

Exploring Mobile Centric Challenges & Opportunities in Pro-IR



Manisha Verma

mverma@cs.ucl.ac.uk

University College London



Outline

- ❖ Mobile Search Landscape
- ❖ Challenges in Mobile Search
 - ❖ Mining Touch Interactions
 - ❖ Measuring User Satisfaction
 - ❖ Local Search
- ❖ Characterizing Relevance on Mobile and Desktop
 - ❖ Research Questions
 - ❖ Contributions
 - ❖ Crowd sourced judgments
 - ❖ Features an their correlation
 - ❖ Results
- ❖ Opportunities in Mobile
- ❖ Information Cards
 - ❖ Card Retrieval
 - ❖ Query Modelling
 - ❖ Open Questions
- ❖ Intelligent Assistants
 - ❖ Understanding user satisfaction
 - ❖ Detecting good abandonment
 - ❖ Open Questions

Mobile Search (Then)

Reactive in nature: Search engines simply serve user's search query.

The screenshot shows a desktop browser window with the Google search interface. The search bar contains the text 'cars'. Below the search bar are navigation links for Web, Images, Local, and Mobile Web (Beta). The search results are displayed in a list format, with the first result being 'Buy new & used cars online, research prices & dealers, sell your ...' from Cars.com. The second result is 'Jaguar - Official worldwide web site of Jaguar Cars.'.

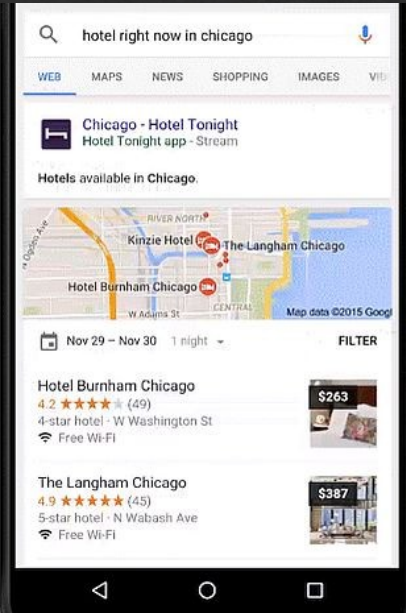
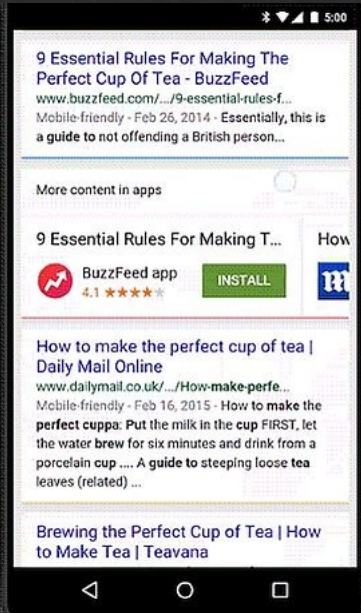
2006



The screenshot shows a mobile browser window with the Google search interface. The search bar contains the text 'Mother's day gift'. Below the search bar are navigation links for Web, Images, Local, and more. The search results are displayed in a list format, with the first result being 'Mother's Day Gift' from PersonalizationMall.com. The second result is 'Mother's Day Gift Ideas' from Gifts.com. The third result is 'Mother's Day gift ideas that give back' from The International Rescue Committee. The fourth result is 'Mother's Day Gift Ideas' from Gifts.com. The fifth result is 'Mother's Day Gift Ideas' from Gifts.com.

2010

Mobile Search (Now)



2016

- ❖ Instant Answers
- ❖ Knowledge Panels
- ❖ Multiple Verticals
 - ❖ Apps
 - ❖ Images
 - ❖ Social
- ❖ Rich Interaction

Mobile Information needs

❖ From Song *et al.* WWW'13

	Mobile (iPhone)
Total Queries	9,732,938
Total Users	1,233,720

	Mobile
Number of words	3.05
Number of characters	18.93

Category	Mobile
Adult	23.5%
Autos	2.4%
Celebrity	8.3%
Commerce	8.6%
Finance	0.4%
Health	1.7%
Image	42.0%
Local	10.3%
Maps	0.1%
Movie	1.6%
Music	3.4%
Name	7.3%
Sports	5.7%
Navigational	15.4%

Mobile
youtube.com
en.wikipedia.org
answers.yahoo.com
ehow.com
imdb.com
amazon.com
wiki.answers.com
chacha.com
facebook.com
myspace.com

CTR for Knowledge Base Sites on

Query	Mobile
hotmail	2.5%
microsoft	16%
usa	33%
facebook	3.5%
louis vuitton	4.3%

Mobile Information needs

❖ From Sohn *et al.*
CHI'08

Need Category	Example
Trivia	"What did Bob Marley die of, and when?"
Directions	"Directions to Sammy's Pizza"
Point of Interest	"Where is the nearest library or bookstore?"
Friend Info	"Where are Sam and Trevor?"
Shopping	"How much does the Pantech phone cost on the AT&T website?"
Business Hours	"What time does the post office close?"
Personal Item	"What is my insurance coverage for cat scans?"
Schedule	"Is there an open date on my family calendar?"
Phone #	"What is the phone # for weight watchers?"
Traffic	"How far does the traffic extend?"
Sports/News/ Stocks	"Did the Miami Heat have any free agent acquisitions?"
Email	"Email update for work"
Movie Times	"Are Harry Potter tickets available tonight?"
Weather	"What will the weather be like this weekend?"
Travel	"Flight status of my Southwest flight"
Recipes	"Needed ingredients for hot and sour soup"

Challenges in Mobile Search

- ❖ Search Result Quality
 - ❖ Mine touch interaction (Lagun *et al.* SIGIR'13)
 - ❖ Understand rich user context (location or preferences etc.)
- ❖ Query Auto Completion (Kamvar *et al.* MobileHCI'07, Vargas *et al.* WSDM'16, Zhang *et al.* WWW'16)
- ❖ Understanding and Predicting Abandonment (Li *et al.* SIGIR'09, Diriye *et al.* CIKM'12, William *et al.* WWW'16)
- ❖ Measuring Search Satisfaction (Lagun *et al.* SIGIR'14)

Mining Touch Interactions

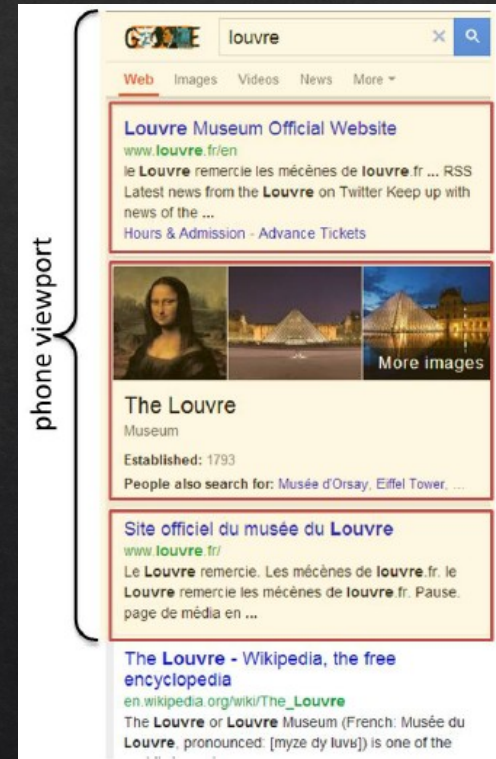
- ❖ Evaluate touch interactions on a smart phone as implicit relevance feedback.
- ❖ Conduct a user study with seven search tasks.
- ❖ Mobile users swipe much more than desktop users scroll.
- ❖ Longer Dwell time and inactivity on Mobile.
- ❖ Gesture count indicative of searcher dissatisfaction.
- ❖ Faster swipe indicates low relevance.

<i>Feature</i>	<i>Mobile</i>
<i>dwell time</i>	44.3 (36.2)
<i>swipe/scroll distance</i>	2808.8 (4007.6)
<i>swipe/scroll speed</i>	66.9 (96.1)
<i>swipe/scroll maximum</i>	1352.6 (1915.4)
<i>inactive percentage</i>	0.59 (0.21)
<i>total inactive time</i>	28.0 (27.2)
<i>average inactive time</i>	4.3 (6.2)
<i>maximum inactive time</i>	14.8 (18.5)
<i>zoom count</i>	0.89 (2.71)
<i>touch size</i>	0.15 (0.06)

<i>Feature</i>	<i>Mobile</i>
<i>dwell time</i>	0.171*
<i>gesture count</i>	-0.042
<i>gesture frequency</i>	-0.201*
<i>touch size</i>	0.032
<i>touch pressure</i>	0.020
<i>swipe/scroll count</i>	-0.043
<i>swipe/scroll frequency</i>	-0.210*
<i>swipe/scroll distance</i>	-0.065
<i>swipe/scroll maximum</i>	-0.052
<i>swipe/scroll speed</i>	-0.165*
<i>zoom count</i>	0.026
<i>zoom frequency</i>	-0.008
<i>zoom distance</i>	0.039
<i>zoom maximum</i>	0.021
<i>inactive percentage</i>	0.244*
<i>total inactive time</i>	0.232*
<i>average inactive time</i>	0.276*
<i>maximum inactive time</i>	0.284*

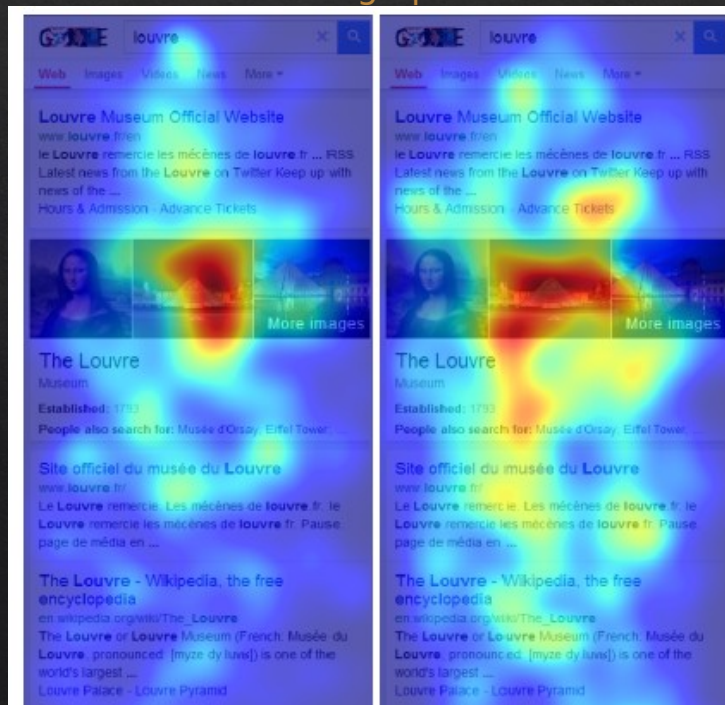
Measuring User Satisfaction

- ❖ Complex search results today do not require users to click.
- ❖ Conduct an eye tracking study to measure attention and satisfaction.
- ❖ Vary instant answer presence and relevance.
- ❖ Six queries for Instant answers and Knowledge panels.
- ❖ Scroll and interaction beyond answer panels indicate dissatisfaction.
- ❖ Knowledge graph panels (presence & relevance) yield higher satisfaction.

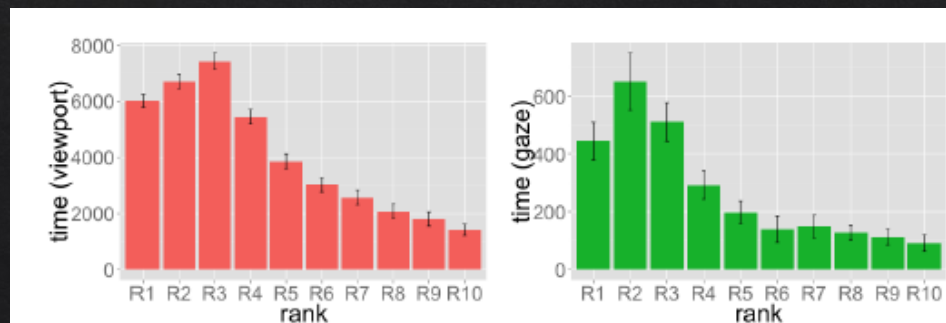


Measuring User Satisfaction

Attention heatmaps for relevant and irrelevant knowledge panel results.



- ❖ Users spend more viewport time and gaze time on KG results when they are irrelevant compared to the relevant KG results.
- ❖ Increased attention does not indicate relevance.
- ❖ User perform short scrolls, results in some results visible on viewport longer.
- ❖ Viewport time poor proxy of user attention.



Local Search

Location (geographic features) and time (temporal aspects) are important in Mobile Search (Teevan et al.)

- ❖ Location-aware Click Prediction (Lymberopoulos *et al.* CIKM'11)
- ❖ Local Search using Community Behavioral Modeling (Lane *et al.* UbiComp'10)
- ❖ Place visits and preferences (Froehlich et al. UbiComp'06)
- ❖ Use of External Logs (Berberich et al. SIGIR'11)
- ❖ Deeper understanding of location based signals (Lv et al. SIGIR'12)

Characterizing Relevance on Mobile

A popular website today has **at least two views -- Mobile and Desktop**

- 1. Mobile : Screen size limits both the layout & amount of visible content.**
- 2. Browser Improvement :** Minimal change in HTML can yield mobile and desktops layout.

A user may see **different** versions of **same website** on **different devices.**

May impact judgment of page with respect to a query

A sneak peek into page interaction

Best selling books

Best-Selling Books: The annual top 100

By Anthony DeBarros, Mary Cadden, Kristin DeRamus and Christopher Schnaars, USA TODAY
Updated 5/20/2011 4:07 PM

The top 100 titles of 2010, compiled from USA TODAY's [Best-Selling Books list](#)

Select another year: 2010

RANK	TITLE	AUTHOR	DESCRIPTION
1	The Girl With the Dragon Tattoo	Stieg Larsson	Journalist is hired to investigate the disappearance of an heir to a wealthy family (F)
2	The Girl Who Played With Fire	Stieg Larsson	An Eastern European sex-trafficking ring involving prominent Swedes propels this sequel to "The Girl With the Dragon Tattoo" (F)
3	The Girl Who Kicked the Hornet's Nest	Stieg Larsson	Lisbeth Salander is accused of murder and turns to friend and journalist Mikael Blomkvist to help prove her innocence; final in series (F)
4	Decision Points	George W. Bush	The former president describes key decisions, including the wars in Iraq and Afghanistan and the financial crisis (NF)
5	The Last Song	Nicholas Sparks	The story of a teenage girl distraught over her parents' divorce; movie (F)
6	Diary of a Wimpy Kid: The Ugly Truth	Jeff Kinney	Youth; Greg learns to deal with boy-girl parties; fifth in series (F)

Showing 1 to 100 of 100 entries
Source: USA TODAY Best-Selling Books

Search: _____

Most Popular

Stories

Ed Baig reviews Kindle Paperwhite
"Pregnant man" struggles through nasty divorce
Tennis Channel Court Report 9-30-2012

Photos

2012 Emmy Awards Red Carpet

Most Popular E-mail Newsletter

Sign up to get:
Top viewed stories, photo galleries and community posts of the day

Most popular right now:
HF Test

Sign up for USA TODAY E-mail newsletters

Who's hot, who's not?

Best-Selling Books: The annual top 100

USA TODAY

Best-Selling Books: The annual top 100

1 The Girl With the Dragon Tattoo

2 The Girl Who Played With Fire

3 The Girl Who Kicked the Hornet's Nest

4 Decision Points

5 The Last Song

6 Diary of a Wimpy Kid: The Ugly Truth

Showing 1 to 100 of 100 entries

Source: USA TODAY Best-Selling Books

A sneak peek into page interaction

- **Worst drought in US**

The screenshot shows the desktop version of the article. The URL is www.livescience.com/21844-worst-droughts-in-u-s-history.html. The page features a navigation bar with categories like TECH, HEALTH, PLANET EARTH, SPACE, STRANGE NEWS, ANIMALS, HISTORY, HUMAN NATURE, and SHOP. The main article title is "The Worst Droughts in U.S. History" by Stephanie Pappas, dated July 25, 2012. A navigation bar below the title shows "Intro" as the selected section, with other sections numbered 5, 4, 3, 2, 1. The article content includes a photo of a cow and a paragraph about drought conditions. On the right side, there are several widgets: a "Science Newsletter: Subscribe" form, "Follow Us" social media icons for Facebook, Twitter, LinkedIn, and Google+, and a "Most Popular" section with four article thumbnails.

The screenshot shows the mobile version of the article. The URL is m.livescience.com/21844-worst-droughts-in-u-s-history.html. The page is optimized for a smaller screen, with a dark navigation bar at the top. The article title "Drought-Stricken Longhorn" is visible, along with a sub-headline "A scrawny Longhorn in Big Bend State Park, West Texas, shows how levels of vegetation have thinned and wildlife struggling to find food." Below the article content, there is a "Ready for a CRM solution?" advertisement. At the bottom, there are social media sharing icons for Facebook, Twitter, and Google+.

Research Questions

1. Can **different web page layouts** (on mobile or desktop) yield **different relevance labels**?
2. How are **pages** on two mediums: **Mobile and Desktop** judged?
3. Can **viewport specific signals** be used to **determine page relevance**?

Contributions

1. Study **crowd sourced judgments** of query-url pairs on mobile and desktop.
2. Study **efficacy of novel viewport oriented features** in predicting page relevance on both mobile and desktop.
3. Analyze **which features are strong signals** to determine relevance on both devices.

Crowdsourced Judgments

DATASET

Judgments on desktop and mobile for **236 query-url pairs**.

PAY

3 cents per hit. For a total of 1416 hits, we paid 45 dollars

WORKERS

Desktop hits completed by **41** workers. **Mobile** hits completed by **28**.

JUDGE SELECTION

Acceptance Rate: 95%
Demographics: US
Masters Requirement

INTER-RATER AGREEMENT

Desktop: **0.28 (Fair)**
Mobile: **0.33 (Fair)**

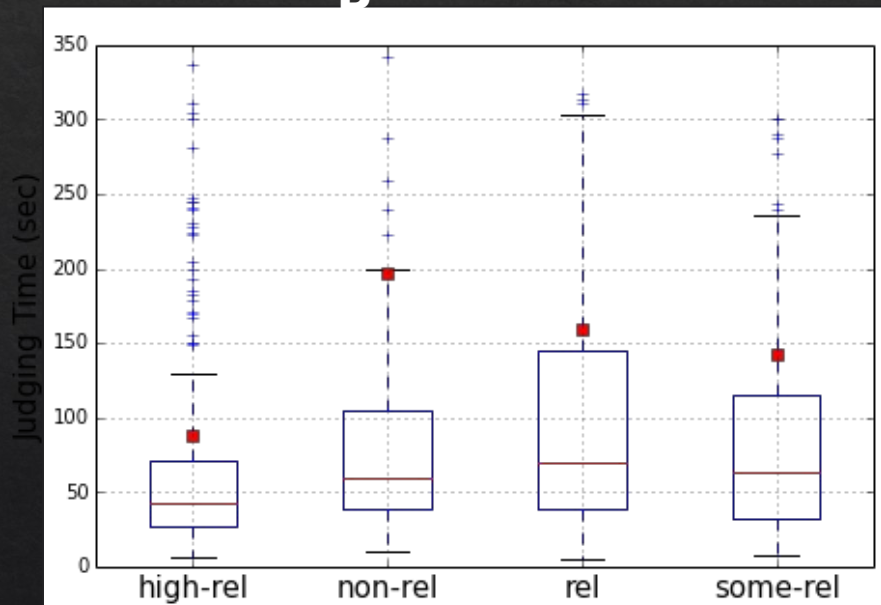
DESKTOP-MOBILE AGREEMENT

Cohen's Kappa: 0.127
Tau: 0.114 (p-val=0.01)

Judging Times vs Relevance Grades

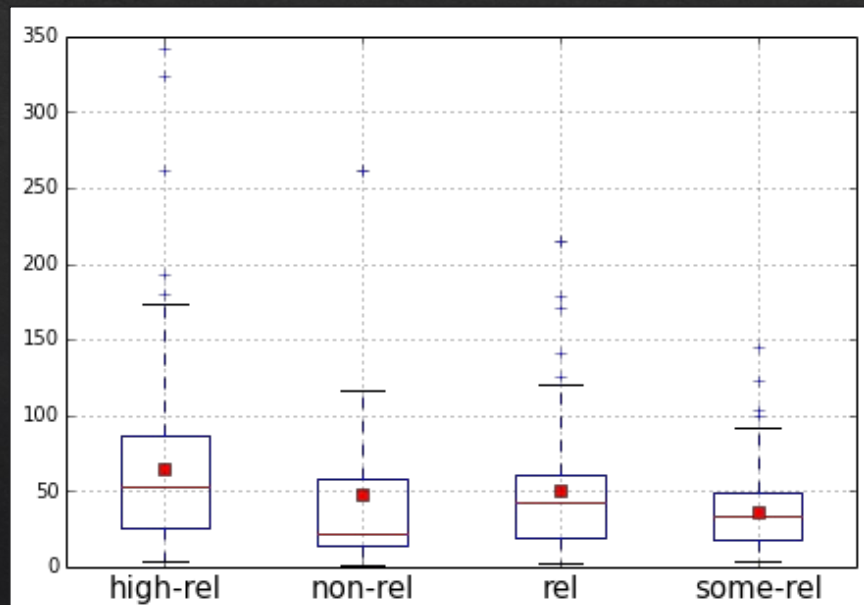
DESKTOP

P



High-rel: 88s
Some-rel: 142s
rel: 159s
Non-rel: 197s

MOBILE



High-rel: 65s
Some-rel: 31s
rel: 51s
Non-rel: 47s

Relevance Prediction

- ❖ **Estimating page relevance** from set of features or user interaction signals is an important task.
- ❖ Given the **observed difference in judgments**, we posit that **different features will be useful** for predicting **relevance on mobiles and desktop**.
- ❖ We **propose two classes of features** and investigate their role in predicting relevance.

Features

VIEWPORT

WHOLE PAGE

Content Specific Features

- ❖ Number of headings
- ❖ Number of images
- ❖ Number of words
- ❖ Number of sentences
- ❖ Query term frequency
- ❖ Unique tokens

Geometric Features

- ❖ Query term position
- ❖ Heading position

Display Specific Features

- ❖ Query term size
- ❖ Heading size
- ❖ Word size

Feature Correlation With Judgments

DESKTOP			MOBILE		
Feature	R	p-val	Feature	R	p-val
Number of sentences (html)	-0.13	0.04	Heading size (max)	0.16	0.01
Number of words (html)	-0.12	0.03	Number of headings (viewport)	0.16	0.02
Unique tokens (html)	-0.12	0.04	Number of words (viewport)	0.14	0.04
Query term size (mean)	0.26	0.00	Number of headings (html)	0.13	0.04
Query term position (min)	-0.17	0.01	Number of images (viewport)	0.10	0.04

Results

	Mobile		Desktop	
	Accuracy	F1-score	Accuracy	F1-score
all	0.76	0.839	0.60	0.71
no.html	0.75	0.82	0.65*	0.75*
no.view	0.73	0.81	0.61	0.73
no.geom	0.79*	0.86*	0.67*	0.75*
no.display	0.78	0.85	0.64*	0.74
only.html	0.74	0.83	0.61	0.72
only.view	0.81*	0.87*	0.64*	0.74
only.geom	0.76	0.84	0.54	0.67
only.display	0.71	0.80	0.64*	0.74
geom.display	0.75	0.82	0.61	0.72
geom.html	0.77	0.85	0.57	0.69
geom.view	0.78	0.85	0.62	0.73
view.html	0.78	0.86*	0.67*	0.76*

Key take aways

- ❖ Desktop relevance grade are not indicative of mobile relevance.
- ❖ Perception of document relevance is not device agnostic. That is, page layout and page rendering influences relevance.
- ❖ Viewport features are useful for Relevance prediction on both devices, more so on Mobiles.

Opportunities in Mobile

- ❖ Users spend increasingly more time on mobiles.
- ❖ Users install multiple apps and browse significantly more pages.
- ❖ Mobiles offer more interaction: Camera, microphone, accelerometer to name a few.

With multiple apps, user context and interaction data, can we move from a reactive to proactive IR?

- ❖ proactive, non-intrusive and contextual

Future information retrieval systems must anticipate user needs and respond with information appropriate to the current context without the user having to enter a query (2012 Workshop)

Proactive IR

❖ Information Cards

- ❖ From Queries to Cards (Shokouhi *et al.* SIGIR'15)
- ❖ Modelling User Interests for Zero-query Ranking (Yang *et al.* ECIR'16)

❖ Intelligent Assistants

- ❖ Understanding User Satisfaction with Intelligent Assistants (Kiseleva *et al.* CHIIR'16)

Information cards

Boarding Pass

Lufthansa 9228
Operated by United Airlines

Name: Mr John Smith Booking number: E12345678



From SFO To JFK Group 7 Seat 23B

Gate 51a Terminal 2 Depart SFO at 11:45...
Tue, May 21, 2013

Ticket type: World Perk Rewards Premier Access

 [View email](#)

Traffic & Transit

57 mins to work

Normal traffic on US - 101

Relevant website

Things to do in
Sausalito: Check out 9
Sausalito Attractions
www.tripadvisor.com



Keep me updated

New content available

Get presale tickets for Lord Huron on
Feb. 25!
The Fillmore - 5 hours ago

Kimo West - Traditional
and Original Slack Key...
Slack Key Show Upcoming
Events - May 7, 2014



Activity Summary

25 miles in August

6 miles less than July

Summary of July & August



Based on your device's location which is periodically sent to Google. [Learn more](#)

TV

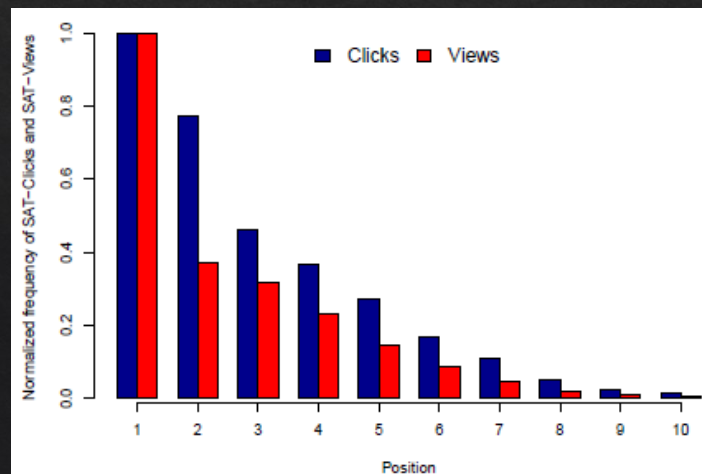
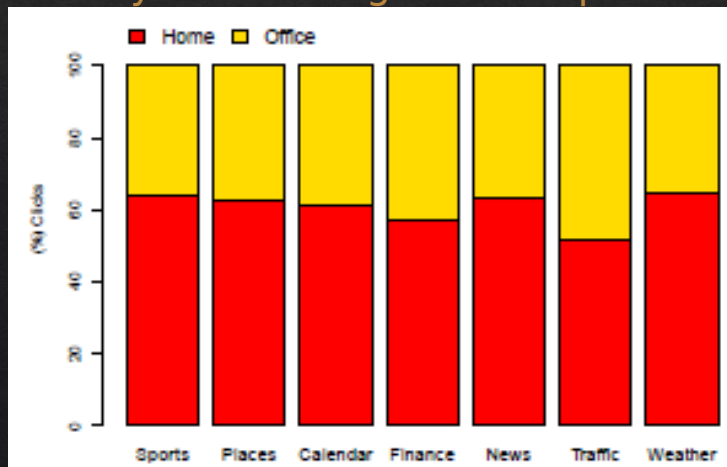
Watch now

The Englishman, North and South, Gator Man, News Today, and more on TV and video on demand



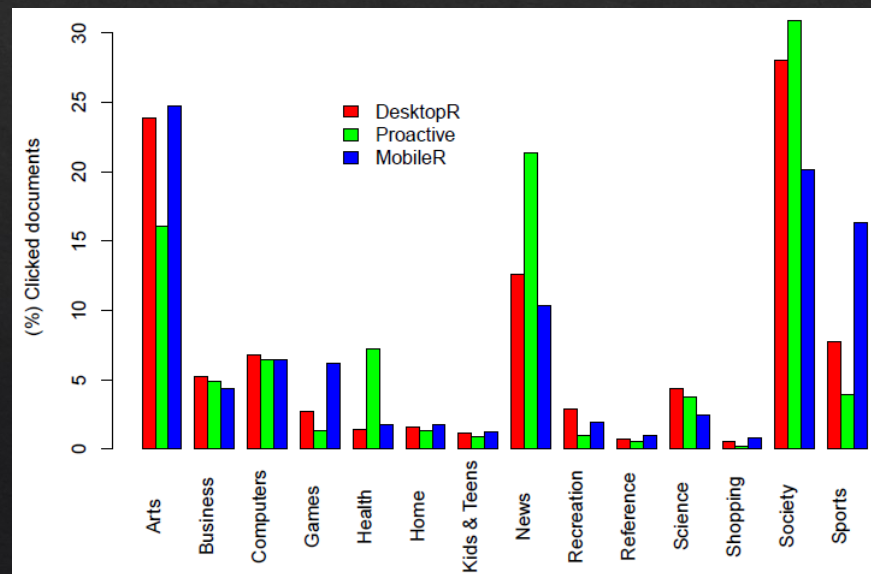
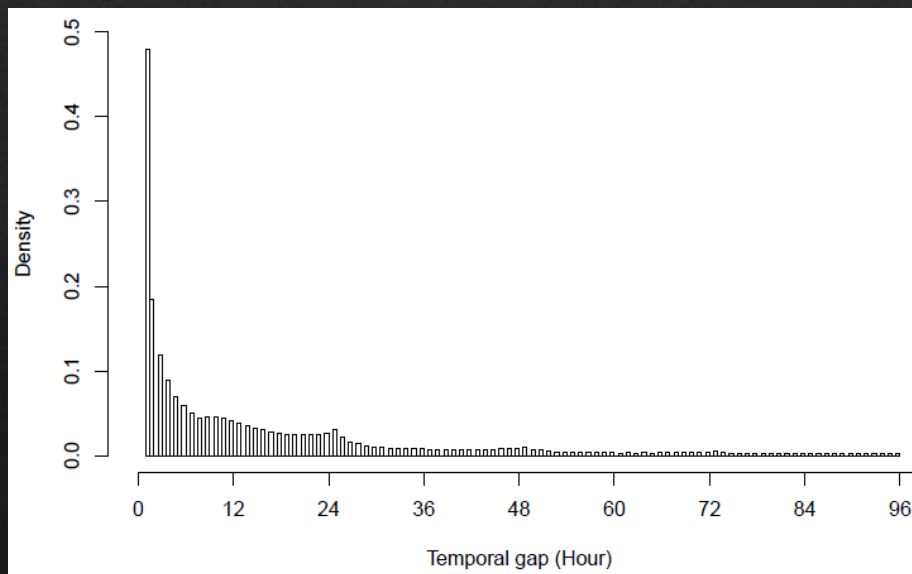
Information Cards

- ❖ Investigate user engagement with different cards.
- ❖ 0.3 million users and 2.6 million impressions.
- ❖ Seven types of cards namely: sports, places, calendar, finance, news, traffic and weather.
- ❖ Card usage varies during the day. Traffic on the weekends, sports on weekdays.
- ❖ Relatively little change in card positions.



Information Cards

- Once the user clicks on a card, the next clicked impression is most likely to appear within an hour.
- Users' reactive history might provide useful signals for ranking their proactive recommendations



Information Cards

- ❖ Click-based (pseudo) relevance labels may not be appropriate for evaluating all types of cards.
- ❖ Focus on predicting clicks for News cards only.
- ❖ Features
 - ❖ Reactive History
 - ❖ Proactive History
 - ❖ Lexical/Topical Features
 - ❖ Local/Temporal Features
 - ❖ Constant Features

Information Cards

	ΔMRR (\mathcal{R}_c)
Static-Ranker	▼ -0.72%
Carré	▲ 10.85%

❖ Lexical features the most important.

	ΔMRR (\mathcal{R}_c)
<i>Head users</i>	
Static-Ranker	▲ 2.12%
Carré	▲ 19.69%
<i>Tail users</i>	
Static-Ranker	▼ -1.57%
Carré	▲ 3.96%

❖ Temporal features have least impact.

❖ Reactive search history useful in predicting clicks.

Research Directions

- ❖ Generating rich information cards
 - ❖ Mine information across several sources to construct rich cards
- ❖ Evaluating engagement with cards
 - ❖ How do we use and port desktop engagement metrics?
- ❖ Personalizing information on cards
 - ❖ How do we combine information from user's environment to timely construct and prompt user with the card ?
- ❖ Social card recommendations
 - ❖ Can we recommend cards from users' network?

Intelligent Assistants

Something I can do for you, Brian?

try "Will it be sunny in Paris tomorrow?"
see more

ask me anything

Here's the current weather.

56572, MN

28° Now Mostly Cloudy

MON	TUE	WED	THU	FRI
32° 16° 0 0%	38° 26° 0 20%	34° 21° 0 40%	38° 25° 0 0%	44° 31° 0 0%

Data from iMap Weather

Search for 'what is the temperature' ...

Thinking...

Who makes the best smart phone in the world?

Seriously?

Yes seriously

Do I need an umbrella today?

It doesn't look like it's going to rain :

North Sydney
Partly Cloudy

24°

Wednesday Today 28 20

Time	Weather	Temp
11 am	Sunny	25
12 pm	Sunny	26

Intelligent Assistants

- ❖ User study designed to measure user satisfaction over a range of typical scenarios of use: controlling a device, web search, and structured search dialogue.
- ❖ Study how user satisfaction varies with different usage scenarios.
- ❖ What signals can be used for modelling satisfaction in the different scenarios.

- ❖ Observe that notion of satisfaction varies across different scenarios.
- ❖ In some scenarios (e.g. making a phone call), task completion is very important while for others (e.g. planning a night out), the amount of effort spent is key.

Research Directions

- ❖ Evaluating Satisfaction (Jiang *et al.* WWW'15)
- ❖ Evaluating Good Abandonment

- ❖ Supporting complex dialogues
- ❖ Task Repetition (Song *et al.* WWW'16)
- ❖ Task based evaluation of agent

- ❖ Long term learning

Thank you!