

YAHOO!

Evaluating the search experience: from Retrieval Effectiveness to User Engagement

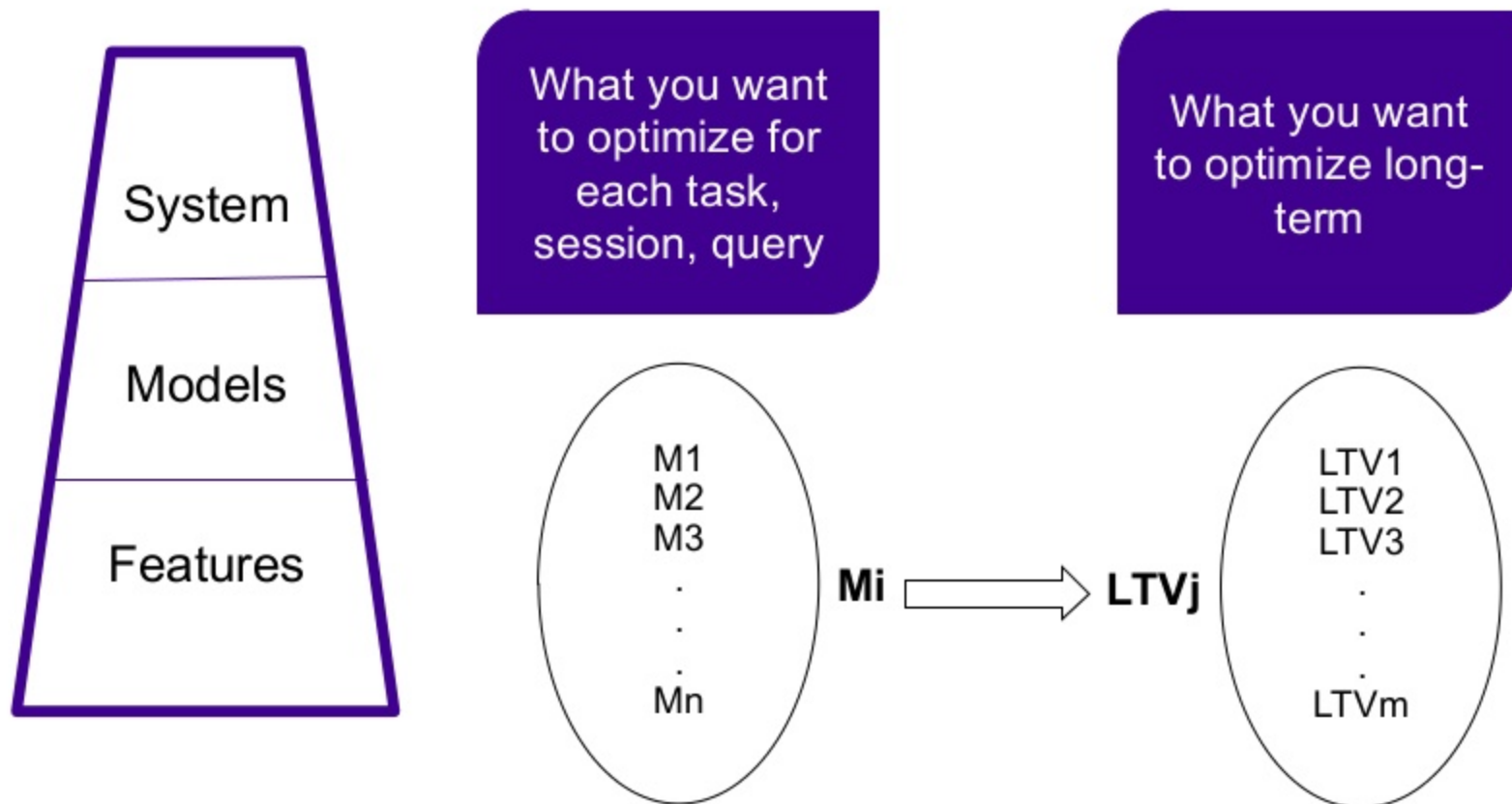
Mounia Lalmas
Yahoo Labs London
mounia@acm.org

This talk

- Evaluation in search
 - (offline evaluation)*
 - (online evaluation)*
 - Interpreting the signals
-

- Introduction to user engagement
 - From retrieval effectiveness to user engagement
 - (from intra-session to inter-session evaluation)*
-

The Message of this talk



Evaluation in search

How to evaluate a search system

- Coverage
- Speed
- Query language
- User interface



- User happiness
Users find what they want and return to the search system
- But let us remember:
In carrying out a search task, search is a means, not an end

Within an online session

- › July 2012
- › 2.5M users
- › 785M page views
- › Categorization of the most frequent accessed sites
 - 11 categories (e.g. news), 33 subcategories (e.g. news finance, news society)
 - 760 sites from 70 countries/regions

Cat.	Subcat.	%Sites	Description
news 22.1%	news	5.79%	society
	news (soc.)	5.13%	
	news (sport)	2.63%	
	news (enter.)	2.24%	music, movies, tv, etc.
	news (finance)	1.97%	health, housing, etc. technology
	news (life)	1.58%	
	news (tech)	1.58%	
search 15.3%	news (weather)	1.18%	
	search	12.63%	search for lyrics, jobs, etc.
	search (special)	1.58%	
service 11.6%	directory	1.05%	
	service	7.63%	translators, banks, etc.
	maps	3.03%	bookmarks, calendar, etc.
sharing 9.6%	organization	0.92%	
	blogging	3.55%	collaborative creation and collection of content
	knowledge	3.55%	
navi 9.3%	sharing	2.50%	sharing of videos, files, etc.
	front page	6.58%	personalized front pages
	front page (pers.)	1.84%	
support 8.7%	sitemap	0.92%	
	support	1.58%	sites that provide products and support for them
	download	7.11%	downloading software
shopping 7.9%	shopping	4.34%	sites to compare prices of products
	auctions	2.11%	
	comparison	1.45%	
leisure 5.7%	adult	2.76%	sites with music, tv, etc.
	games	1.97%	
	entertainment	0.92%	
mail 3.9%	mail	3.95%	
social 3.0%	social media	1.97%	
	dating	1.05%	
settings 2.9%	login	1.71%	profile setting, site person- alization
	settings	1.18%	

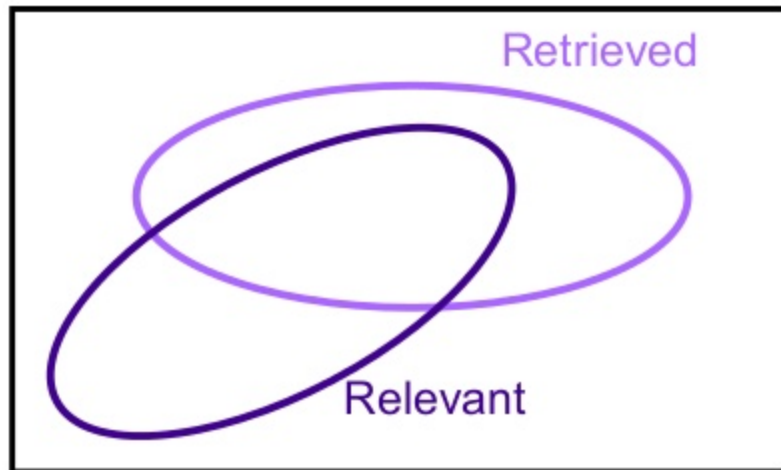
short sessions: average 3.01 distinct sites visited with revisitation rate 10%

long sessions: average 9.62 distinct sites visited with revisitation rate 22%

YAHOO!

Measuring user happiness

Most common proxy: *relevance* of retrieved results

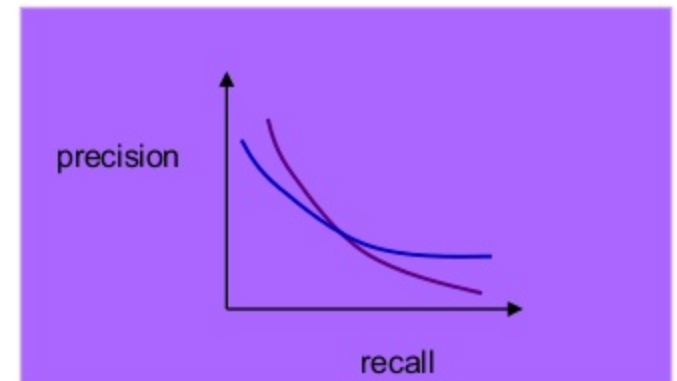


all items

- User **information need** translated into a **query**
- Relevance assessed relative to **information need** *not* the **query**
- Example:
 - › Information need: *I am looking for tennis holiday in a country with no rain*
 - › Query: *tennis academy good weather*

Evaluation measures:

- precision, recall, R-precision; precision@n; average precision; F-measure; ...
- bpref; cumulative gains, rank-biased precision, expected reciprocal rank, Q-measure, ...



Measuring user happiness

Most common proxy: *relevance* of retrieval results



Explicit signals

- Test collection methodology (TREC, CLEF, ...)

- Human labeled corpora

Implicit signals

- User behavior in online settings (clicks, skips, ...)

Explicit and implicit signals can be used together

Examples of implicit signals

- Number of clicks
- SAT click
- Quick-back click
- Click at given position
- Time to first click
- Skipping
- Abandonment rate
- Number of query reformulations
- Dwell time
- Hover


YAHOO! FRANCE

toulouse france × Rechercher

[Toulouse — Wikipédia](#)
fr.wikipedia.org/wiki/Toulouse En cache
Toulouse est située dans le Midi de la France, au nord du département de la Haute-Garonne, sur l'axe de communication entre la mer Méditerranée et l'océan Atlantique

[Tourisme à Toulouse : Vacances, Séjours et Sorties...](#)
www.toulouse-tourisme.com En cache
Préparez votre séjour sur le site officiel du Tourisme à Toulouse : Séjour, activités, sorties... Des centaines d'idées et bons plans pour réussir vos vacances !

[Toulouse France - Images](#)



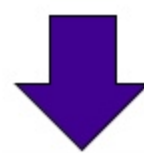
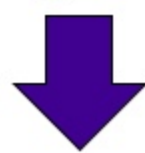
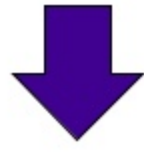
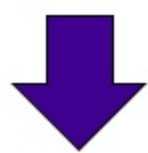
Plus d'images pour Toulouse France

[Toulouse - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/wiki/Toulouse En cache
Toulouse is the fourth largest city in France, after Paris, Marseille and Lyon, and the fourth-largest metropolitan area after Paris, Lyon, and Marseille.

YAHOO!

What is a happy user in search

1. The user information need is satisfied
2. The user has learned about a topic and even about other topics
3. The system was inviting and even fun to use



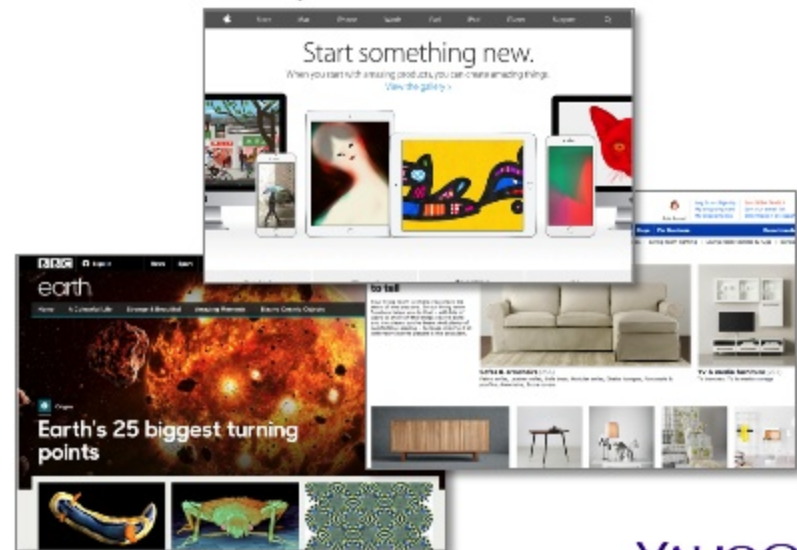
USER ENGAGEMENT

In-the-moment engagement

Users on a site

Long-term engagement

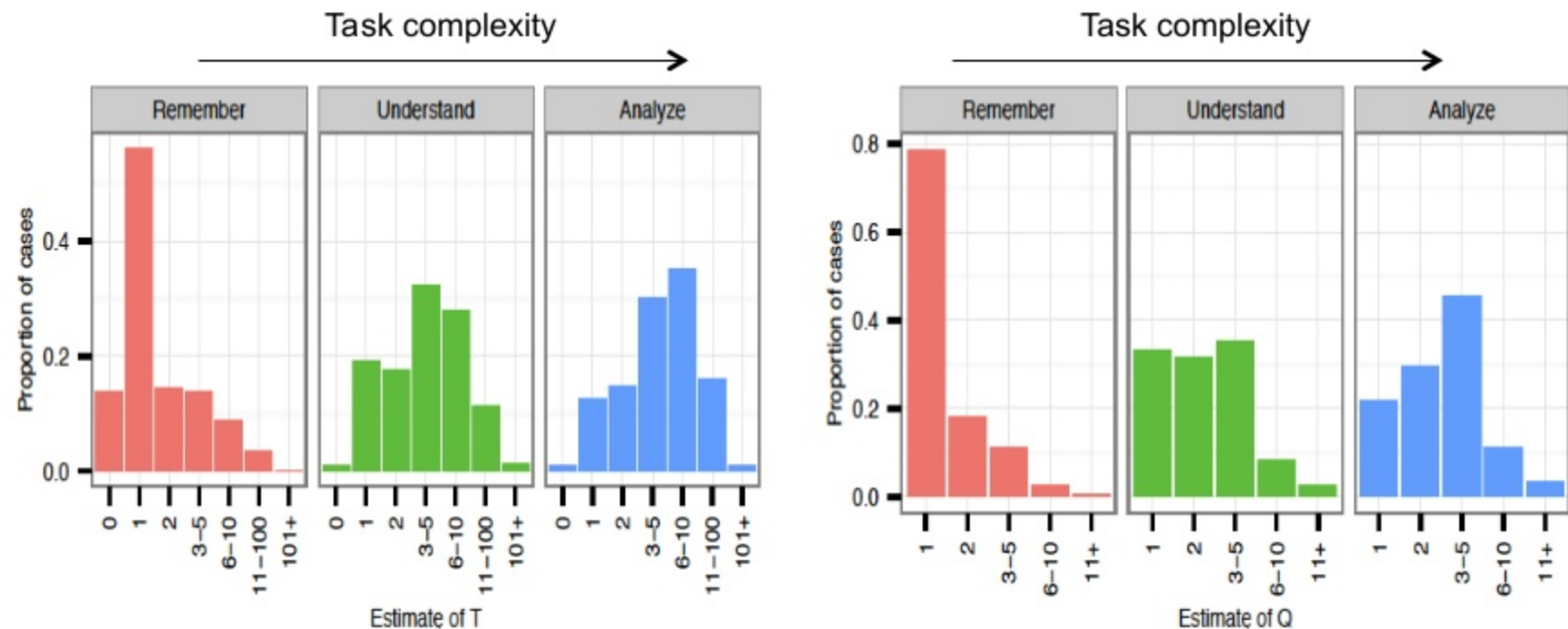
Users come back frequently



YAHOO!

Interpreting the signals

User variability



T: number of documents users (judges) expected to read
Q: number of queries users (judges) expected to issue

Explicit signal: MAP

PRECISION-BASED SEARCH

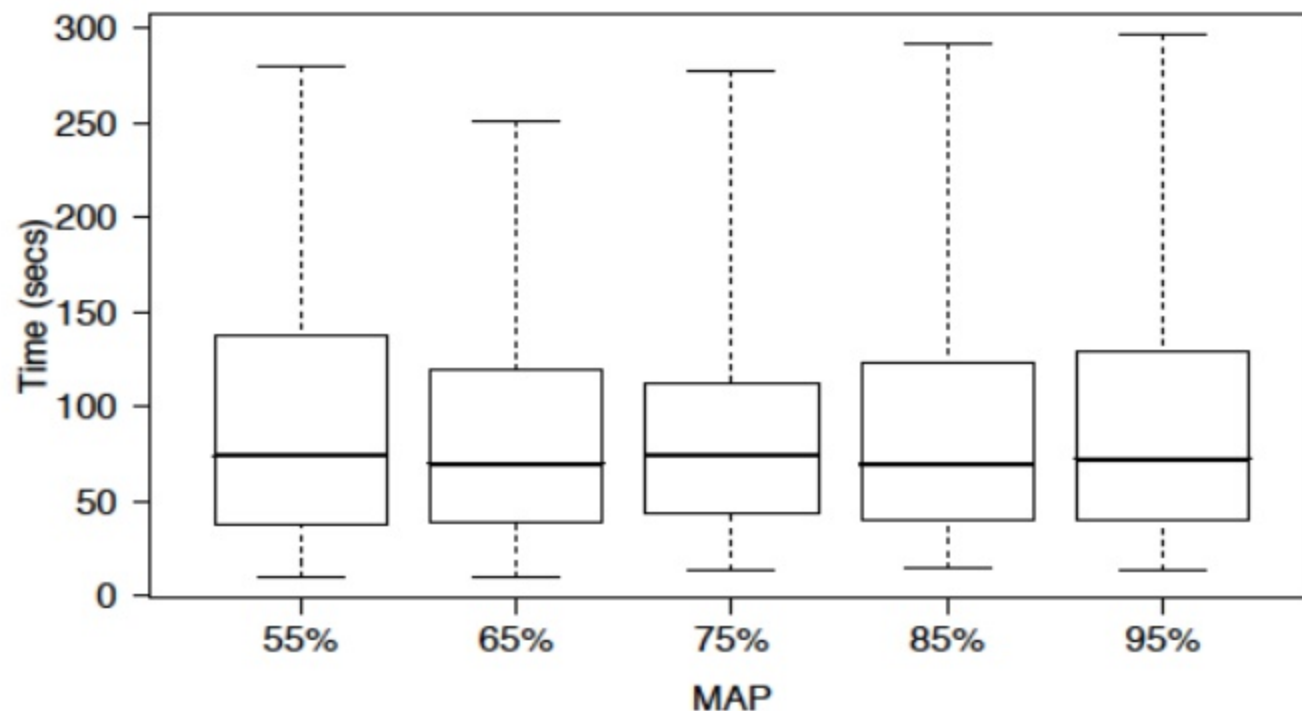


Figure 3: Time taken to find the first relevant document versus the mean average precision of the system used.

Similar results obtained with $P@2$, $P@3$, $P@4$ and $P@10$

Explicit signal: MAP (2)

RECALL-BASED SEARCH

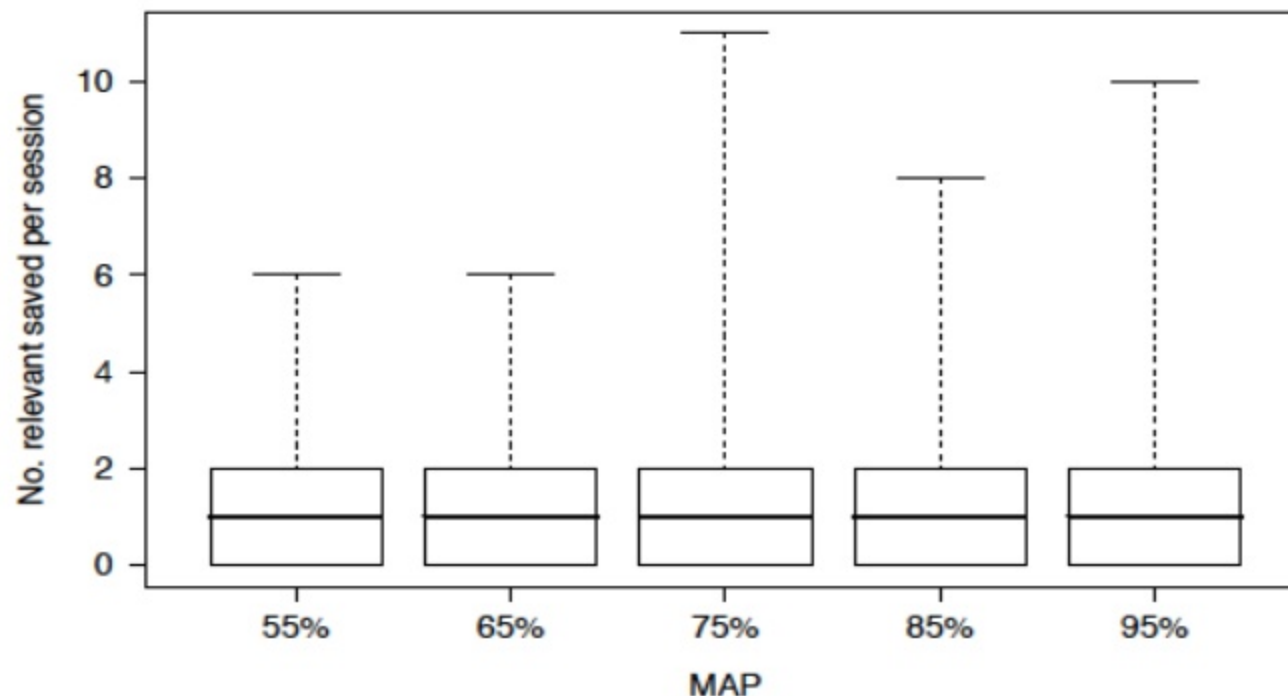


Figure 7: Number of relevant documents found by users within five minutes for systems with differing MAP.

Explicit signal: “Diversity”

Rosario Góngora @rosarioge Situación insólita en Liceo B 36 de Recoleta, 0 votos en 16 mesas. #vota2012 http://t.co/QVszZBFY	21:38, 2012-10-28
Alberto Fouillieux M @afouillieux Chile Rating de los canales demuestra el desinterés de la gente. CHV 6,4 puntos Canal 13 5,4 pts; TVN 5, Mega 3,4. Mucho despliegue por poco rating	20:57, 2012-10-28
fran torres carrasco @frantorresc Santiago -Chile bueno si gana josefa , labbe tiene pega segura en los 80's	22:10, 2012-10-28
Juan Manuel Astorga @jumastorga Santiago, Chile Cero votos en tres mesas de Recoleta a una hora del cierre: http://t.co/Luu09gA4 (Vía @Cooperativa) #abstención #municipales2012	20:44, 2012-10-28
CNN Chile @cnnchile Santiago, Chile #ProyeccionesCNNBioBio para Providencia: Josefa Errázuriz 52,58% - Cristián Labbé 47,42% http://t.co/oD1tJeCX	22:05, 2012-10-28
LORETO ARAVENA @loretoaravena CHILE Montón de propaganda electoral por acá en Santiago Centro! Ojo, si los candidatos no cumplen la ley no esperemos que cumplan sus promesas!!	18:15, 2012-10-28

top most popular tweets

Evelyn Espinoza @evecielito Tanta queja, tanta marcha y casi nada de gente joven votando, después no se queje #Municipales2012 #adnelecciones	19:53, 2012-10-28
SalvadorSchwartzmann @s_schwartzmann Concepción Primera mesa termina en Concepción 90 votos, de 350 electores (75% abstención) Armstrong 39, Ortriz 36, Córdoba 8 y 7 votos nulos	21:29, 2012-10-28
Mario Waissbluth @mwaissbluth Santiago, Chile Bamoh a tener Arcardeh shuper legitimao por la halta botación recibía, N preocupaos por la edukasión y esas otras cosas como la vasurah	20:22, 2012-10-28
Radio Santa Maria @radiosantamaria Coyhaique, Patagonia Chilena Coyhaique: 4037 votos escrutados Huala 57,7, Muñoz 39,3, Acevedo 3,1%	21:54, 2012-10-28
Christian Pino L. @christianpino Santiago de Chile VALPARAISO: Vía @mauropalma: Edo de la Barra, 23 mesas, Castro 1061 Pinto 1062...un voto!!!!	21:59, 2012-10-28
Cristian Nuñez Fica @hombrederadio Iquique - Chile AHORA: Hay pelea en Luis Cruz Martinez. Es agredido el Camarógrafo de Tarapaca Televisión, por el Diputado Hugo Gutierrez #IQUIQUE	22:00, 2012-10-28

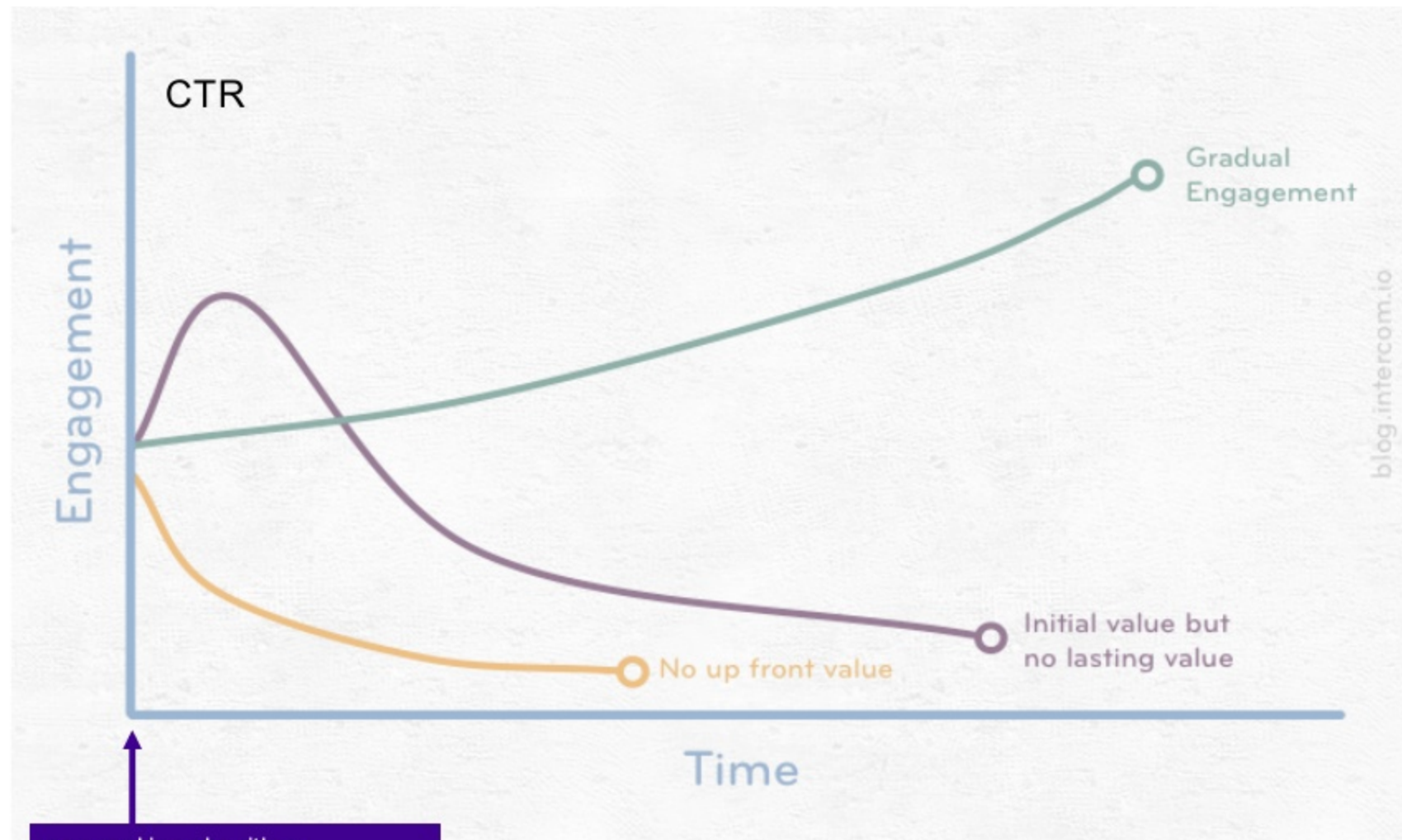
top most popular tweets + geographical diverse

Being from a central or peripheral location makes a difference.

Peripheral users did not perceive the timeline as being diverse

It should never be just about the algorithm, but also how users respond to what the algorithm returns to them

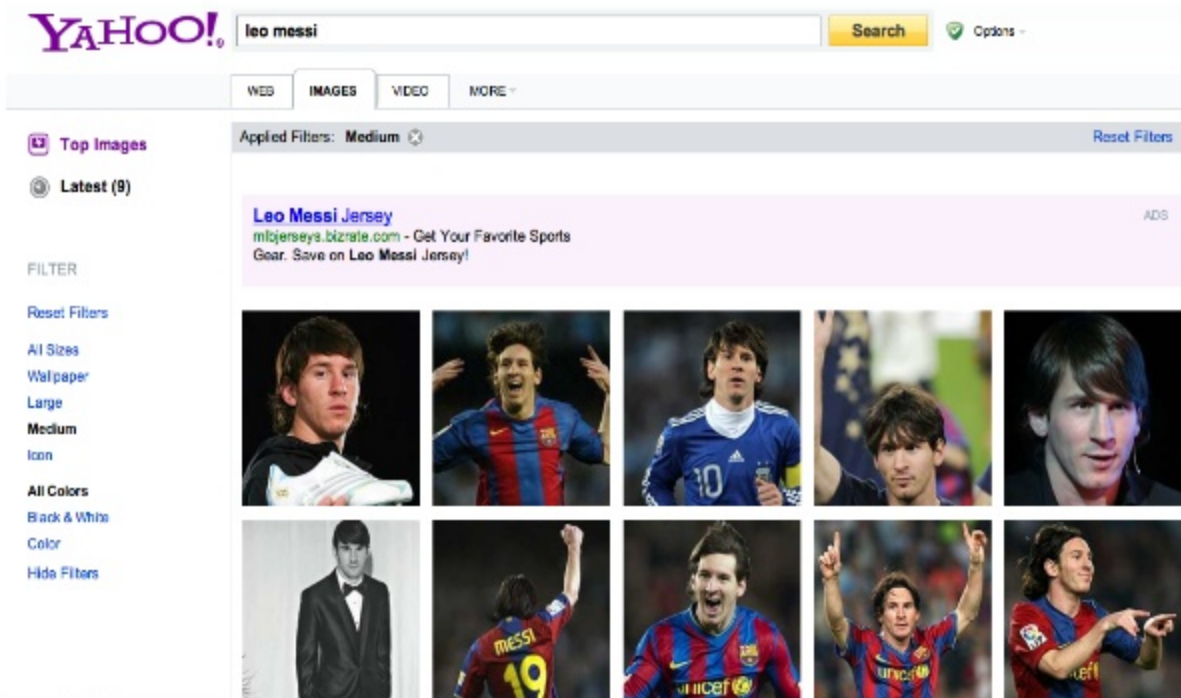
Implicit signal: Click-through rate



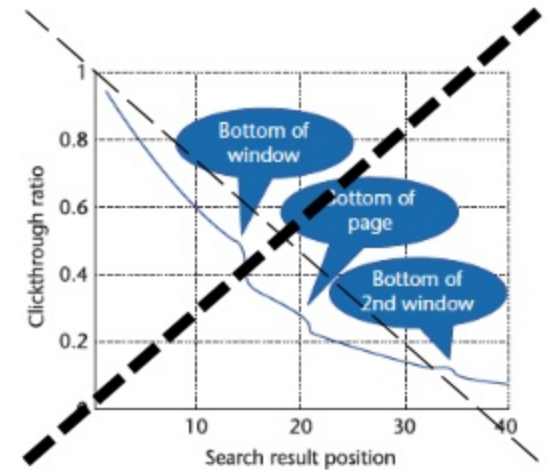
new ranking algorithm
new design of search result page
...

YAHOO!

Signal signal: Clicks (I)



PEOPLE WHO SEARCHED FOR LEO MESSI ALSO SEARCHED FOR



Multimedia search activities often driven by entertainment needs, not by information needs

YAHOO!

Implicit signal: Clicks (II)

Explorative and serendipitous search

YAHOO! barcelona spain

Web
Images
Video
News
Local
Shopping
Maps
More

Anytime
Past day
Past week
Past month

City Guide Hotels Restaurants Things To Do Flights

Barcelona, Spain
travel.yahoo.com
Mon Jul 21 10:35 am CEST Partly Cloudy, 77°F

Originally established by the Romans, Barcelona is a city with a rich cultural history. Today, it serves as the capital of Catalonia and is still one of the busiest port towns in the Mediterranean. In addition to being an economic powerhouse, it is one of the top tourist destinations in Europe, drawing ... [more](#)

Barcelona Spain - Image Results

Related Places

- Sagrada Familia
- Las Ramblas
- La Boqueria
- Casa Milà
- Park Guell
- Camp Nou
- La Pedrera
- Casa Batlló
- Les Corts



YAHOO!

(Miliaraki, Blanco & Lalmas, 2015)

Implicit signal: No click

YAHOO!
UK & IRELAND

Angel Lane surgery Dunmow

Search

3,110 results

WEB IMAGES VIDEO NEWS SHOPPING MORE

Search: ☒ the Web ☐ only in UK ☐ only in Ireland

FILTER BY TIME

- Anytime
- Past day
- Past week
- Past month

[Angel Lane Surgery - Essex](#)
Welcome to **Angel Lane Surgery** . We aim to provide you and your family with the best possible healthcare. **Angel Lane Surgery Angel Lane Great Dunmow Essex CM6 1AQ**
www.angellanesurgery.co.uk - [Cached](#)

[Useful Contacts - Angel Lane Surgery - Essex](#)
Angel Lane Surgery: Appointments/Enquiries: 01371 872 122: Hospitals ... Registrar of Births, Deaths and Marriages (**Dunmow**) ...
www.angellanesurgery.co.uk/useful.asp - [Cached](#)

[Overview - Angel Lane Surgery - NHS Choices](#)
Angel Lane Surgery. Telephone: 01371 872122 Address: **Angel Lane** , Great Dunmow, Dunmow, Essex, CM6 1AQ Website: Website address not added
www.nhs.uk/Services/gp/Overview/DefaultView.aspx?id=6F... - [Cached](#)
[More results from nhs.uk »](#)

I just wanted the phone number ... I am totally happy 😊

Implicit signal: No click



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

Clickthrough rate:
% of clicks when URL shown (per query)

Hover rate:
% hover over URL (per query)

Unclicked hover:
Median time user hovers over URL but no click (per query)

Max hover time:
Maximum time user hovers over a result (per SERP)