Similar to the collection-based approach, I'd like to find out the quality of each document.</p>	<p>I am mostly interested in the <i>rankings</i></p> <p>I'm trying to evaluate retrieval functions. I don't need to be able to drill down to individual documents.</p>

2. Absolute or Relative?

Absolute Judgments	Relative Judgments
<p>I want a score on an absolute scale</p> <p>Similar to the Cranfield approach, I'd like a number that I can compare to many methods, over time.</p>	<p>I am mostly interested in a comparison</p> <p>It's enough if I know which document, or which ranking, is better. Its not necessary to know the absolute value.</p>

Interpreting Clicks

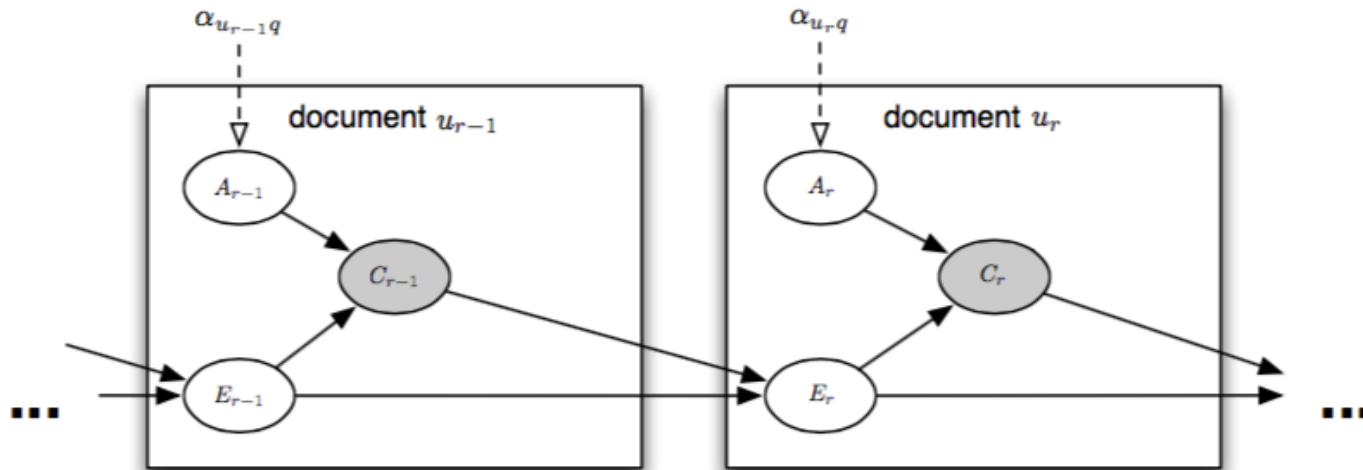
	Absolute	Relative
Item level	Click rate ...	Click-Skip ...
SERP level	Abandonment ...	A/B testing, Interleaving

Interpreting Clicks

	Absolute	Relative
Item level	Click rate ...	Click-Skip ...
SERP level	Abandonment ...	A/B testing, Interleaving

Modeling user behavior

- Straightforward interpretation of clicks
 - Use click-through rate
 - May be **biased**
- Can absolute **document relevance** be recovered



Interpreting Clicks

	Absolute	Relative
Item level	Click rate ...	Click-Skip ...
SERP level	Abandonment ...	A/B testing, interleaving

Document Level Preferences

The image shows a Google search results page for the query "PhD advice". The browser's address bar shows the URL "https://www.google.nl/webhp?sourceid=chrome-instant&ion...". The search bar contains the text "PhD advice". Below the search bar, there are tabs for "Web", "Images", "Videos", "News", "Shopping", "More", and "Search tools". The search results are displayed in a list format. The first result is "Philip Guo - Advice for new Ph.D. students" from "pgbovine.net/early-stage-PhD-advice.htm". The second result is "PhD Advice - Find a PhD" from "www.findaphd.com/advice/". The third result is "6 Essential Study Tips for the PhD Student | Top Universities" from "www.topuniversities.com/blog/6-essential-study-tips-phd-student". The fourth result is "Surviving a PhD – 10 Top Tips... | The Thesis Whisperer" from "thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/". The fifth result is "Graduate School Advice: 10 Things To Know Before Starting ..." from "www.nextscientist.com/graduate-school-advice-series-starting-phd/". The sixth result is "15 Tips For PhD Students In Their First Week - Next Scientist" from "www.nextscientist.com/15-tips-phd-students-start/". The seventh result is "Some Modest Advice for Graduate Students | Stearns Lab" from "stearnslab.yale.edu/some-modest-advice-graduate-students". The eighth result is "PhD Talk: 20 Tips for Surviving your PhD" from "phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html". The ninth result is "How to stay sane through a PhD: get survival tips from fellow ..." from "www.theguardian.com". The tenth result is "Finishing your PhD thesis: 15 top tips from those in the know ..." from "www.theguardian.com".

1 Philip Guo - Advice for new Ph.D. students
pgbovine.net/early-stage-PhD-advice.htm
Nov 24, 2013 - I know this sounds presumptuous, but if you just started a Ph.D. program, especially in science or engineering, bookmark this page and read it ...

2 PhD Advice - Find a PhD
www.findaphd.com/advice/
Welcome to the FindAPhD advice section. Whether you're looking for a PhD, or you're a current PhD student, this section has something for you. Check out the ...

3 6 Essential Study Tips for the PhD Student | Top Universities
www.topuniversities.com/blog/6-essential-study-tips-phd-student
Feb 12, 2014 - What makes a PhD student successful, productive and happy? Check out these six essential study tips for the PhD student.

4 Surviving a PhD – 10 Top Tips... | The Thesis Whisperer
thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/
Jul 16, 2012 - Alex is also on Twitter where he tweets about sustainability, academia, PhD advice and life. I hope you will head on over there and check out ...

1 Graduate School Advice: 10 Things To Know Before Starting ...
www.nextscientist.com/graduate-school-advice-series-starting-phd/
From a senior PhD student to a starting PhD student, this is the graduate school advice nobody will tell you but that you need to succeed and get your PhD title.

5 15 Tips For PhD Students In Their First Week - Next Scientist
www.nextscientist.com/15-tips-phd-students-start/
During the first week of a PhD you feel lost. Follow these tips for PhD students that are at the beginning of their PhDs and get a head start.

6 Some Modest Advice for Graduate Students | Stearns Lab
stearnslab.yale.edu/some-modest-advice-graduate-students
Secondly, your PhD work will shape your future. It is your choice of a field in which to carry out a life's work. It is also important to the dynamic of science that you ...

PhD Talk: 20 Tips for Surviving your PhD
phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html
Sep 19, 2013 - PhD studies are the highest level of education, and the road can be frustrating and exhausting at times, but the final result (your dissertation) is ...

How to stay sane through a PhD: get survival tips from fellow ...
www.theguardian.com › Higher Education Network › Academics
Mar 20, 2014 - You might have chosen to take a look at this blog because you are currently feeling overwhelmed by your PhD, or perhaps you just know of ...

Finishing your PhD thesis: 15 top tips from those in the know ...
www.theguardian.com › Higher Education Network › Impact of research
Aug 27, 2014 - Trying to complete a PhD thesis in time for the October deadline? We share some advice on getting over that final hurdle.

Click > Skip Heuristics

- **CLICK > SKIP ABOVE**
- LAST CLICK > SKIP ABOVE
- CLICK > EARLIER CLICK
- LAST CLICK > SKIP PREVIOUS
- CLICK > NO-CLICK NEXT

The screenshot shows a Google search for "PhD advice" with approximately 111,000,000 results. The search results are annotated with large blue numbers 1 through 6, indicating a sequence of clicks. Annotations 1 and 2 are enclosed in blue rounded rectangles.

1 Philip Guo - Advice for new Ph.D. students
pgbovine.net/early-stage-PhD-advice.htm
Nov 24, 2013 - I know this sounds presumptuous, but if you just started a Ph.D. program, especially in science or engineering, bookmark this page and read it ...

2 PhD Advice - Find a PhD
www.findaphd.com/advice/
Welcome to the FindAPhD advice section. Whether you're looking for a PhD, or you're a current PhD student, this section has something for you. Check out the ...

3 6 Essential Study Tips for the PhD Student | Top Universities
www.topuniversities.com/blog/6-essential-study-tips-phd-student
Feb 12, 2014 - What makes a PhD student successful, productive and happy? Check out these six essential study tips for the PhD student.

4 Surviving a PhD – 10 Top Tips... | The Thesis Whisperer
thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/
Jul 16, 2012 - Alex is also on Twitter where he tweets about sustainability, academia, PhD advice and life. I hope you will head on over there and check out ...

1 Graduate School Advice: 10 Things To Know Before Starting ...
www.nextscientist.com/graduate-school-advice-series-starting-phd/
From a senior PhD student to a starting PhD student, this is the graduate school advice nobody will tell you but that you need to succeed and get your PhD title.

5 15 Tips For PhD Students In Their First Week - Next Scientist
www.nextscientist.com/15-tips-phd-students-start/
During the first week of a PhD you feel lost. Follow these tips for PhD students that are at the beginning of their PhDs and get a head start.

6 Some Modest Advice for Graduate Students | Stearns Lab
stearnslab.yale.edu/some-modest-advice-graduate-students
Secondly, your PhD work will shape your future. It is your choice of a field in which to carry out a life's work. It is also important to the dynamic of science that your ...

PhD Talk: 20 Tips for Surviving your PhD
phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html
Sep 19, 2013 - PhD studies are the highest level of education, and the road can be frustrating and exhausting at times, but the final result (your dissertation) is ...

How to stay sane through a PhD: get survival tips from fellow ...
www.theguardian.com › Higher Education Network › Academics
Mar 20, 2014 - You might have chosen to take a look at this blog because you are currently feeling overwhelmed by your PhD, or perhaps you just know of ...

Finishing your PhD thesis: 15 top tips from those in the know ...
www.theguardian.com › Higher Education Network › Impact of research
Aug 27, 2014 - Trying to complete a PhD thesis in time for the October deadline? We share some advice on getting over that final hurdle.

Click > Skip Heuristics

- CLICK > SKIP ABOVE
- **LAST CLICK > SKIP ABOVE**
- CLICK > EARLIER CLICK
- LAST CLICK > SKIP PREVIOUS
- CLICK > NO-CLICK NEXT

The screenshot shows a Google search for "PhD advice" with approximately 111,000,000 results. The search results are listed on the right side of the page. On the left side of the results, there are blue circles with numbers 1, 2, and 1, indicating specific heuristics. On the right side, there are large blue numbers 1 through 6, indicating a ranking or sequence. The search results include:

- Philip Guo - Advice for new Ph.D. students**
pgbovine.net/early-stage-PhD-advice.htm
Nov 24, 2013 - I know this sounds presumptuous, but if you just started a Ph.D. program, especially in science or engineering, bookmark this page and read it ...
- PhD Advice - Find a PhD**
www.findaphd.com/advice/
Welcome to the FindAPhD advice section. Whether you're looking for a PhD, or you're a current PhD student, this section has something for you. Check out the ...
- 6 Essential Study Tips for the PhD Student | Top Universities**
www.topuniversities.com/blog/6-essential-study-tips-phd-student
Feb 12, 2014 - What makes a PhD student successful, productive and happy? Check out these six essential study tips for the PhD student.
- Surviving a PhD – 10 Top Tips... | The Thesis Whisperer**
thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/
Jul 16, 2012 - Alex is also on Twitter where he tweets about sustainability, academia, PhD advice and life. I hope you will head on over there and check out ...
- Graduate School Advice: 10 Things To Know Before Starting ...**
www.nextscientist.com/graduate-school-advice-series-starting-phd/
From a senior PhD student to a starting PhD student, this is the graduate school advice nobody will tell you but that you need to succeed and get your PhD title.
- 15 Tips For PhD Students In Their First Week - Next Scientist**
www.nextscientist.com/15-tips-phd-students-start/
During the first week of a PhD you feel lost. Follow these tips for PhD students that are at the beginning of their PhDs and get a head start.
- Some Modest Advice for Graduate Students | Stearns Lab**
stearnslab.yale.edu/some-modest-advice-graduate-students
Secondly, your PhD work will shape your future. It is your choice of a field in which to carry out a life's work. It is also important to the dynamic of science that your ...
- PhD Talk: 20 Tips for Surviving your PhD**
phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html
Sep 19, 2013 - PhD studies are the highest level of education, and the road can be frustrating and exhausting at times, but the final result (your dissertation) is ...
- How to stay sane through a PhD: get survival tips from fellow ...**
www.theguardian.com › Higher Education Network › Academics
Mar 20, 2014 - You might have chosen to take a look at this blog because you are currently feeling overwhelmed by your PhD, or perhaps you just know of ...
- Finishing your PhD thesis: 15 top tips from those in the know ...**
www.theguardian.com › Higher Education Network › Impact of research
Aug 27, 2014 - Trying to complete a PhD thesis in time for the October deadline? We share some advice on getting over that final hurdle.

Click > Skip Heuristics

- CLICK > SKIP ABOVE
- LAST CLICK > SKIP ABOVE
- **CLICK > EARLIER CLICK**
- LAST CLICK > SKIP PREVIOUS
- CLICK > NO-CLICK NEXT

The screenshot shows a Google search for "PhD advice" with approximately 111,000,000 results. The search results are annotated with large blue numbers 1 through 6, indicating a sequence of clicks. Annotations 1 and 2 are enclosed in blue rounded rectangles.

1 Philip Guo - Advice for new Ph.D. students
pgbovine.net/early-stage-PhD-advice.htm
Nov 24, 2013 - I know this sounds presumptuous, but if you just started a Ph.D. program, especially in science or engineering, bookmark this page and read it ...

2 PhD Advice - Find a PhD
www.findaphd.com/advice/
Welcome to the FindAPhD advice section. Whether you're looking for a PhD, or you're a current PhD student, this section has something for you. Check out the ...

3 6 Essential Study Tips for the PhD Student | Top Universities
www.topuniversities.com/blog/6-essential-study-tips-phd-student
Feb 12, 2014 - What makes a PhD student successful, productive and happy? Check out these six essential study tips for the PhD student.

4 Surviving a PhD – 10 Top Tips... | The Thesis Whisperer
thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/
Jul 16, 2012 - Alex is also on Twitter where he tweets about sustainability, academia, PhD advice and life. I hope you will head on over there and check out ...

1 Graduate School Advice: 10 Things To Know Before Starting ...
www.nextscientist.com/graduate-school-advice-series-starting-phd/
From a senior PhD student to a starting PhD student, this is the graduate school advice nobody will tell you but that you need to succeed and get your PhD title.

5 15 Tips For PhD Students In Their First Week - Next Scientist
www.nextscientist.com/15-tips-phd-students-start/
During the first week of a PhD you feel lost. Follow these tips for PhD students that are at the beginning of their PhDs and get a head start.

6 Some Modest Advice for Graduate Students | Stearns Lab
stearnslab.yale.edu/some-modest-advice-graduate-students
Secondly, your PhD work will shape your future. It is your choice of a field in which to carry out a life's work. It is also important to the dynamic of science that your ...

PhD Talk: 20 Tips for Surviving your PhD
phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html
Sep 19, 2013 - PhD studies are the highest level of education, and the road can be frustrating and exhausting at times, but the final result (your dissertation) is ...

How to stay sane through a PhD: get survival tips from fellow ...
www.theguardian.com › Higher Education Network › Academics
Mar 20, 2014 - You might have chosen to take a look at this blog because you are currently feeling overwhelmed by your PhD, or perhaps you just know of ...

Finishing your PhD thesis: 15 top tips from those in the know ...
www.theguardian.com › Higher Education Network › Impact of research
Aug 27, 2014 - Trying to complete a PhD thesis in time for the October deadline? We share some advice on getting over that final hurdle.

Click > Skip Heuristics

- CLICK > SKIP ABOVE
- LAST CLICK > SKIP ABOVE
- CLICK > EARLIER CLICK
- **LAST CLICK > SKIP PREVIOUS**
- CLICK > NO-CLICK NEXT

The screenshot shows a Google search for "PhD advice" with approximately 111,000,000 results. The search results are annotated with large blue numbers 1 through 6, indicating a sequence of clicks. Annotations 1 and 2 are circled in blue.

1 Philip Guo - Advice for new Ph.D. students
pgbovine.net/early-stage-PhD-advice.htm
Nov 24, 2013 - I know this sounds presumptuous, but if you just started a Ph.D. program, especially in science or engineering, bookmark this page and read it ...

2 PhD Advice - Find a PhD
www.findaphd.com/advice/
Welcome to the FindAPhD advice section. Whether you're looking for a PhD, or you're a current PhD student, this section has something for you. Check out the ...

3 6 Essential Study Tips for the PhD Student | Top Universities
www.topuniversities.com/blog/6-essential-study-tips-phd-student
Feb 12, 2014 - What makes a PhD student successful, productive and happy? Check out these six essential study tips for the PhD student.

4 Surviving a PhD – 10 Top Tips... | The Thesis Whisperer
thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/
Jul 16, 2012 - Alex is also on Twitter where he tweets about sustainability, academia, PhD advice and life. I hope you will head on over there and check out ...

1 Graduate School Advice: 10 Things To Know Before Starting ...
www.nextscientist.com/graduate-school-advice-series-starting-phd/
From a senior PhD student to a starting PhD student, this is the graduate school advice nobody will tell you but that you need to succeed and get your PhD title.

5 15 Tips For PhD Students In Their First Week - Next Scientist
www.nextscientist.com/15-tips-phd-students-start/
During the first week of a PhD you feel lost. Follow these tips for PhD students that are at the beginning of their PhDs and get a head start.

6 Some Modest Advice for Graduate Students | Stearns Lab
stearnslab.yale.edu/some-modest-advice-graduate-students
Secondly, your PhD work will shape your future. It is your choice of a field in which to carry out a life's work. It is also important to the dynamic of science that your ...

PhD Talk: 20 Tips for Surviving your PhD
phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html
Sep 19, 2013 - PhD studies are the highest level of education, and the road can be frustrating and exhausting at times, but the final result (your dissertation) is ...

How to stay sane through a PhD: get survival tips from fellow ...
www.theguardian.com › Higher Education Network › Academics
Mar 20, 2014 - You might have chosen to take a look at this blog because you are currently feeling overwhelmed by your PhD, or perhaps you just know of ...

Finishing your PhD thesis: 15 top tips from those in the know ...
www.theguardian.com › Higher Education Network › Impact of research
Aug 27, 2014 - Trying to complete a PhD thesis in time for the October deadline? We share some advice on getting over that final hurdle.

Click > Skip Heuristics

- CLICK > SKIP ABOVE
- LAST CLICK > SKIP ABOVE
- CLICK > EARLIER CLICK
- LAST CLICK > SKIP PREVIOUS
- **CLICK > NO-CLICK NEXT**

The screenshot shows a Google search for "PhD advice" with approximately 111,000,000 results. The search results are annotated with large blue numbers 1 through 6, indicating a sequence of clicks. Annotations 1 and 2 are circled in blue.

1 Philip Guo - Advice for new Ph.D. students
pgbovine.net/early-stage-PhD-advice.htm
Nov 24, 2013 - I know this sounds presumptuous, but if you just started a Ph.D. program, especially in science or engineering, bookmark this page and read it ...

2 PhD Advice - Find a PhD
www.findaphd.com/advice/
Welcome to the FindAPhD advice section. Whether you're looking for a PhD, or you're a current PhD student, this section has something for you. Check out the ...

3 6 Essential Study Tips for the PhD Student | Top Universities
www.topuniversities.com/blog/6-essential-study-tips-phd-student
Feb 12, 2014 - What makes a PhD student successful, productive and happy? Check out these six essential study tips for the PhD student.

4 Surviving a PhD – 10 Top Tips... | The Thesis Whisperer
thesiswhisperer.com/2012/07/16/surviving-a-phd-10-top-tips/
Jul 16, 2012 - Alex is also on Twitter where he tweets about sustainability, academia, PhD advice and life. I hope you will head on over there and check out ...

1 Graduate School Advice: 10 Things To Know Before Starting ...
www.nextscientist.com/graduate-school-advice-series-starting-phd/
From a senior PhD student to a starting PhD student, this is the graduate school advice nobody will tell you but that you need to succeed and get your PhD title.

5 15 Tips For PhD Students In Their First Week - Next Scientist
www.nextscientist.com/15-tips-phd-students-start/
During the first week of a PhD you feel lost. Follow these tips for PhD students that are at the beginning of their PhDs and get a head start.

6 Some Modest Advice for Graduate Students | Stearns Lab
stearnslab.yale.edu/some-modest-advice-graduate-students
Secondly, your PhD work will shape your future. It is your choice of a field in which to carry out a life's work. It is also important to the dynamic of science that you ...

PhD Talk: 20 Tips for Surviving your PhD
phdtalk.blogspot.com/2013/09/20-tips-for-surviving-your-phd.html
Sep 19, 2013 - PhD studies are the highest level of education, and the road can be frustrating and exhausting at times, but the final result (your dissertation) is ...

How to stay sane through a PhD: get survival tips from fellow ...
www.theguardian.com › Higher Education Network › Academics
Mar 20, 2014 - You might have chosen to take a look at this blog because you are currently feeling overwhelmed by your PhD, or perhaps you just know of ...

Finishing your PhD thesis: 15 top tips from those in the know ...
www.theguardian.com › Higher Education Network › Impact of research
Aug 27, 2014 - Trying to complete a PhD thesis in time for the October deadline? We share some advice on getting over that final hurdle.

Click > Skip Heuristics – Evaluation

Evaluation against explicit manual preference judgments:

Method	Accuracy
Click > Skip Above	78.2 ± 5.6
Last Click > Skip Above	80.9 ± 5.1
Click > Earlier Click	64.3 ± 15.4
Click > Skip Previous	80.7 ± 9.6
Click > No Click Next	67.4 ± 8.2
Inter-Judge Agreement	86.4

- High accuracy (up to 80%)
- May suffer from position bias

Interpreting Clicks

	Absolute	Relative
Item level	Click rate ...	Click-Skip ...
SERP level	Abandonment ...	A/B testing, interleaving

Absolute SERP Quality

- Document-level feedback requires converting judgments to evaluation metric (of a ranking)
- Ranking-level judgments directly define such a metric

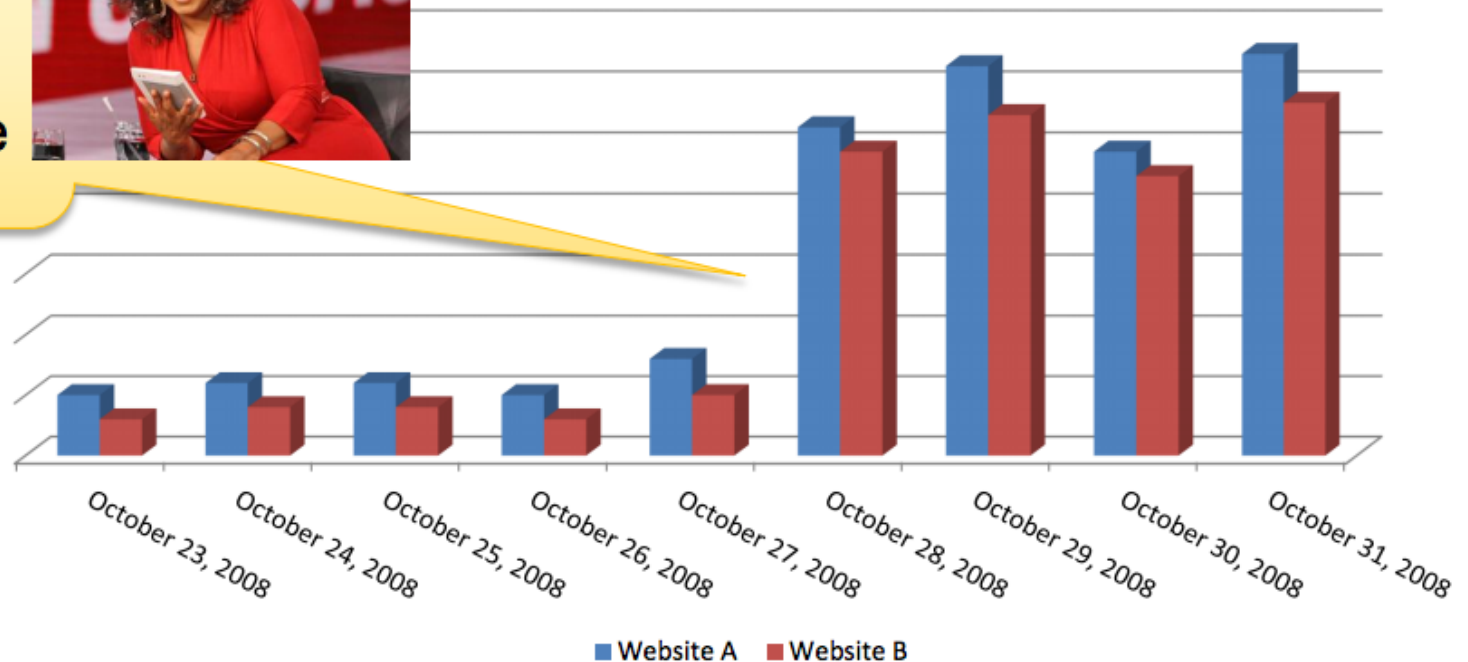
Some Absolute Metrics	
Abandonment Rate	Reformulation Rate
Queries per Session	Clicks per Query
Click rate on first result	Max Reciprocal Rank
Time to first click	Time to last click
% of viewed documents skipped (pSkip)	

Compare against historical data

Oprah calls
Kindle "her
new favorite
thing"



Amazon Kindle Sales



Interpreting Clicks

	Absolute	Relative
Item level	Click rate ...	Click-Skip ...
SERP level	Abandonment ...	A/B testing, interleaving

In-situ evaluation in one slide

- See how normal **users interact** with your **live search engine** when just using it
- Observe **implicit behavior**
 - Clicks, skips, saves, forwards, bookmarks, “likes”, etc.
- Try to **infer differences** in behavior from different flavors of the live system
 - A/B testing
 - Have x% of query traffic use system A and y% of query traffic use system B
 - Interleaving
 - Expose a combination of system versions to users

4. A/B Testing

A/B Testing

Baseline (control)

Ads related to **valencia spain** ⓘ

250 Hotels in Valencia - Lowest price guarantee - booking.com
www.booking.com/Valencia-Hotels ★★★★★ 59,565 seller reviews
Book your Hotel in Valencia online
40 people in Zurich, Switzerland +1'd this
Hotels in Eixample - Bioparc Valencia Hotels - Hotels near Oceanografic

Valencia Spain: Beaches - visitvalencia.com
www.visitvalencia.com/ ▶
Guide to the beaches of Valencia Enjoy everything!

Valencia Tourism Official Site | Tourist Info in Valencia Spain
www.visitvalencia.com/en/home ▶
2 days ago - Valencia's tourist information in one place. What to see, special discounts and restaurant promos. Find all the information you need for visiting Valencia.

Valencia, Spain Travel Guide - Must-See Attractions - YouTube
www.youtube.com/watch?v=_o9jZrj42A
Apr 26, 2013 - Uploaded by BookingHunterTV
<http://bookinghunter.com> Valencia is the third largest city in Spain after Madrid and Barcelona. Valencia is ...

Tomatina 2013 - Valencia, Spain | Europe | Travel | Toronto Sun
www.torontosun.com/2013/08/28/tomatina-2013---valencia-spain ▶
Aug 28, 2013 - The origin of the Tomatina tomato fight is disputed - everyone in Bunol seems to have a favourite story - but most agree it started around 1940, in the early years ...

Images for valencia spain - Report images



Experimental (treatment)

Ad related to **valencia spain** ⓘ

250 Hotels in Valencia - Lowest price guarantee - booking.com
www.booking.com/Valencia-Hotels ★★★★★ 59,565 seller reviews
Book your Hotel in Valencia online
40 people in Zurich, Switzerland +1'd this
Hotels in Eixample - Bioparc Valencia Hotels - Hotels near Oceanografic

Valencia - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Valencia ▶
Valencia (Spanish: [baˈleŋja]), or València (Valencian: [vaˈlensja]), is the capital of the autonomous community of Valencia and the third largest city in Spain ...
Valencia (disambiguation) - Valencia CF - Valencian Community - Valencian

Valencia Tourism and Vacations: 219 Things to Do in ... - TripAdvisor
www.tripadvisor.com > ... > Valencia Province ▶
Jan 21, 2013
Valencia Tourism: TripAdvisor has 56831 reviews of Valencia Hotels, Attractions, and Restaurants ... Flights to ...

Images for valencia spain - Report images



Valencia Tourism Official Site | Tourist Info in Valencia Spain
www.visitvalencia.com/en/home ▶
Valencia's tourist information in one place. What to see, special discounts and restaurant promos. Find all the information you need for visiting Valencia.

Tourism in Valencia, Spain | Spain.info in english
www.spain.info/en/que-quieren/ciudades-pueblos.../valencia.html ▶
Spain.info in english: Information on Valencia in Spain. Holidayss in Valencia. Sights in Valencia in Spain, accommodation in Valencia, events in Valencia and ...



Valencia

City in Spain

Valencia, or València, is the capital of the autonomous community of Valencia and the third largest city in Spain after Madrid and Barcelona, with around 808,000 inhabitants in the administrative centre. Wikipedia

Area: 51.99 sq miles (134.6 km²)

Weather: 26°C, Wind N at 6 km/h, 72% Humidity

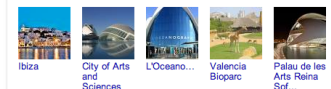
Local time: Sunday 12:48 PM

Province: Province of Valencia

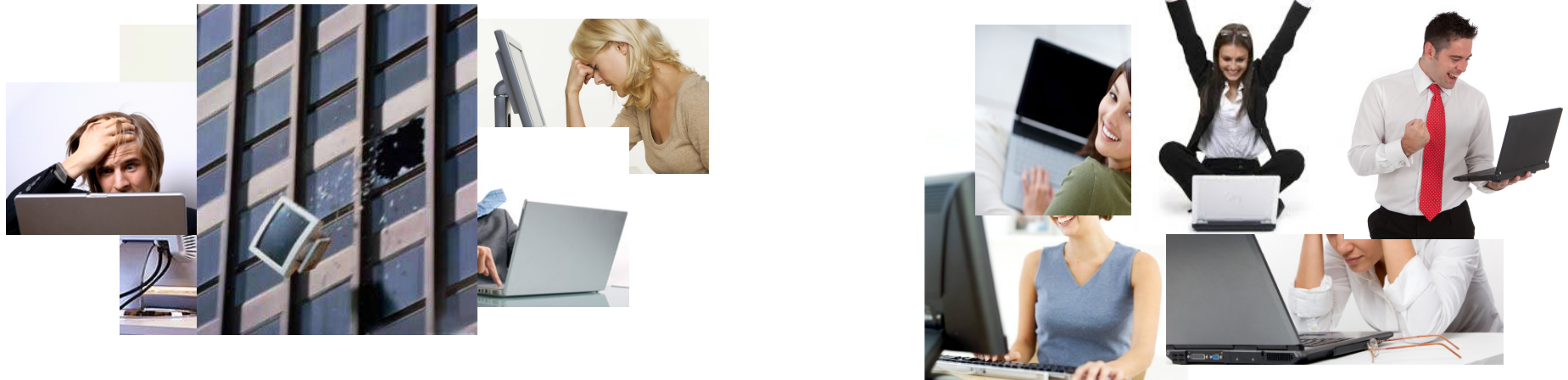
Population: 797,028 (2012) Instituto Nacional de Estadística

Colleges and Universities: Universitat de València, Møre

Points of interest

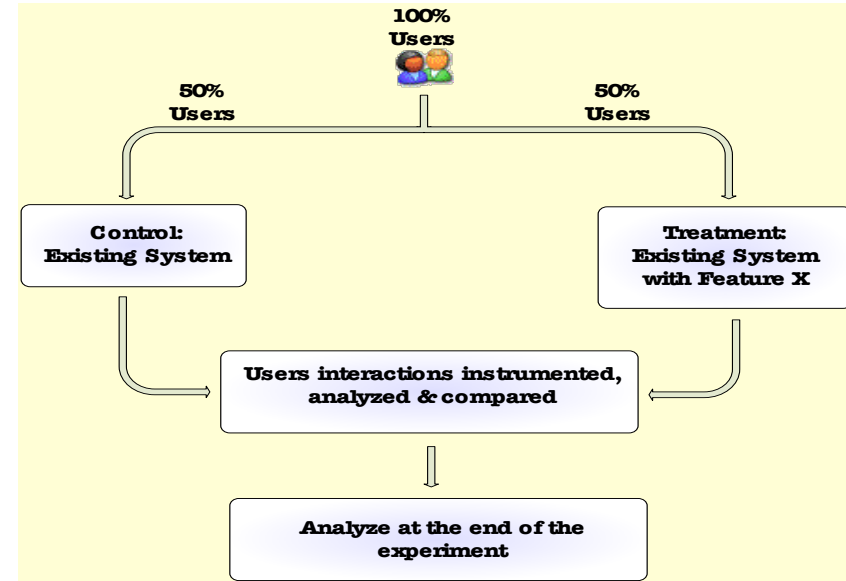


[Feedback](#) / [More info](#)



A/B Testing

- Concept is trivial
 - Randomly split traffic between two (or more) versions
 - A (Control)
 - B (Treatment)
 - Collect metrics of interest
 - Analyze
- Sample of real users
 - Not **WEIRD** (Western, Educated, Industrialized, Rich, and Democratic) like many academic research samples
- A/B test is the simplest controlled experiment
- Must run **statistical tests** to confirm differences are not due to **chance**
- Best scientific way to prove **causality**, i.e., the changes in metrics are caused by changes introduced in the treatment(s)



Experimental Setup

- Evaluate **one factor** with **two levels**
 - A/B test
 - Any percentage; but 50–50 gives maximum **power**
 - Fixed percentage throughout experiment to avoid **Simpson's paradox**

• $E \setminus -$	(a)	Combined	E	$\neg E$	Recovery Rate
		drug (C)	20	20	40
		no-drug ($\neg C$)	16	24	40
			36	44	80
• $E \setminus -$	(b)	Males	E	$\neg E$	Recovery Rate
		drug (C)	18	12	30
		no-drug ($\neg C$)	7	3	10
			25	15	40
• $E \setminus -$	(c)	Females	E	$\neg E$	Recovery Rate
		drug (C)	2	8	10
		no-drug ($\neg C$)	9	21	30
			11	29	40

th **multiple levels**

Understanding Simpson's Paradox

Judea Pearl

Computer Science Department

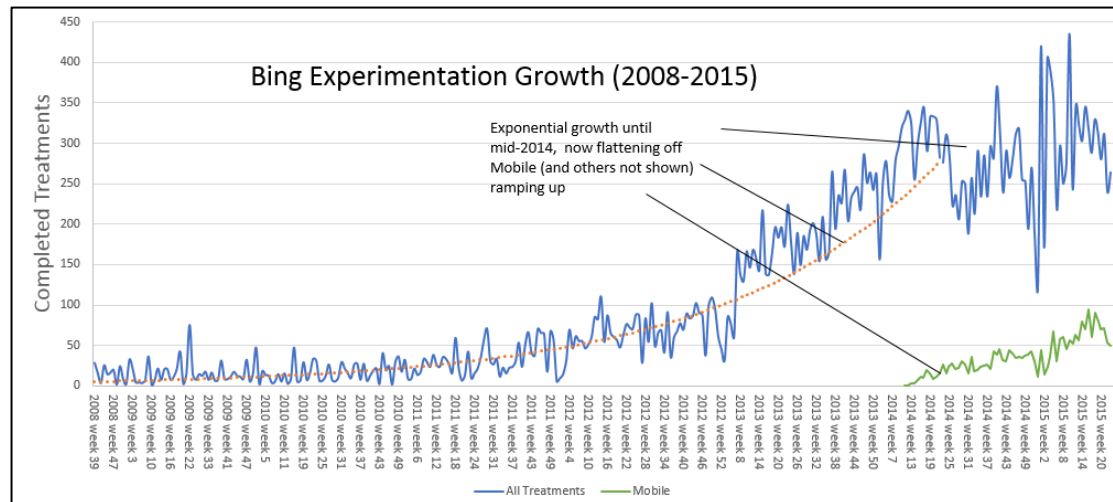
University of California, Los Angeles

Los Angeles, CA, 90095-1596

or

Experimentation at Scale

- At **Bing** they run ~300 experiment treatments every week
- Each variant is exposed to between 100K and millions of users, sometimes tens of millions
- 90% of eligible users are in experiments
 - 10% are a global holdout changed once a year
- There is no single Bing
 - Each user is exposed to 15 concurrent experiments, they get one of $5^{15} = 30$ billion variants



Overlapping Experiments

Designing and Deploying Online Field Experiments

Eytan Bakshy
Facebook
Menlo Park, CA
eytan@fb.com

Dean Eckles
Facebook
Menlo Park, CA
deaneckles@fb.com

Michael S. Bernstein
Stanford University
Palo Alto, CA
msb@cs.stanford.edu

Online Controlled Experiments at Large Scale

Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, Nils Pohlmann
Microsoft, One Microsoft Way, Redmond, WA 98052
{ronnyk, alexdeng, brianfra, towalker, yaxu, nilsp}@microsoft.com

Overlapping Experiment Infrastructure: More, Better, Faster Experimentation

Diane Tang, Ashish Agarwal, Deirdre O'Brien, Mike Meyer
Google, Inc.
Mountain View, CA
[diane,agarwal,deirdre,mimm]@google.com

- Single layer
 - Each randomization unit in a single experiment
 - Easy-to-use, flexible, but insufficiently scalable
- Multi-factorial
 - Full factorial design (independent factors)
 - N factors, k values each $\Rightarrow N^k$ experiments
 - Each randomization unit in N experiments (for N factors)
- Reality: Not all parameters are independent
 - Partition parameters into subsets (layers) of dependent parameters
 - Each randomization unit in M experiments (for M layers)

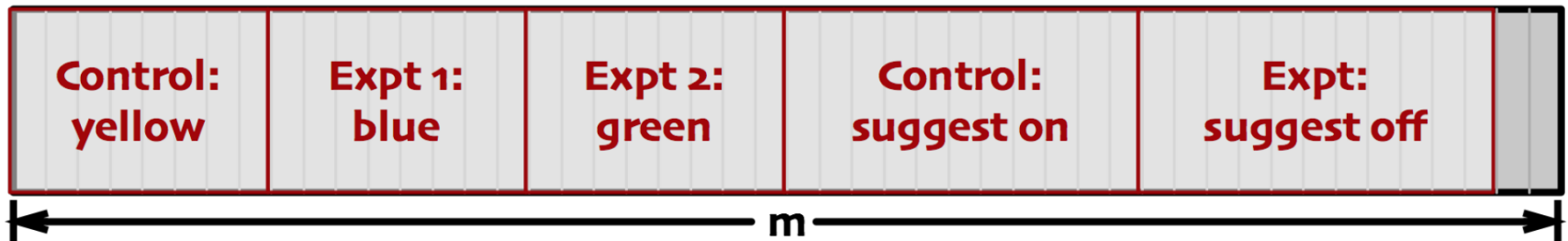
Traffic Diversion

- Random traffic
 - user-visible changes
 - Inconsistent user experience
- Cookie as the basis of diversion
 - Used to track unique users
 - Reality: machine/browser specific and easily cleared
 - Allows consistent user experience over successive queries
- Randomize traffic over cookie mods
 - Easier to specify
 - E.g. cookie mod 1000: Exp1 uses mods 1 and 2, Exp2 uses mods 3 and 4, etc.

Overlapping Experiments

- Extreme 1: **Single Layer**
 - Every request in at most one experiment
 - Straightforward, but insufficiently scalable

Incoming request R
has cookie C
 $f(C) \% 1000 = m$



**Overlapping Experiment Infrastructure:
More, Better, Faster Experimentation**

Diane Tang, Ashish Agarwal, Deirdre O'Brien, Mike Meyer
Google, Inc.
Mountain View, CA
[diane,agarwal,deirdre,mmm]@google.com

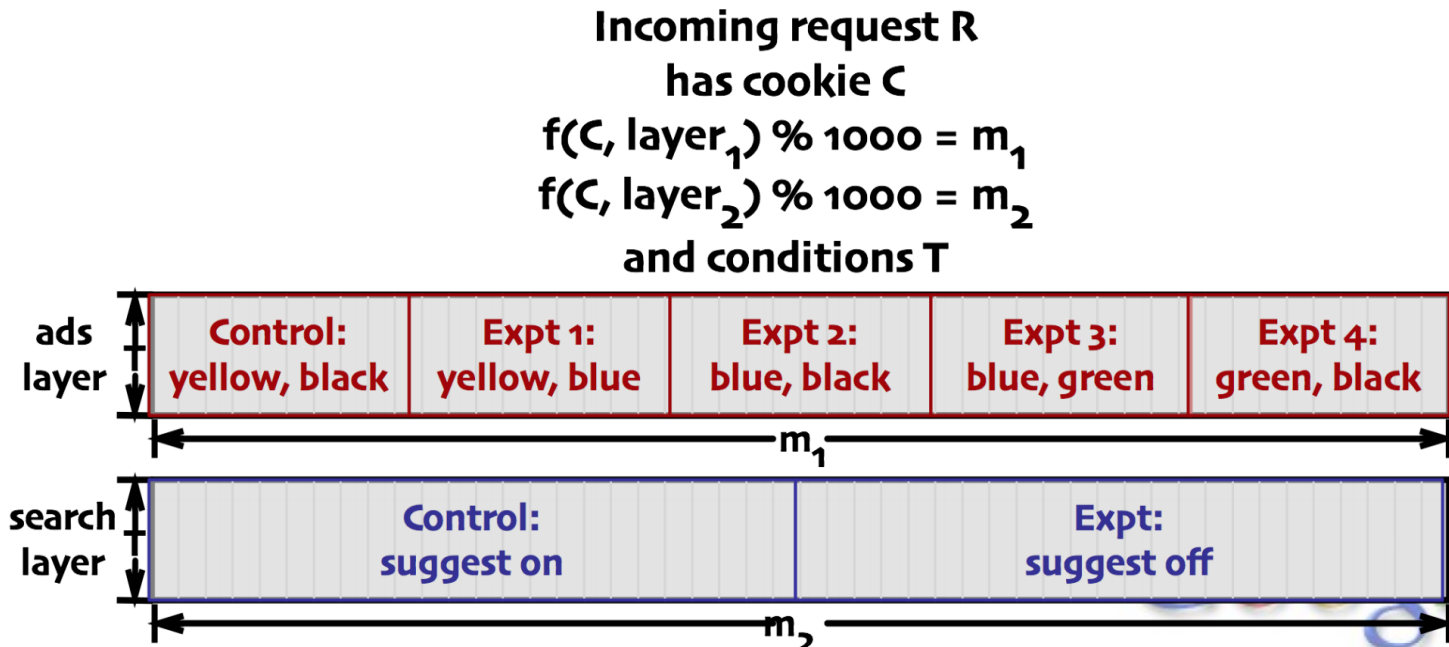
Overlapping Experiments: Extreme 2

- Extreme 2: **Multi-factorial**
 - Vary each parameter independently
 - Issues: Must serve valid pages only e.g., blue text on blue background








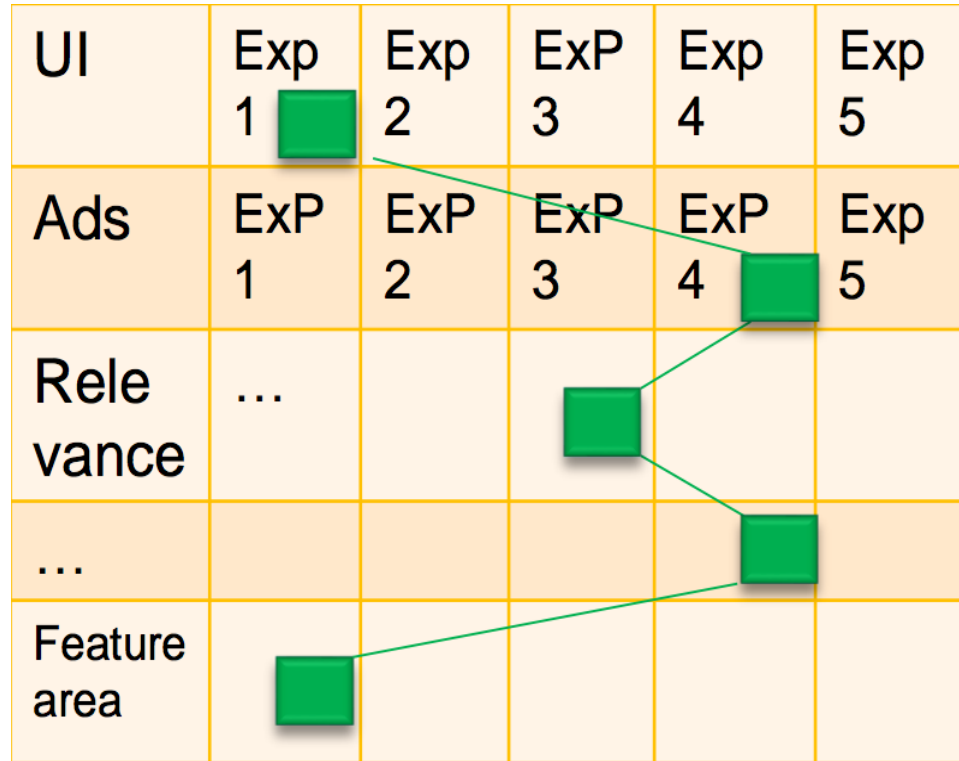
Overlapping experiments

- Partition parameters into **layers**
 - Each layer independent of every other layer
 - Controls and experiments must be in same layer



Overlapping experiments

UI	Exp 1 	Exp 2	Exp 3	Exp 4	Exp 5
Ads	Exp 1	Exp 2	Exp 3	Exp 4 	Exp 5
Relevance	...				
...					
Feature area					

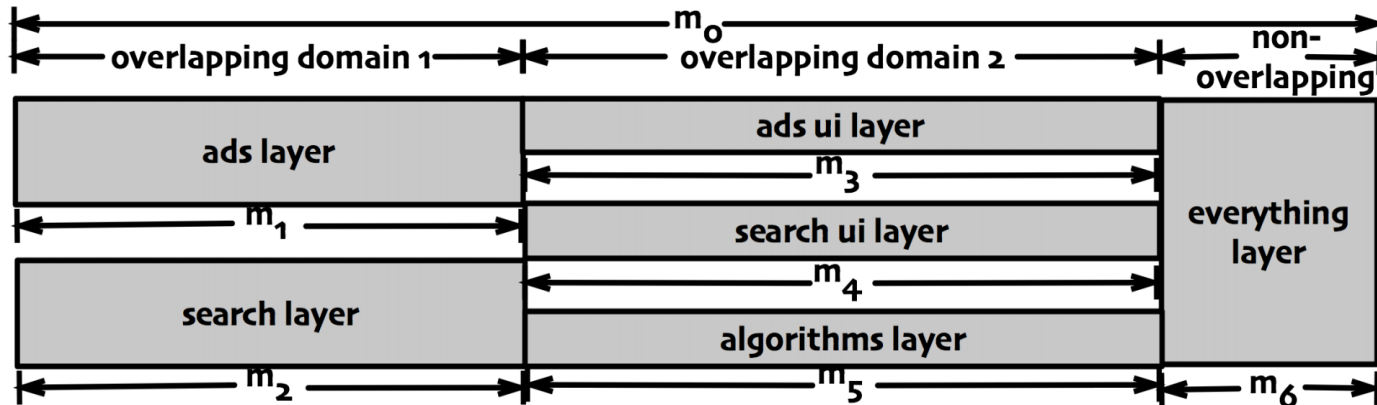


Online Controlled Experiments at Large Scale

Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, Nils Pohlmann
Microsoft, One Microsoft Way, Redmond, WA 98052
{ronnyk, alexdeng, brianfra, towalker, yaxu, nilsp}@microsoft.com

Overlapping experiments

Incoming request R
has cookie C
 $f(C, \text{layer}_i) \% 1000 = m_i$
and conditions T



**Overlapping Experiment Infrastructure:
More, Better, Faster Experimentation**

Diane Tang, Ashish Agarwal, Deirdre O'Brien, Mike Meyer
Google, Inc.
Mountain View, CA
[diane,agarwal,deirdre,mmm]@google.com

A/B Testing

- Running an A/B Test
 - Planning
 - Validation
 - Diagnostics
 - Analysis
- Improving Sensitivity
- Predicting the outcome of an experiment

Planning

- Control **extraneous factors**
 - Test vs. non-test factors
 - Fixing: impact external validity
 - e.g. weekend days are different from week days => run only week days
 - **Randomize**
 - **Blocking**: stratification over non-testing factors => improves **statistical power**

Planning

- Randomization unit
 - Typically: the user
 - Consistent experience
 - Evaluate metrics at user level: sessions or clicks per user
 - In reality: cookie (or login)
 - Affects the power
 - For page-level metrics, more power if randomization at page level

Planning

- Estimate adequate sample size
 - Sample size
 - Percent of users admitted into the experiment variants (control and treatments)
 - Length of the experiment
 - Sample size => statistical power
 - Statistical power
 - Probability of correctly rejecting the null hypothesis when it is false

Statistical Power

- Statistical significance testing:
 1. sample size
 2. effect size = diff of means / st. dev.
 3. significance level = $P(\text{Type I error})$ = probability of finding an effect that is not there
 4. power = $1 - P(\text{Type II error})$ = probability of finding an effect that is there
- Given any three, we can determine the fourth
 - Easier under normality assumption

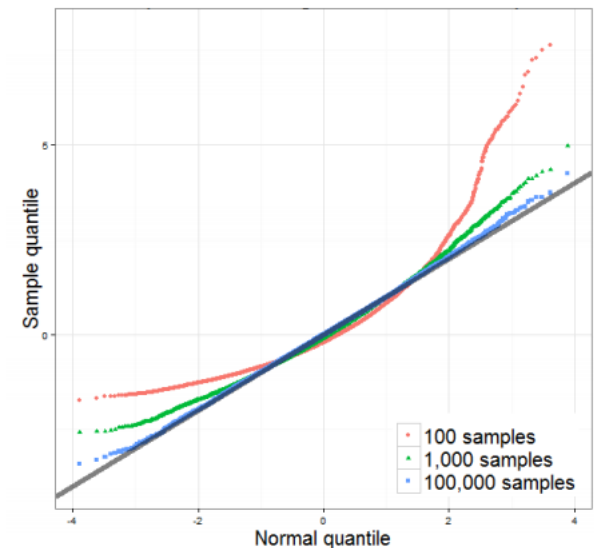
Variance estimation

- Run an A/A test
 - Collect data to assess variability
 - High vs. low variance measures
- Issues with variance estimates
 - Novelty effects
 - Small differences at first
 - Reduces power => longer experiments
 - Skewed measures
 - Not normally distributed

Variance estimation

- Often metrics are skewed
 - Metric transformation
 - Bootstrap estimation

Metric	<i>skewness squared</i>	Min Sample Size	Sensitivity: % change detectable at 80% power
Revenue/User	322.4	114k	4.4%
Revenue/User(Truncated)	27.4	9.7k	10.5%
Sessions/User	13.2	4.70k	5.4%
TimeToSuccess	4.4	1.55k	12.3%
TimeToSuccess (Truncated)	0.15	0.05k	27.9%



QQ-norm plot for averages of different sample sizes showing convergence to Normal when skewness is about 18

Alex Deng
Microsoft
One Microsoft Way
Redmond, WA 98052
alex deng@microsoft.com

Victor Hu
Microsoft
One Microsoft Way
Redmond, WA 98052
vihu@microsoft.com

Planning

- Triggering
 - Track the right users
 - Analyze only the subset of population that was potentially impacted
- Dilution
 - Translates measurements from triggered to overall population
 - Reduces variance

Choose Measures

- On query-SERP
 - Click Through Rate
 - Time to click
 - Reciprocal rank of first click
- On overall activity
 - Number of sessions per user
 - Absence time
- Easy-to-improve measures vs. all-up organizational measures
 - E.g. click to a feature vs. session/user or time-to-success
- Different measures, different variance





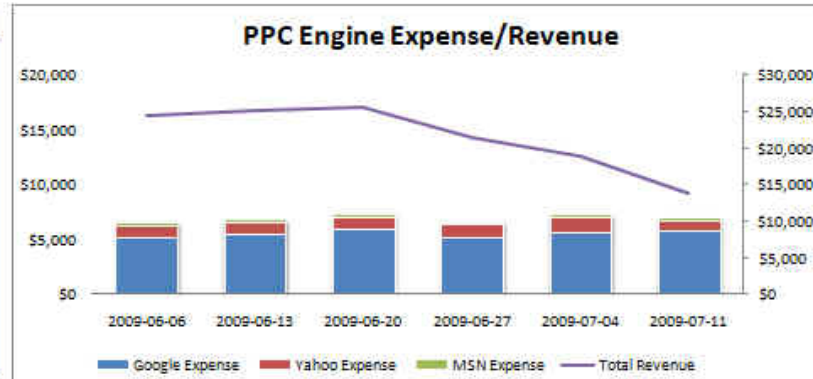
Overall Evaluation Criterion



All Engines Weekly Dashboard

Performance Summary Report for Google, Yahoo, and MSN

All Engines	Prior	This Period	% Change
Impressions	2,300,469	2,088,221	-9.23%
Clicks	18,476	17,245	-6.66%
CTR	0.80%	0.83%	2.82%
CPC	\$0.39	\$0.40	2.44%
Keyword Cost	\$7,258	\$6,939	-4.38%
Revenue	\$18,805	\$13,704	-27.13%
Gross Profit	\$11,547	\$6,764	-41.42%
Orders	361	256	-29.09%
CPO	\$20	\$27	34.83%
AOV	\$52	\$54	2.76%
ROAS	259%	197.48%	-23.78%
Conversion %	1.95%	1.48%	-24.02%



GOOGLE	Prior	This Period	% Change
Impressions	1,227,187	1,382,126	12.63%
Clicks	13,077	13,425	2.66%
CTR	1.07%	0.97%	-9.35%
CPC	\$0.43	\$0.43	-0.09%
Keyword Cost	\$5,684	\$5,831	2.58%
Revenue	\$14,225	\$10,681	-24.91%
Gross Profit	\$8,541	\$4,851	-43.21%
Orders	268	195	-27.24%
CPO	\$21.21	\$29.90	40.98%
AOV	\$53.08	\$54.78	3.20%
ROAS	250%	183%	-26.80%
Conversion %	2.05%	1.45%	-29.27%
% Expense: 84.0%		% Rev: 77.9%	

YAHOO!	Prior	This Period	% Change
Impressions	1,029,098	676,080	-34.30%
Clicks	4,412	3,135	-28.94%
CTR	0.43%	0.46%	6.98%
CPC	\$0.29	\$0.29	-0.31%
Keyword Cost	\$1,288	\$912	-29.17%
Revenue	\$3,334	\$2,442	-26.75%
Gross Profit	\$2,045	\$1,530	-25.22%
Orders	69	47	-31.88%
CPO	\$19	\$19	3.98%
AOV	\$48	\$52	7.54%
ROAS	259%	268%	3.43%
Conversion %	1.56%	1.50%	-3.85%
% Expense: 17.7%		% Rev: 17.7%	

MSN	Prior	This Period	% Change
Impressions	44,184	30,015	-32.07%
Clicks	987	685	-30.60%
CTR	2.23%	2.28%	2.24%
CPC	\$0.29	\$0.29	-0.90%
Keyword Cost	\$286	\$196	-31.21%
Revenue	\$1,245	\$580	-53.43%
Gross Profit	\$960	\$384	-60.04%
Orders	24	14	-41.67%
CPO	\$12	\$14	17.93%
AOV	\$52	\$41	-20.16%
ROAS	436%	295%	-32.30%
Conversion %	2.43%	2.04%	-16.05%
% Expense: 3.9%		% Rev: 6.6%	

Revenue By Engine

Gross Profit / ROAS

Validation

- A/A test (or Null test)
 - Test the experimentation system
 - The null hypothesis should be rejected ~5% of the time if 95% confidence levels are used

Diagnostics

Luo Lu
Twitter Inc.
1355 Market Street, Suite 900
San Francisco, California, USA
llu@twitter.com

Chuang Liu
Twitter Inc.
1355 Market Street, Suite 900
San Francisco, California, USA
chuang@twitter.com

- Carry over effect

- Experiments running in the past may affect users' behavior in the new experiments
- A special case: iterative experimentation with the population in the case buckets dropping off the experiment
 - test for bucket size abnormality
 - if abnormality occurs, **shuffle** users

- Novelty impact

- short term user behavior may not be a good indicator of long term user behavior
- bias can be due to
 - Curiosity
 - learning curve
 - User type structure
- test stability of ratio of control/case metric **throughout time**

Analysis

- Treatment effect and percent change with 95% confidence intervals
 - Law of large numbers: **Normality assumption**
 - **Fieller theorem** for percent change

$$\text{Var} \left(\frac{a}{b} \right) = \left(\frac{a}{b} \right)^2 \left(\frac{\text{Var}(a)}{a^2} + \frac{\text{Var}(b)}{b^2} \right)$$

Analysis

Hypothesis Testing

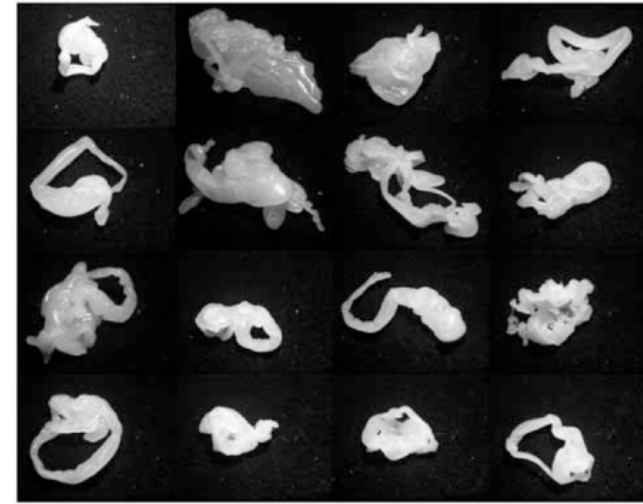
- **Statistical distribution** of the treatment **different** from that of the control
- Simplification: **means** are different
- **Normal distribution** o.w.
 - Transformation of the data
 - Non-parametric tests

Analysis

False positives: 5% expected from Statistics

- Under: one dataset, one outcome, one analysis
- All assumptions are violated
 - Multiple testing
 - Multiple treatments
 - Multiple metrics
 - Slicing and dicing analysis

Sequential testing



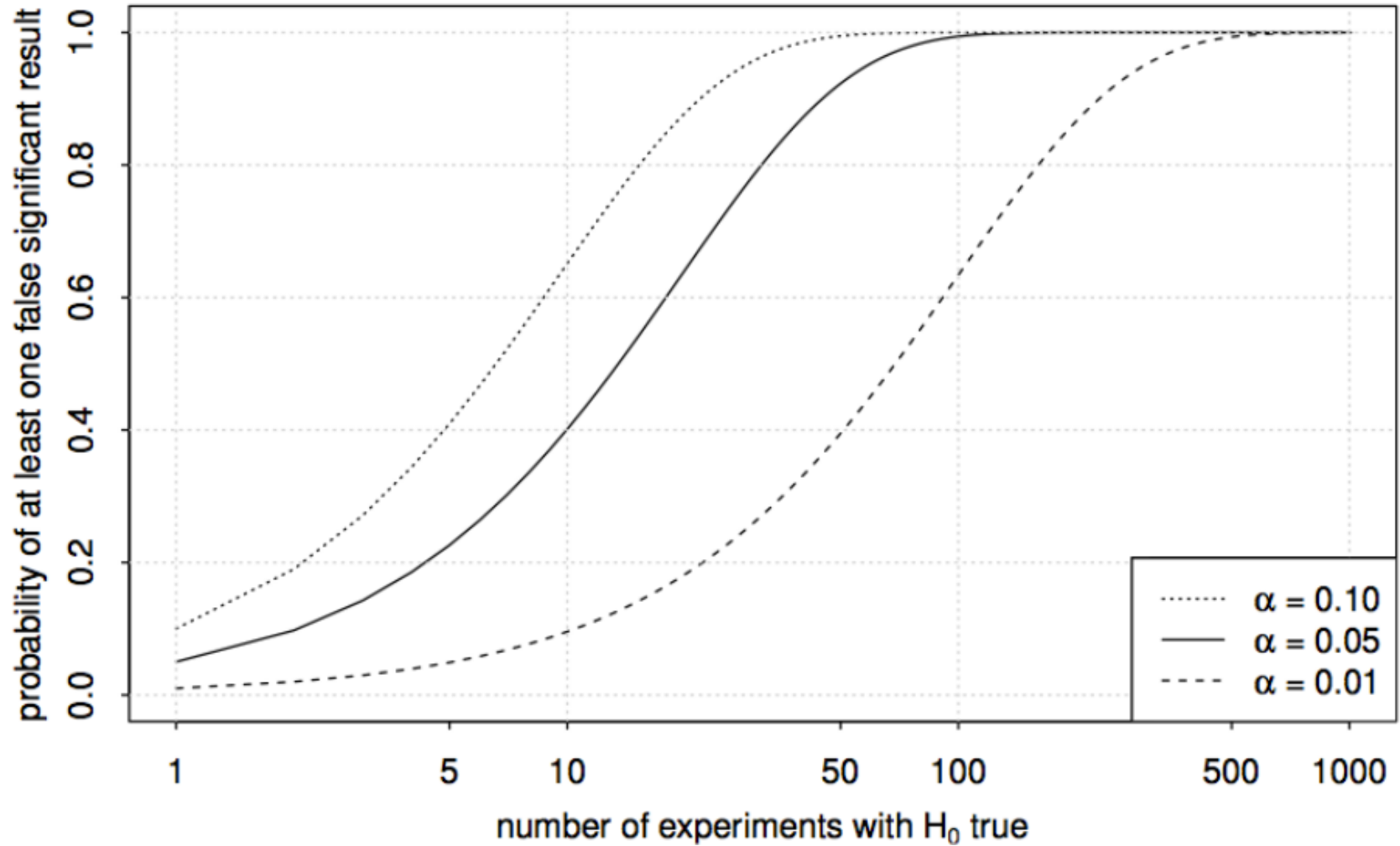
Sequential testing

- Suppose (hypothetically) that the null hypothesis is actually true
- The probability of concluding it is false after one test is α (normally 0.05)
 - The probability of concluding it is false after two tests is $.05 + .95*.05 = .0975$
 - After three tests, $.05 + .95*.05 + .95*.95*.05 = .143$
 - After 14 tests, ~ 0.5
 - After 27 tests, ~ 0.75
 - After 90 tests, ~ 0.99

Multiple testing

- Suppose three different people have the same null hypothesis
 - If each of them does one experiment, probability that there will be one false positive is 0.143
 - If each of them does three experiments, probability goes to ~ 0.4
- Result: very high probability that any given published result is false!
 - “Why Most Published Research Findings Are False”, Ioannidis, PLoS Medicine, 2005

Multiple testing



Correcting for multiple testing

- We should adjust our p-values up for the fact that we have made multiple comparisons
- Many different approaches:
 - Bonferroni correction
 - Tukey's Honest Significant Differences
 - Multivariate t test

A/B Testing

- Running an A/B Test
 - Planning
 - Validation
 - Diagnostics
 - Analysis
- Improving Sensitivity
- Predicting the outcome of an experiment

Statistical Power

- The following four quantities have an intimate relationship:
 1. sample size = # of units * length of exp.
 2. effect size = diff of means / st. dev.
 3. significance level = $P(\text{Type I error})$ = probability of finding an effect that is not there
 4. power = $1 - P(\text{Type II error})$ = probability of finding an effect that is there
- Given any three, we can determine the fourth

Alex Deng*
Microsoft
One Microsoft Way
Redmond, WA 98052
alex deng@microsoft.com

Ya Xu*
Microsoft
1020 Enterprise Way
Sunnyvale, CA 94089
yaxu@microsoft.com

Ron Kohavi
Microsoft
One Microsoft Way
Redmond, WA 98052
ronnyk@microsoft.com

Toby Walker
Microsoft
One Microsoft Way
Redmond, WA 98052
towalker@microsoft.com

- Stratification

1. Divide the sampling region into strata
2. Sample within each stratum separately
3. Combine results from individual strata

- Still obtain an unbiased estimator

- Reduce variance

- Variance
 - Variance within strata
 - Variance between strata

- How can we stratify?

- Use pre-experiment variables to construct strata

Alex Deng*
Microsoft
One Microsoft Way
Redmond, WA 98052
alex deng@microsoft.com

Ya Xu*
Microsoft
1020 Enterprise Way
Sunnyvale, CA 94089
yaxu@microsoft.com

Ron Kohavi
Microsoft
One Microsoft Way
Redmond, WA 98052
ronnyk@microsoft.com

Toby Walker
Microsoft
One Microsoft Way
Redmond, WA 98052
towalker@microsoft.com

- Control variates

1. Choose a random variable Y , with known $E[Y]$
2. Estimate difference in control/exp. X as

$$\hat{X} = \bar{X} - \theta \bar{Y} + \theta E[Y]$$

- Still obtain an unbiased estimator
- Reduce variance
 - By a factor of ρ^2 , with $\rho = \text{cor}(X, Y)$
- How can we find a control variate with known expectation and high correlation?
 - Use pre-experiment variables to construct strata

Variance reduction

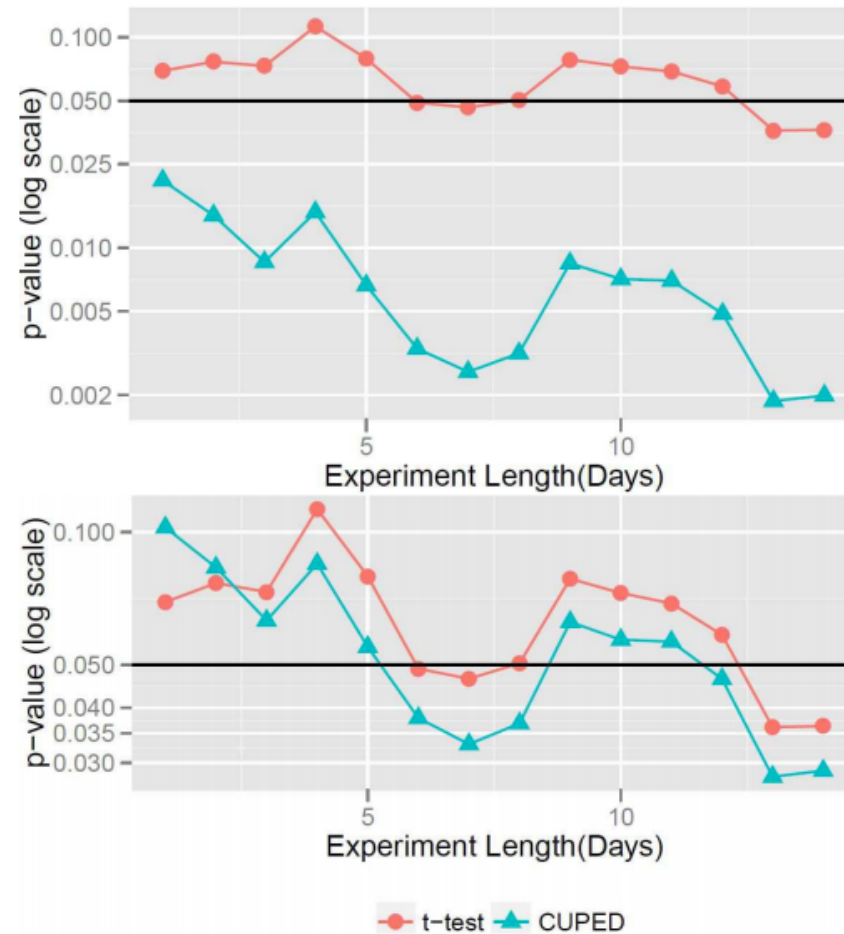


Figure 2: Slowdown experiment. Top: p-value. Bottom: p-value when using only half the users for CUPED.

Increase sample size

- Pseudo-sample size
 1. Consider a number of user engagement measures
 2. Run the experiment and record these measures as time series
 3. Generate a number of features based on time series signals
 - Statistics, totals, derivatives, periodicity, entropy, etc.
 4. Predict the future
 5. Use observed and predicted data for testing

Increase sample size

- Pseudo-sample size

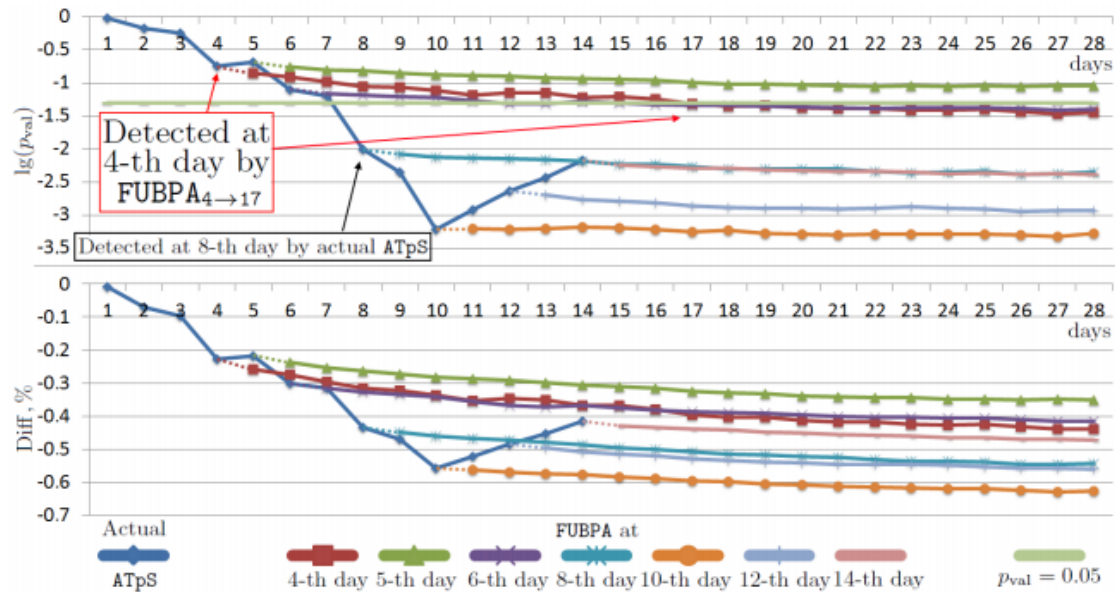


Figure 4: The Diff and p_{val} of ATPs observed during an example A/B test and of the estimations of ATPs by the FUBPA_{X→Y} with different values of X and Y.

- Reduce the duration of an experiment
 - Stop early
- Sequential testing
 - Repeated significance tests
 - Pocock
 - O'Brien & Fleming
 - Sequential Probability Ratio Test (SPRT)
 - Likelihood ratio:
 - likelihood of observed data under H_1 divided by likelihood of observed data under H_0
 - The likelihood under H_1 is unknown
 - replace it with the maximum likelihood estimate before the i -th step.

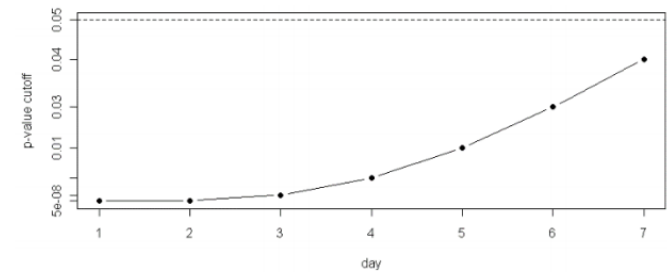


Figure 4: O'Brien-Fleming p-value thresholds as the experiment progresses, with 7 check points

A/B Testing

- Running an A/B Test
 - Planning
 - Validation
 - Diagnostics
 - Analysis
- Improving Sensitivity
- Predicting the outcome of an experiment

Predicting Experimental Results

- Predict the outcome of **online experiments** using **offline log data**
 - We want the best from both worlds

Predicting online metrics

- **Accurate user model** [Artem et al, SIGIR 2015]

Bayesian Ranker Comparison Based on Historical User Interactions

Artem Grotov
a.groto@uva.nl

Shimon Whiteson
s.a.whiteson@uva.nl

Maarten de Rijke
derijke@uva.nl

University of Amsterdam, Amsterdam, The Netherlands

- **Randomization** [Li et al, WSDM 2015]

Toward Predicting the Outcome of an A/B Experiment for Search Relevance

Lihong Li *
Microsoft Corp
One Microsoft Way
Redmond, WA 98052
lihongli@microsoft.com

Jin Young Kim †
Microsoft Corp
One Microsoft Way
Redmond, WA 98052
jink@microsoft.com

Imed Zitouni
Microsoft Corp
One Microsoft Way
Redmond, WA 98052
izitouni@microsoft.com

Running an experiment

- A search engine

1. observes a query q from Q (iid)

$$q \sim \pi$$

2. takes an action a from $A_q = \{\text{SERP}_q\}$ (iid)

$$a \sim \pi(\cdot|q)$$

3. receives a reward r in $[0, R]$

- any evaluation measure

Expected reward

- If we run the experiment

$$\{ \langle q_i, a_i, r_i \rangle \}_{1 \leq i \leq m}$$

$$\mathbf{E}[r] = \frac{1}{m} \sum_i r_i$$

$$\begin{aligned} \mathbf{E}[r] &= \mathbf{E}_{q \sim \pi} [\mathbf{E}[r | q]] \\ &= \mathbf{E}_{q \sim \pi} \mathbf{E}_{a \sim \pi(\cdot | q)} [\mathbf{E}[r | a, q]] \end{aligned}$$

Expected reward

- Instead, consider the query log

$$\{ \langle q_i, a_i, r_i \rangle \}_{1 \leq i \leq n}$$

- Assumption 1: $n(q, a)$ is not 0 in the log

$$\mathbf{E}[r|a, q] = \frac{1}{n(q, a)} \sum_{1 \leq i \leq n} \mathbf{I}(q_i = q, a_i = a) r_i = \hat{r}(q, a)$$

Expected reward

- Instead, consider the query log

$$\{ \langle q_i, a_i, r_i \rangle \}_{1 \leq i \leq n}$$

- Assumption 2: the distribution of rewards is stationary

$$\begin{aligned} & \mathbf{E}_{q \sim \pi} [\mathbf{E}_{a \sim \pi(\cdot|q)} [\hat{r}(q, a)]] \\ &= \mathbf{E}_{q \sim \pi} \left[\sum_{a \in A_q} \pi(a|q) \hat{r}(q, a) \right] \\ &= \sum_{q \in Q} \pi(q) \sum_{a \in A_q} \pi(a|q) \hat{r}(q, a) \end{aligned}$$

Expected reward

- Instead, consider the query log

$$\{ \langle q_i, a_i, r_i \rangle \}_{1 \leq i \leq n}$$

- Reality: actions are deterministic

$$= \sum_{q \in Q} \pi(q) \sum_{a \in A_q} \pi(a|q) \hat{r}(q, a)$$

$$= \sum_{q \in Q} \pi(q) \hat{r}(q, a)$$

Expected reward

- Instead, consider the query log

$$\{ \langle q_i, a_i, r_i \rangle \}_{1 \leq i \leq n}$$

- If distribution of queries remains the same then,

$$\pi(q) = \frac{n(q)}{n}$$

- Otherwise, use live statistics

Variance of reward

- If a variable is bounded, $r \in [m, M]$

$$\mathbf{Var}[r] \leq \frac{(M - m)^2}{4}$$

- Based on this, we can compute:

$$\mathbf{Var}[\hat{r}] \leq \frac{R^2}{4} \sum_{q \in Q} \frac{n^2(q)}{n^2} \frac{1}{n(q, a_q)}$$

Fuzzy matching

- Limitations

1. Variance grows linearly to the cardinality of A_q
2. Very likely that $n(q, a) = 0$

- Fuzzy matching

$$a \sim a' \text{ if } a_{[1..j]} = a'_{[1..j]}$$

Results: Predicting absolute values

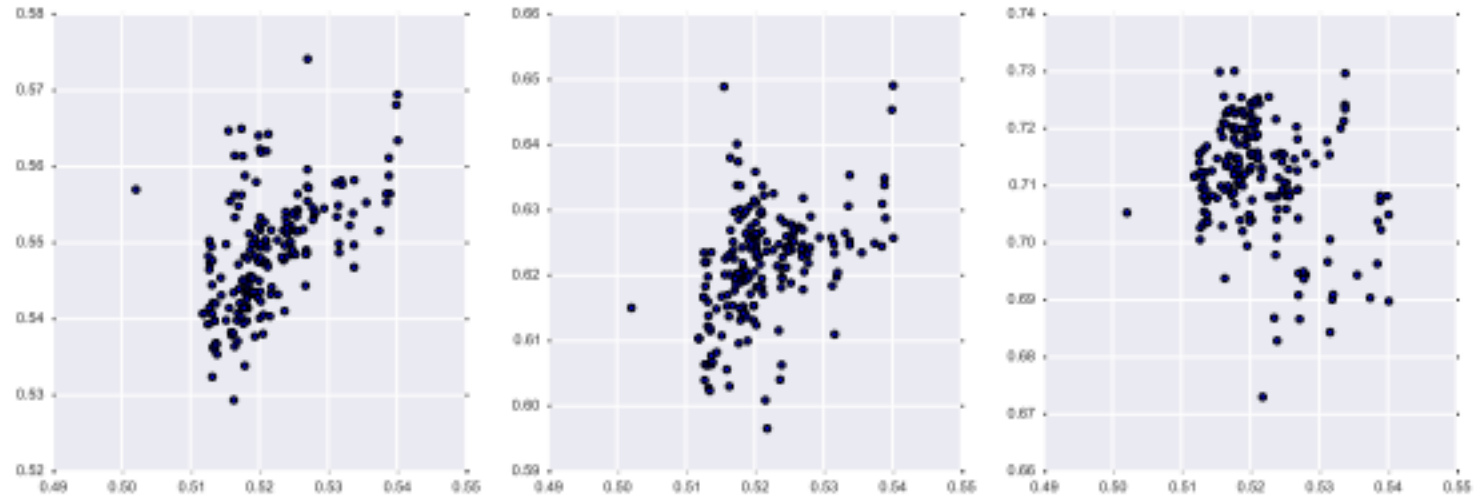


Figure 1: Scatterplot for actual (X) vs. predicted (Y) click ratio based on the 1st estimator (\hat{v}_1), where the action is defined by Top 3, 5 and 8 web results, respectively.

Table 1: Results for predicting absolute metric values.

TopK	$\text{Cor}(\hat{v}_1)$	$\text{Cor}(\hat{v}_2)$	$\#(\text{Query}, \text{Action})$
3	0.549	0.596	645,749,791
5	0.431	0.438	244,046,777
8	-0.271	-0.254	63,646,334

Results: Predicting Deltas

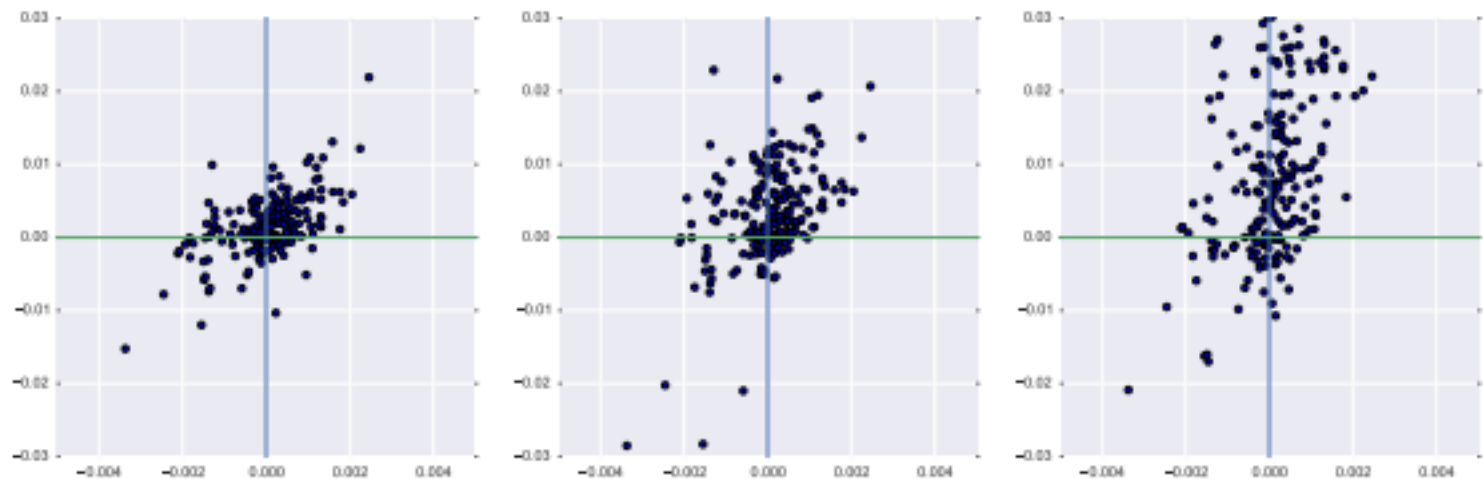


Figure 3: Scatterplot for actual (X) vs. predicted (Y) delta in click ratio between two rankers in the same experiment where the action is defined by Top 3, 5 and 8 web results, respectively.

Results: Predicting Decisions

Table 3: WIN/TIE/LOSS confusion table between actual outcomes (columns) and predicted outcomes (rows) for different K values of topK fuzzy matching. Accuracy/recall are also included for each outcome.

K=3	LOSS	TIE	WIN	Accuracy	Recall
WIN	3.4%	19.0%	13.7%	37.8%	65.1%
TIE	8.8%	38.0%	6.8%	70.9%	63.4%
LOSS	6.8%	2.9%	0.5%	66.7%	35.9%
K=5	LOSS	TIE	WIN	Accuracy	Recall
WIN	3.9%	11.7%	5.9%	27.3%	27.9%
TIE	12.2%	47.3%	15.1%	63.4%	78.9%
LOSS	2.9%	1.0%	0.0%	75.0%	15.4%
K=8	LOSS	TIE	WIN	Accuracy	Recall
WIN	4.4%	18.5%	7.3%	24.2%	34.9%
TIE	12.7%	41.5%	13.7%	61.2%	69.1%
LOSS	2.0%	0.0%	0.0%	100%	10.3%

Table 2: Results for predicting the delta in metric values between two rankers.

TopK	Accuracy	Correlation
3	58.5%	0.450
5	56.1%	0.396
8	50.7%	0.370

Analysis: Optimistic Bias vs. Coverage

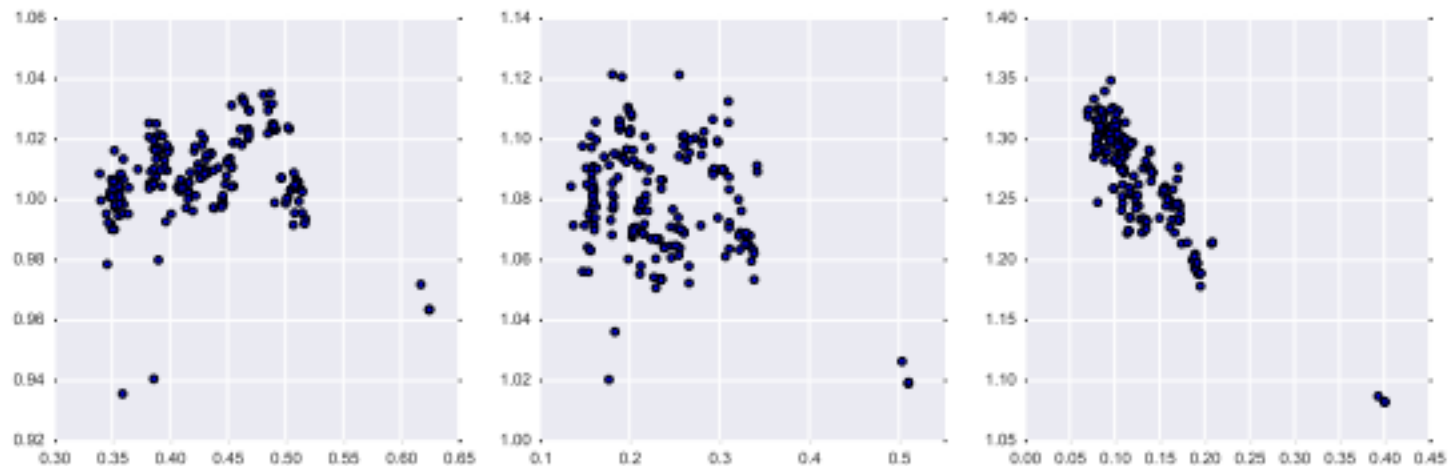


Figure 4: Scatterplot for the ratio of predicted metric value against actual metric value, plotted against % of matching records, where the action is defined by Top 3, 5 and 8 web results, respectively.

Analysis: Assumptions

1. No confounding effects
2. Sufficient amount of data for each query
3. Enough observations for (query, SERP) pair
4. Consistent users behavior (i.e. stable rewards distribution)

Analysis: Query Segments

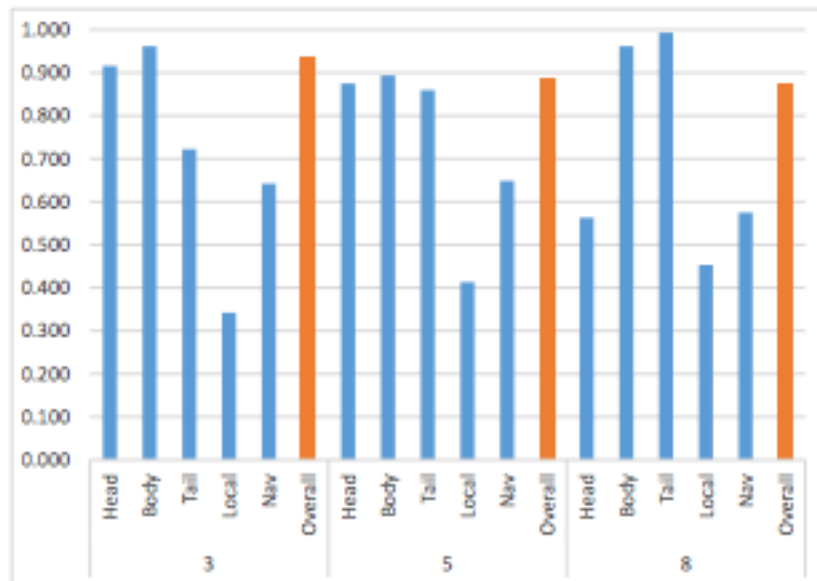


Figure 5: Correlation between predicted and actual metric values for different segments with varying K values in fuzzy matching.

Table 5: Correlation between predicted and actual metric values for different segments with different K values in fuzzy matching.

TopK	Segment	Cor	Count
3	Head	0.916	364,757
	Body	0.962	553,892,081
	Tail	0.722	90,867,404
	Local	0.342	60,448,865
	Navigational	0.642	120,343,701
	Overall	0.934	645,124,406
5	Head	0.875	345,640
	Body	0.894	229,862,589
	Tail	0.860	13,636,733
	Local	0.413	24,049,649
	Navigational	0.648	58,680,950
	Overall	0.885	243,845,041
8	Head	0.563	304,142
	Body	0.962	61,433,473
	Tail	0.993	1,809,956
	Local	0.453	6,369,383
	Navigational	0.575	23,107,195
	Overall	0.872	63,547,643

Prediction for Learning

Counterfactual Estimation and Optimization of Click Metrics in Search Engines: A Case Study

Lihong Li¹

Shunbao Chen¹

Jim Kleban^{2*}

Ankur Gupta¹

¹Microsoft Corp.
Redmond, WA 98052

²Facebook Inc.
Seattle, WA 98101

Counterfactual Risk Minimization: Learning from Logged Bandit Feedback

Adith Swaminathan

Cornell University, Ithaca, NY 14853 USA

ADITH@CS.CORNELL.EDU

Thorsten Joachims

Cornell University, Ithaca, NY 14853 USA

TJ@CS.CORNELL.EDU

A/B Test Types

- Experiment
 - To validate a new idea (algorithm, feature, interface, etc.)
- Calibration test
 - Degrade production system deliberately with a known quantity (i.e., remove top document), to calibrate metrics
- A/A test
 - No differences should be measured (95% of the time)
- Reverse test
 - Test a previous experiment again by reversely applying changes
- Random bucket
 - To collect data

Summary of A/B testing

- When the variants run **concurrently**, only two things could explain a change in metrics:
 - Actual difference in the **quality** of the algorithms
 - **Random chance**
- Everything else happening affects both the variants
- For random chance, conduct **statistical tests** for significance

Challenges in A/B Testing

- One metric to rule them all
 - Overall evaluation criterion (OEC)
 - Many metrics; typically improve one but hurt the others
 - Higher level metric to incorporate tradeoffs among metrics
 - Measurable over short duration (e.g. two weeks)
 - Predictive of long term-goals

Challenges in A/B Testing

- OEC: Market share, aka number of queries
 - Making the search engine worse will lead to more queries short term
 - But push users to alternatives in long-term
- Better: sessions per user; repeated visits

Challenges in A/B Testing

- Long turn-around time
 - => improve sensitivity
 - reduce variance by stratification
 - pseudo-increase data points by predicting future user behaviour
 - => use interleaving
- Non-guaranteed quality of the experimental system
 - => off-line evaluation
 - collection-based evaluation
 - side-by-side experiments
- Many experiments competing for traffic
 - => prioritize

6. Interleaving

Interleaving

Ranker A

bing Interleaving for information retrieval

1 [Information Retrieval - University of Glasgow :: School of ...](#)
www.gla.ac.uk/...research/researchgroups/informationretrieval
The group has a long and strong research history in the process of information retrieval as a ... information in. retrieval ... Interleaving for Retrieval ...

2 [Optimized Interleaving for Online Retrieval Evaluation](#)
research.microsoft.com/pubs/179433/Radlinski_Optimized_WSDM2013... - PDF
Optimized Interleaving for Online Retrieval Evaluation Filip Radlinski Microsoft Cambridge, UK filiprad@microsoft.com Nick Craswell Microsoft Bellevue, WA, U...

3 [Optimized Interleaving for Online Retrieval Evaluation ...](#)
research.microsoft.com/apps/pubs/default.aspx?id=179433
Abstract. Interleaving is an online evaluation technique for comparing the relative of information retrieval functions by combining their result lists and ...

4 [CiteSeerX — Content-Aware DataGuides: Interleaving IR ...](#)
citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.4.4377
Content-Aware DataGuides: Interleaving IR and DB Indexing Techniques for Efficient Retrieval of Textual XML Data (2004)

5 [Large-scale validation and analysis of interleaved search ...](#)
dl.acm.org/citation.cfm?id=2094078
Interleaving is an increasingly popular technique for evaluating information retrieval systems based on implicit user feedback. While a number of isolated studies ...

Which ranker is better?

Several ways to find out:

- Ask assessors which documents are relevant.
- Split user population, observe user interactions (clicks) with ranker A and B.
- Interleave ranker A and ranker B

Ranker B

bing Interleaving for information retrieval

1 [research on online learning to rank for information retrieval](#)
khofm.wordpress.com
research on online learning to rank for information retrieval
contributions of this thesis

Expensive
Labels don't come from users
...

Between subject design
A and B seen by different users
A and B seen by different queries

Within subject design
A and B seen by same users with
same queries

Interleaving

Ranker A

wins

bing

Interleaving for information retrieval

Information Retrieval - University of Glasgow :: School of ...

www.gla.ac.uk/.../research/researchgroups/informationretrieval

The group has a long and strong research history in the process of information r

as a ... information in. retrieval ... Interleaving for Retrieval ...

1

Optimized Interleaving for Online Retrieval Evaluation

research.microsoft.com/pubs/179433/Radlinski_Optimized_WSDM2013...

PDF

Optimized Interleaving for Online Retrieval Evaluation Filip Radlinski Microsoft

Cambridge, UK filiprad@microsoft.com Nick Craswell Microsoft Bellevue, WA, US

2

Optimized Interleaving for Online Retrieval Evaluation ...

research.microsoft.com/apps/pubs/default.aspx?id=179433

Abstract. Interleaving is an online evaluation technique for comparing the relativ

of information retrieval functions by combining their result lists and ...

3

CiteSeerX — Content-Aware DataGuides: Interleaving IR ...

citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.4.4377

Content-Aware DataGuides: Interleaving IR and DB Indexing Techniques for Effi

Retrieval of Textual XML Data (2004)

4

Large-scale validation and analysis of interleaved search ...

dl.acm.org/citation.cfm?id=2094078

Interleaving is an increasingly popular technique for evaluating information retr

systems based on implicit user feedback. While a number of isolated studies ...

5

bing Interleaving for information retrieval



Ranker B

loses

bing

Interleaving for information retrieval

research on online learning to rank for information retrieval

khofm.wordpress.com

research on online learning to rank for information retrieval (by Katja Hofmann)

contributions of this thesis include a novel interleaved comparison method, ...

6

Optimized Interleaving for Online Retrieval Evaluation

research.microsoft.com/pubs/179433/Radlinski_Optimized_WSDM2013...

PDF

Optimized Interleaving for Online Retrieval Evaluation Filip Radlinski Microsoft

Cambridge, UK filiprad@microsoft.com Nick Craswell Microsoft Bellevue, WA, US

2

Information Retrieval - University of Glasgow :: School of ...

www.gla.ac.uk/.../research/researchgroups/informationretrieval

The group has a long and strong research history in the process of information r

as a ... information in. retrieval ... Interleaving for Retrieval ...

1

CiteSeerX — Content-Aware DataGuides: Interleaving IR ...

citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.4.4377

Content-Aware DataGuides: Interleaving IR and DB Indexing Techniques for Effi

Retrieval of Textual XML Data (2004)

4

Large-scale validation and analysis of interleaved search ...

dl.acm.org/citation.cfm?id=2094078

Interleaving is an increasingly popular technique for evaluating information retr

systems based on implicit user feedback. While a number of isolated studies ...

5

Why do interleaving?

- **Within subject** design
... as opposed to **between subject** of A/B testing
 - **Reduces variance** (same users/queries for both A and B)
 - Need 1 to 2 orders of magnitude less data
 - ~100K queries for interleaving in a mature web search engine (>>1M for A/B testing)

Interleaving

- Running an Interleaving Test
 - Method
 - Analysis
- Improving Sensitivity
- Predicting the outcome of an experiment

Interleaving Methods

- Balanced interleave (*Joachims et al 2006, Radlinski et al 2008*)
- Team Draft interleave (*Radlinski et al 2008*)
- Document constraints interleave (*He et al 2009*)
- Probabilistic interleave (*Hofmann et al 2011*)
- Optimized interleave (*Radlinski and Craswell 2013*)
- Vertical aware team draft interleave (*Chuklin et al 2013*)
- Team draft multileave (*Schuth et al 2014*)
- Optimized multileave (*Schuth et al 2014*)
- Probabilistic multileave (*Schuth et al 2015*)

Team Draft Interleaving

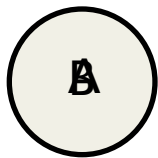
Ranking A

1. ~~Napa Valley - The authority for lodging...~~
~~www.napavalley.com~~
2. ~~Napa Valley Wineries - Plan your wine...~~
~~www.napavalley.com/wineries~~
3. Napa Valley College
www.napavalley.edu/homex.asp
4. Been There | Tips | Napa Valley
www.ivebeenthere.co.u
5. Napa Valley Wineries ar
www.napavintners.com
6. ~~Napa Country, California~~
~~en.wikipedia.org/wiki/~~

Ranking B

1. ~~Napa Country, California - Wikipedia~~
 2. ~~en.wikipedia.org/wiki/Napa_Valley~~
~~Napa Valley - The authority for lodging...~~
~~www.napavalley.com~~
 3. ~~Napa: The Story of an American Eden...~~
~~books.google.co.uk/books?isbn=...~~
 4. ~~Napa valley Hotels - Bed and Breakfast...~~
- s.com
g
ey.org
ey Marathon
eymarathon.org

Presented Ranking



Team Draft Interleaving

Ranking A

1. Napa Valley – The authority for lodging...
www.napavalley.com
2. Napa Valley Wineries – Plan your wine...
www.napavalley.com/wineries
3. Napa Valley College
www.napavalley.edu/homex.asp
4. Been There | Tips | Napa Valley
www.ivebeenthere.co.u
5. Napa Valley Wineries a
www.napavintners.com
6. Napa Country, Californ
en.wikipedia.org/wiki/

Ranking B

1. Napa Country, California – Wikipedia
en.wikipedia.org/wiki/Napa_Valley
2. Napa Valley – The authority for lodging...
www.napavalley.com
3. Napa: The Story of an American Eden...
books.google.co.uk/books?isbn=...
4. Napa Valley Hotels – Bed and

Presented Ranking

1. Napa Valley – The authority for lodging...
www.napavalley.com
2. Napa Country, California – Wikipedia
en.wikipedia.org/wiki/Napa_Valley
3. Napa: The Story of an American Eden...
books.google.co.uk/books?isbn=...
4. Napa Valley Wineries – Plan your wine...
www.napavalley.com/wineries
5. Napa valley Hotels – Bed and Breakfast...
www.napalinks.com
6. Napa Balley College
www.napavalley.edu/homex.asp
7. NapaValley.org
www.napavalley.org

Click

Click

Tie!

Probabilistic Interleaving

- An alternative interleaving algorithm
- Similar to Team Draft, but
 - Toss a coin every pick
 - Don't always pick the top result
 - Assign credit based on all possible assignments

A Probabilistic Method for Inferring Preferences from Clicks

Katja Hofmann, Shimon Whiteson and Maarten de Rijke
ISLA, University of Amsterdam
{K.Hofmann, S.A.Whiteson, deRijke}@uva.nl

Probabilistic Interleaving (1)

Ranking A

1. ~~Napa Valley – The authority for lodging...~~
~~www.napavalley.com~~
2. ~~Napa Valley Wineries – Plan your wine...~~
~~www.napavalley.com/wineries~~
3. Napa Valley College
www.napavalley.edu/homex.asp
4. Been There | Tips | Napa Valley
www.ivebeenthere.co.u
5. Napa Valley Wineries ar
www.napavintners.com
6. ~~Napa Country, California – Wikipedia~~
~~en.wikipedia.org/wiki/~~

Ranking B

1. ~~Napa Country, California – Wikipedia~~
~~en.wikipedia.org/wiki/Napa_Valley~~
 2. ~~Napa Valley – The authority for lodging...~~
~~www.napavalley.com~~
 3. ~~Napa: The Story of an American Eden...~~
~~books.google.co.uk/books?isbn=...~~
 4. Napa Valley Hotels – Bed and
- s.com
g
ey.org
ey Marathon
eymarathon.org

Presented Ranking

A

Probabilistic Interleaving (2)

Ranking A

1. ~~Napa Valley - The authority for lodging...~~
~~www.napavalley.com~~
2. Napa Valley Wineries - Plan your wine...
www.napavalley.com/wineries
3. Napa Valley College
www.napavalley.edu/homex.asp
4. ~~Been There | Tips | Napa Valley~~
~~www.iivebeenthere.co.uk~~
5. Napa Valley Wineries &...
www.napavintners.com
6. ~~Napa Country, California - Wikipedia~~
~~en.wikipedia.org/wiki/~~

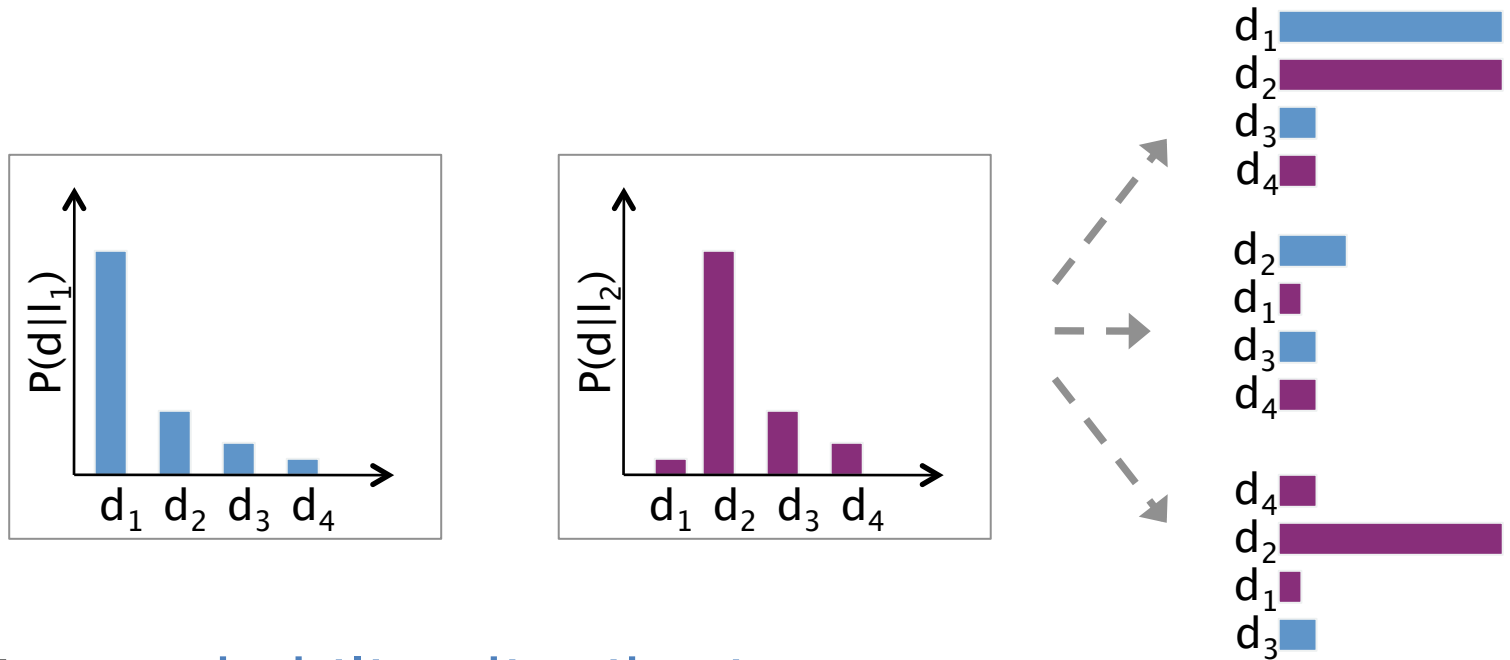
Ranking B

1. ~~Napa Country, California - Wikipedia~~
~~en.wikipedia.org/wiki/Napa_Valley~~
2. ~~Napa Valley - The authority for lodging...~~
~~www.napavalley.com~~
3. ~~Napa: The Story of an American Eden...~~
~~books.google.co.uk/books?isbn=...~~
4. Napa Valley Hotels - Bed and

Presented Ranking

1. Napa Valley - The authority for lodging...
www.napavalley.com
2. Been There | Tips | Napa Valley
www.iivebeenthere.co.uk
3. Napa: The Story of an American Eden...
books.google.co.uk/books?isbn=...
4. Napa Country, California - Wikipedia
en.wikipedia.org/wiki/Napa_Valley
5. Napa Valley Hotels - Bed and Breakfast...
www.napalinks.com
6. Napa Valley College
www.napavalley.edu/homex.asp
7. NapaValley.org
www.napavalley.org

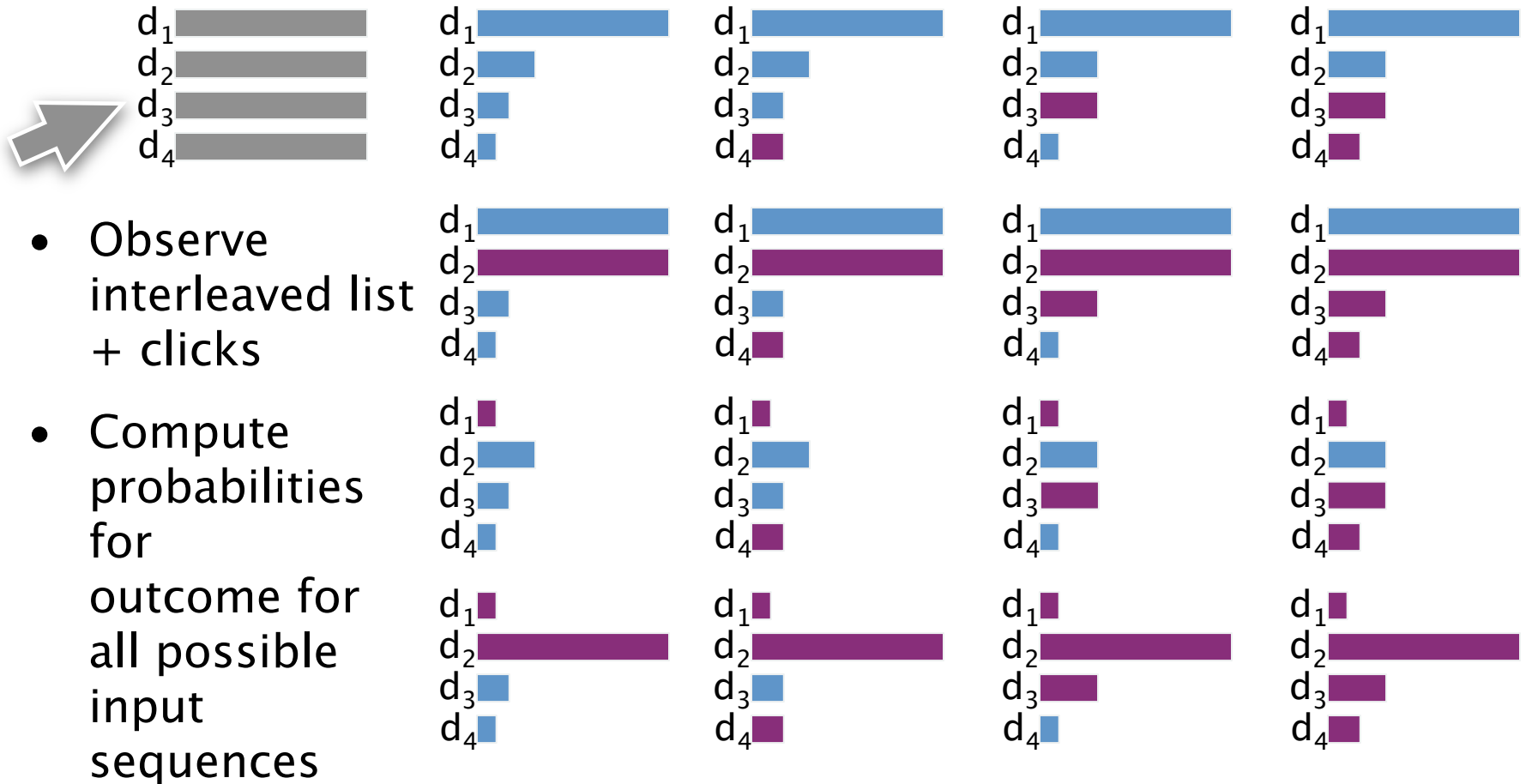
Probabilistic Interleaving (2)



- Define **probability distributions** over documents, based on the lists to be compared
- During interleaving draw documents randomly

➡ Any permutation of documents is possible

Probabilistic Interleaving (3)



Interleaving

- Running an Interleaving Test
 - Method
 - Analysis
- Improving Sensitivity
- Predicting the outcome of an experiment

Predicting Experimental Results

- Probabilistic interleaving
 - Applied on historical data
 - For two rankers, some permutation of the interleaved list might be in the logs
- Importance sampling
 - Correct for bias

Estimating Interleaved Comparison Outcomes from Historical Click Data

Katja Hofmann
k.hofmann@uva.nl

Shimon Whiteson
s.a.whiteson@uva.nl

Maarten de Rijke
derijke@uva.nl

Interleaving

- Running an Interleaving Test
 - Method
 - Analysis
- Improving Sensitivity
- Predicting the outcome of an experiment

Improving Sensitivity

- Optimized interleaving
 - Set constraints and desirable properties
 - Among them, high sensitivity
 - Obtain an interleaving algorithm as a solution to an optimization problem

Optimized Interleaving for Online Retrieval Evaluation

Filip Radlinski
Microsoft
Cambridge, UK
filiprad@microsoft.com

Nick Craswell
Microsoft
Bellevue, WA, USA
nickcr@microsoft.com

Beyond Click Count

Learning More Powerful Test Statistics for Click-Based Retrieval Evaluation

Yisong Yue
Cornell University
Ithaca, NY, USA
yyue@cs.cornell.edu

Yue Gao
Cornell University
Ithaca, NY, USA
ygao@cs.cornell.edu

Olivier Chapelle
Yahoo! Research
Santa Clara, CA, USA
chap@yahoo-inc.com

Ya Zhang
Shanghai Jiao Tong University
Shanghai, China
ya_zhang@sjtu.edu.cn

Thorsten Joachims
Cornell University
Ithaca, NY, USA
tj@cs.cornell.edu

- Not every click in the interleaved ranking is equally informative
- Instead of $\delta(q, C, C') = |C| - |C'|$ use

$$\delta(q, C, C') = \sum_{c \in C} \text{score}(q, c) - \sum_{c' \in C'} \text{score}(q, c')$$

- Score = linear combination of features
 - learned from training pairs of known retrieval quality

7. Comparative Studies

Quantitative Analysis

- Can we quantify **how well** Interleaving performs?
 1. Compared to **offline judgments**
 2. Compared to **absolute ranking**-level Metrics
- How **reliable** is it?
 - Does Interleaving correctly identify the better retrieval function?
- How **sensitive** is it?
 - How much data is required to achieve a target confidence level (p-value)?

Quantitative Analysis

How Does Clickthrough Data Reflect Retrieval Quality?

Filip Radlinski
Dept. of Computer Science
Cornell University
Ithaca, NY, USA
filip@cs.cornell.edu

Madhu Kurup
Dept. of Computer Science
Cornell University
Ithaca, NY, USA
mmk222@cs.cornell.edu

Thorsten Joachims
Dept. of Computer Science
Cornell University
Ithaca, NY, USA
tj@cs.cornell.edu

Comparing the Sensitivity of Information Retrieval Metrics

Filip Radlinski
Microsoft
Cambridge, UK
filiprad@microsoft.com

Nick Craswell
Microsoft
Redmond, WA, USA
nickcr@microsoft.com

Large-Scale Validation and Analysis of Interleaved Search Evaluation

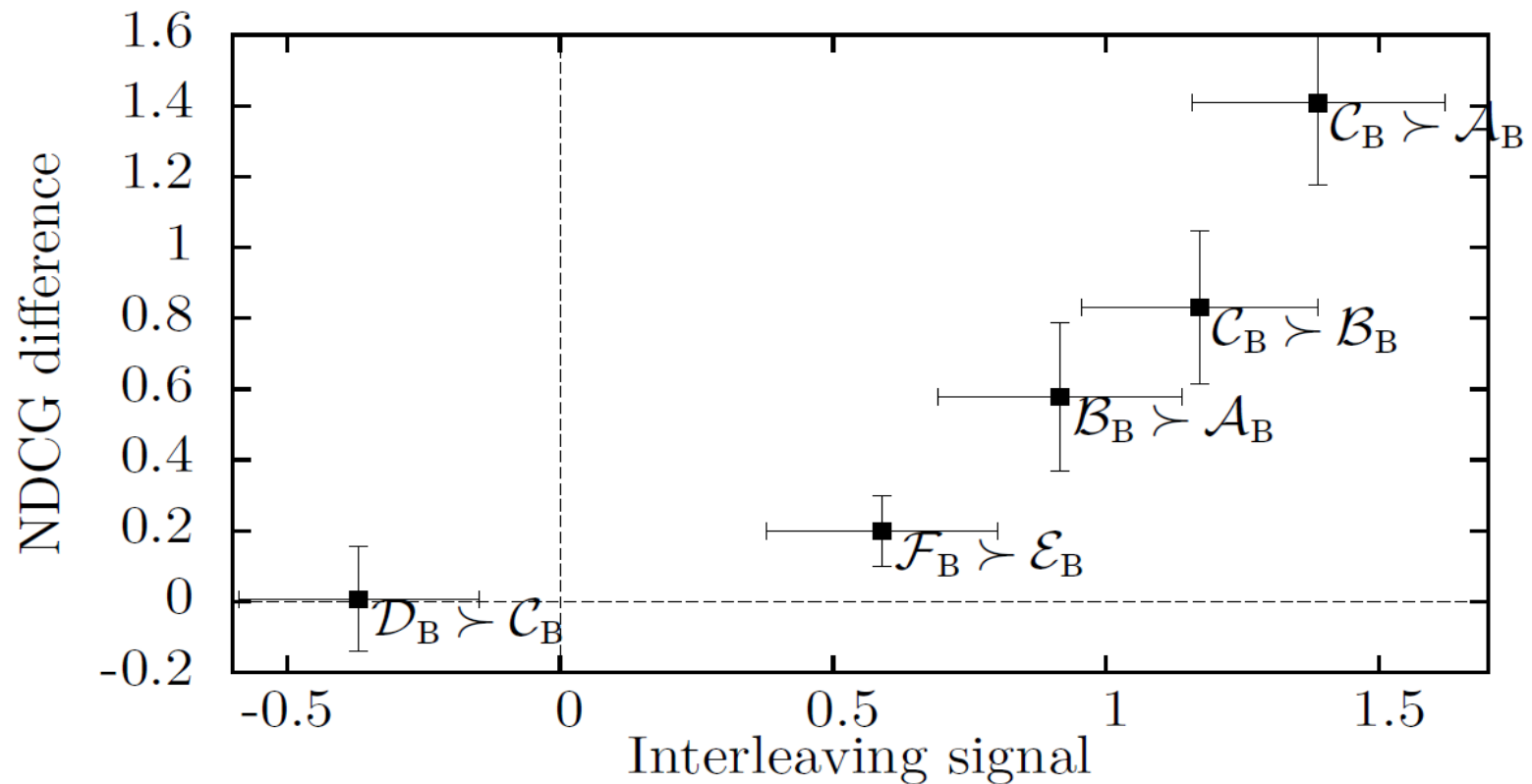
OLIVIER CHAPELLE, Yahoo! Research
THORSTEN JOACHIMS, Cornell University
FILIP RADLINSKI, Microsoft
YISONG YUE, Carnegie Mellon University

Interleaving against Collection-based Evaluation

Experimental Setup

- Selected 4–6 pairs of ranking functions to compare in different settings
 - Known retrieval quality, by construction or by judged evaluation
- Observed user behavior in two experimental conditions
 - Randomly used one of the two individual ranking functions
 - Presented an interleaving of the two ranking functions
- Evaluation performed on three different search platforms
 - arXiv.org (academic paper repository)
 - Bing Web search
 - Yahoo! Web search

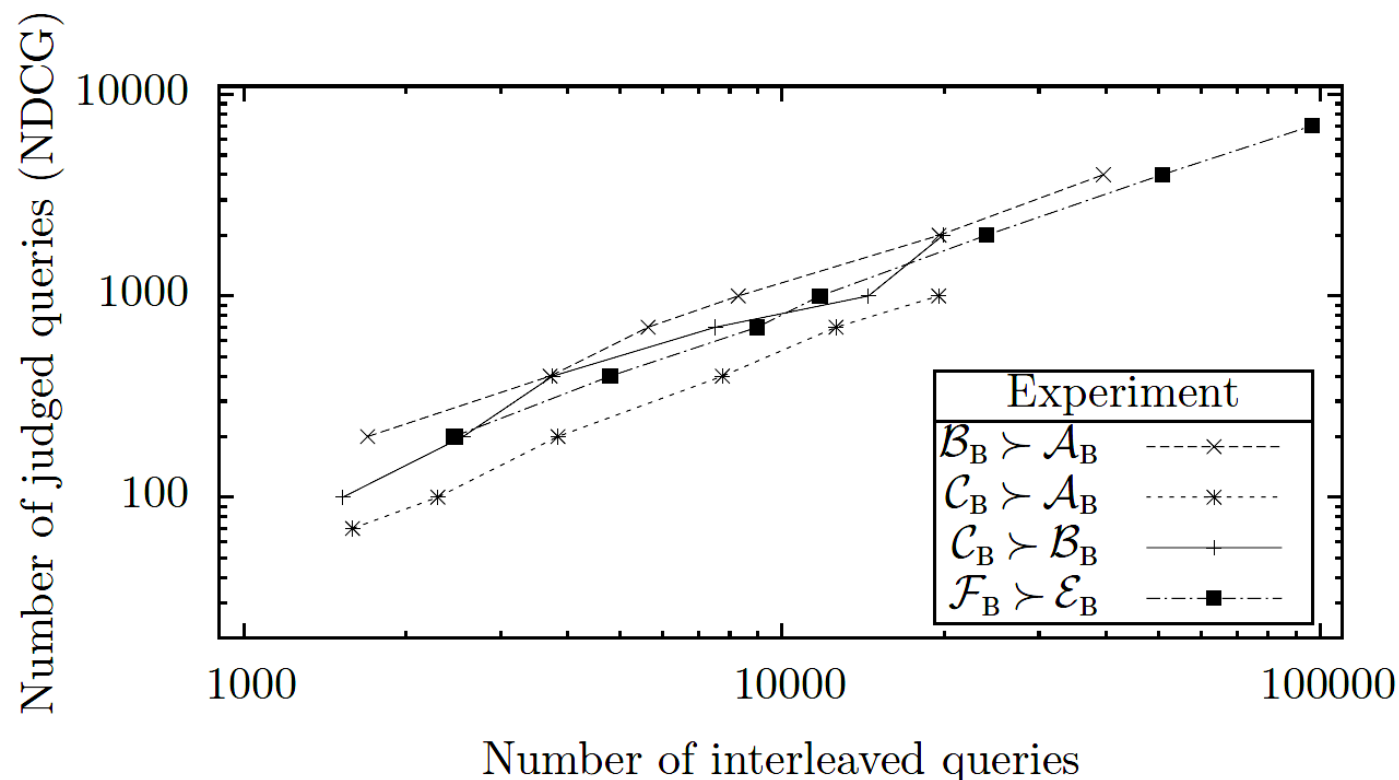
Comparison with Offline Judgments



- Experiments on Bing (large scale experiment)
- Plotted interleaving preference vs NDCG difference
- Good calibration between expert judgments and interleaving

[Radlinski & Craswell 2010; Chapelle et al. 2012]

Comparison with Offline Judgments



- Experiments on Bing (large-scale experiment)
- Plotted queries required vs expert judgments required (for different p-values)
- Linear relationship between queries and expert judgments required
- **One expert judged query is worth ~10 queries with clicks**

Interleaving against A/B Testing

Monotonicity Assumption

- Consider two sets of results: A & B
 - A is high quality
 - B is medium quality
- Which will get more clicks from users, A or B?
 - A has more good results: Users may be **more** likely to click when presented results from A.
 - B has fewer good results: Users may need to click on **more** results from ranking B to be satisfied.
- Need to test with real data
 - If either direction happens consistently, with a reasonable amount of data, we can use this to evaluate online

Testing Monotonicity

How Does Clickthrough Data Reflect Retrieval Quality?

Filip Radlinski
Dept. of Computer Science
Cornell University
Ithaca, NY, USA
filip@cs.cornell.edu

Madhu Kurup*
Dept. of Computer Science
Cornell University
Ithaca, NY, USA
mmk222@cs.cornell.edu

Thorsten Joachims
Dept. of Computer Science
Cornell University
Ithaca, NY, USA
tj@cs.cornell.edu

- Contacted on ArXiv.org, an academic search engine.
- Real users looking for real documents.
- Relevance direction known by construction

ORIG > SWAP2 > SWAP4

- ORIG: Hand-tuned ranking function
- SWAP2: ORIG with 2 pairs swapped
- SWAP4: ORIG with 4 pairs swapped

ORIG > FLAT > RAND

- ORIG: Hand-tuned ranking function, over many fields
- FLAT: No field weights
- RAND : Top 10 of FLAT randomly reordered shuffled

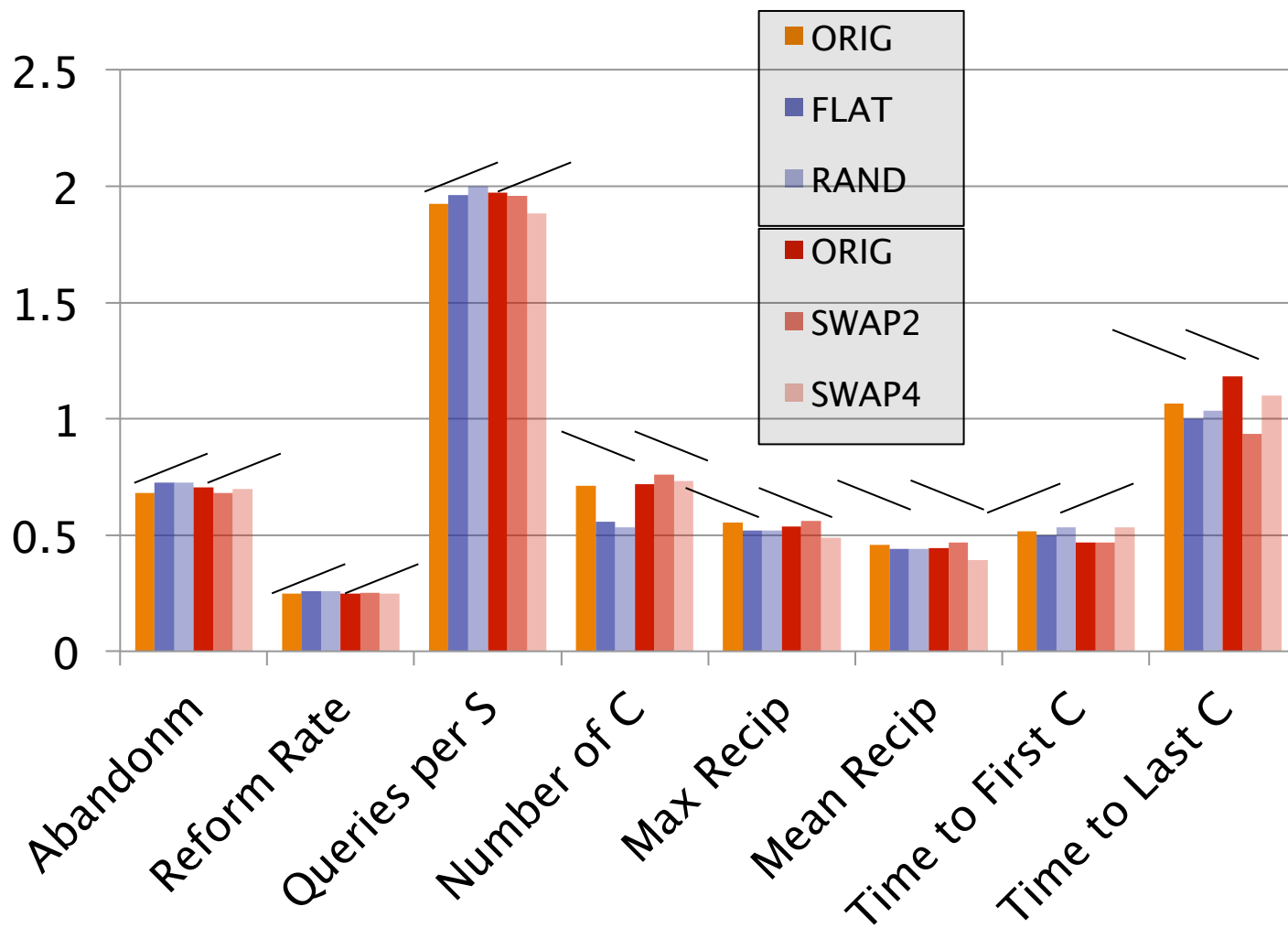
- Evaluation on 3500 x 6 queries

Absolute Metrics

Name	Description	Hypothesized Change as Quality Falls
Abandonment Rate	% of queries with no click	Increase
Reformulation Rate	% of queries that are followed by reformulation	Increase
Queries per Session	Session = no interruption of more than 30 minutes	Increase
Clicks per Query	Number of clicks	Decrease
Clicks @ 1	Clicks on top results	Decrease
pSkip [Wang et al '09]	Probability of skipping	Increase
Max Reciprocal Rank*	1/rank for highest click	Decrease
Mean Reciprocal Rank*	Mean of 1/rank for all clicks	Decrease
Time to First Click*	Seconds before first click	Increase
Time to Last Click*	Seconds before final click	Decrease

(*) only queries with at least one click count

Evaluation of Absolute Metrics on ArXiv.org



Comparative Summary

Method	Consistent (weak)	Inconsistent (weak)	Consistent (strong)	Inconsistent (strong)
Abandonment Rate	4	2	2	0
Clicks per Query	4	2	2	0
Clicks @ 1	4	2	4	0
pSkip	5	1	2	0
Max Reciprocal Rank	5	1	3	0
Mean Reciprocal Rank	5	1	2	0
Time to First Click	4	1	0	0
Time to Last Click	3	3	1	0
Interleaving	6	0	6	0

- Comparison on arXiv.org experiments
- Results on Yahoo! qualitatively similar

Different approaches to evaluation

- User-studies
- Collection-based evaluation
- In-situ evaluation
 - A/B Testing
 - Interleaving

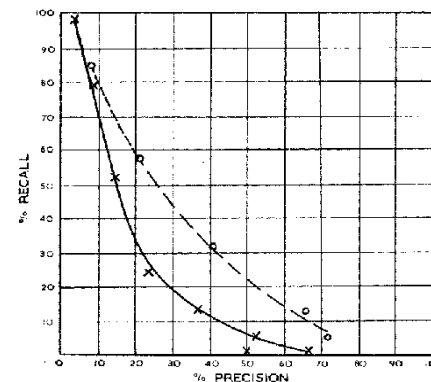
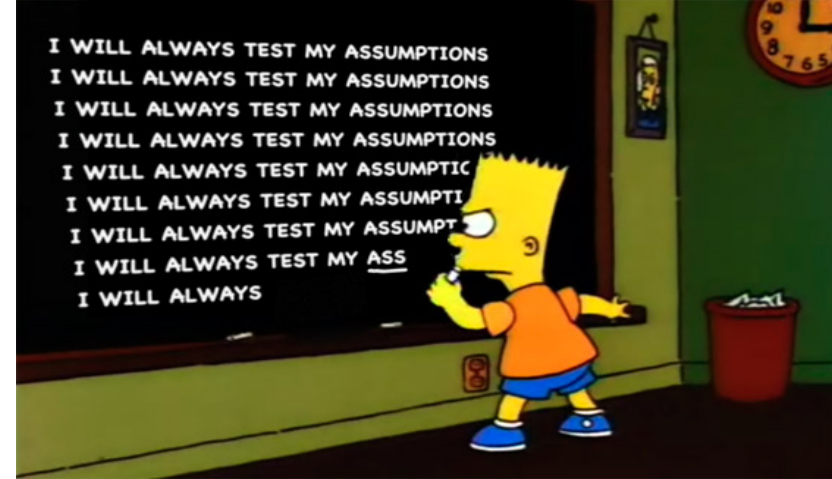


FIGURE 4.814P INDEX LANGUAGE III, 5, 8 SEARCH E
200 DOCUMENTS
(Index 2-language III, 1, 8 Broken line)

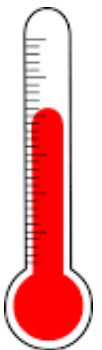


Takeaways

- Don't trust the HiPPO



- Trust the data; hence experiment often
- If you torture the data enough they will confess to anything
- The measure defines the problem



Acknowledgements

- For the material on offline evaluation:
 - Emine Yilmaz (University College London)
 - Ben Carterette (University of Delaware)
- For the material on online evaluation:
 - Anne Schuth (University of Amsterdam)
 - Katja Hofmann (Microsoft Research)
 - Filip Radlinski (Microsoft Bing)

Thank You!

