

YAHOO!

Impact of Response Latency on User Behavior in Web Search

Ioannis Arapakis, Xiao Bai, B. Barla Cambazoglu

Yahoo Labs, Barcelona

Background Information

- The core research in IR has been on improving the **quality** of search results with the eventual goal of satisfying the **information needs** of users
- This often requires sophisticated and costly solutions
  - more information stored in the inverted index
  - machine-learned ranking strategies
  - fusing results from multiple resources

Trade-off between the **speed** of a search system and the **quality** of its results

Too **slow** or too **fast** may result in financial consequences for the search engine

User Side

- Web users
  - are impatient
  - have limited time
  - expect sub-second response times
- High response latency
  - can distract users
  - results in fewer query submissions
  - decreases user engagement over time

Search Engine Side

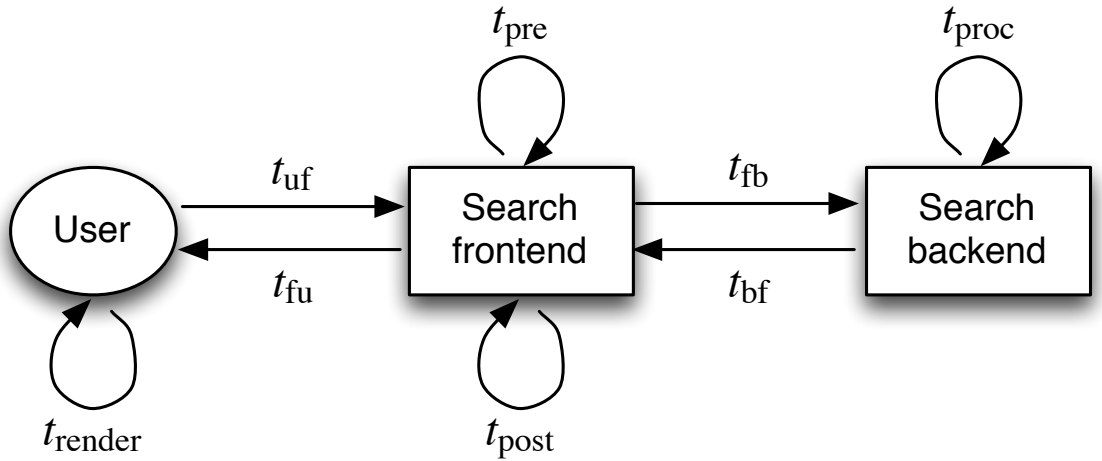
- Search engines
  - have large user bases and query volumes
  - make heavy investments on H/W infrastructure
  - try to maintain query response times at reasonable levels
- These investments
  - incur a financial burden on search engine companies
  - result in financial losses

Research Questions

1. What are the main components in the response latency of a search engine?
2. How sensitive are users to response latency?
3. How much does response latency affect user behavior?

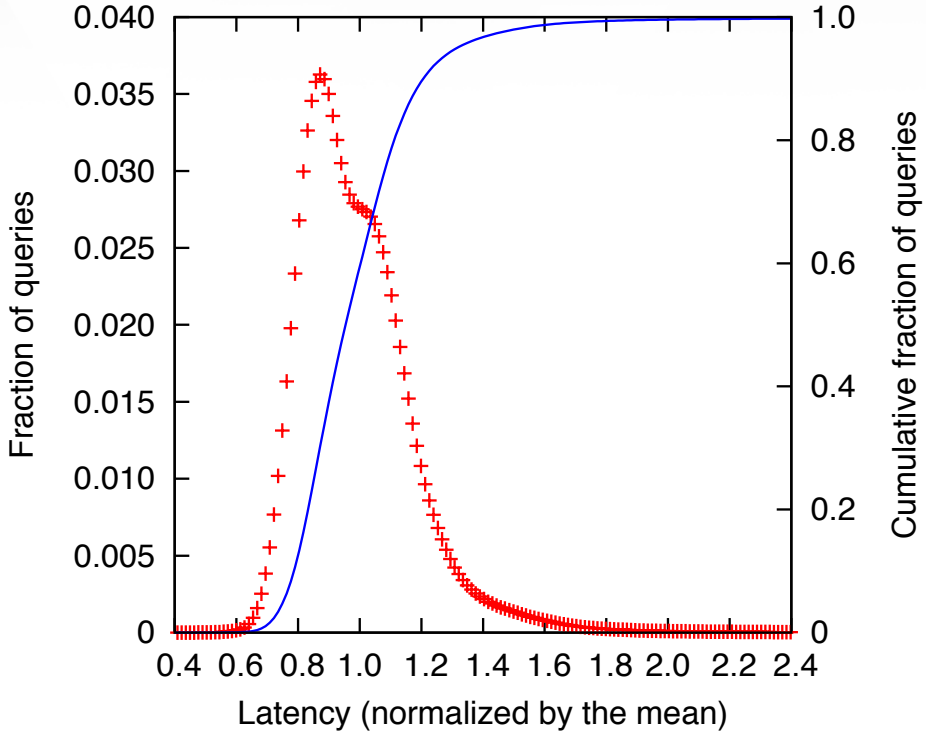
- We conduct two studies
  - a small-scale user study
  - a large-scale query log analysis

Components of User-Perceived Response Latency



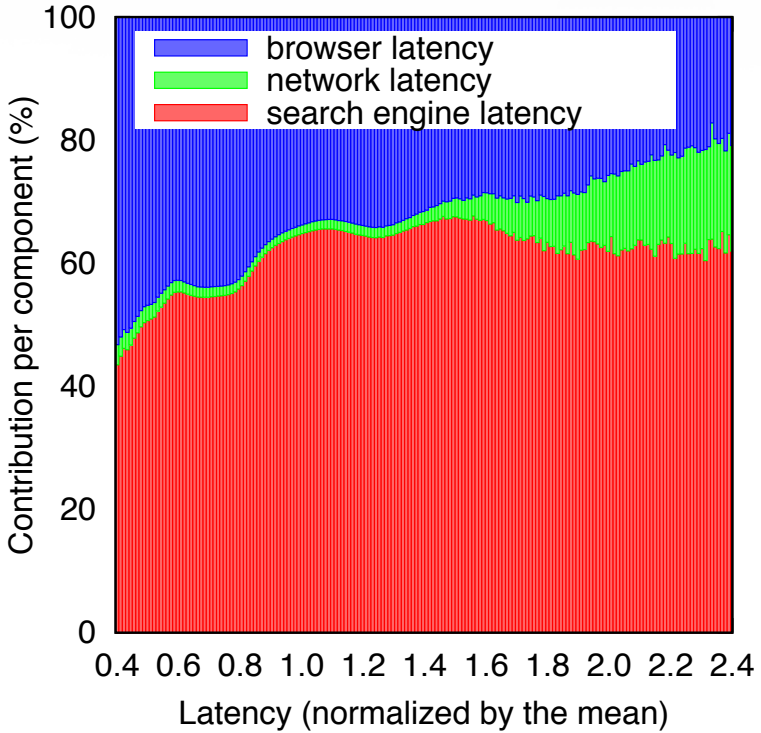
- **network latency**: $t_{uf} + t_{fu}$
- **search engine latency**: $t_{pre} + t_{fb} + t_{proc} + t_{bf} + t_{post}$
- **browser latency**: t_{render}

Distribution of Latency Values



YAHOO!

Contribution of Latency Components



YAHOO!

Study 1:
User Sensitivity to Latency

Experimental Method (Task 1 & 2)

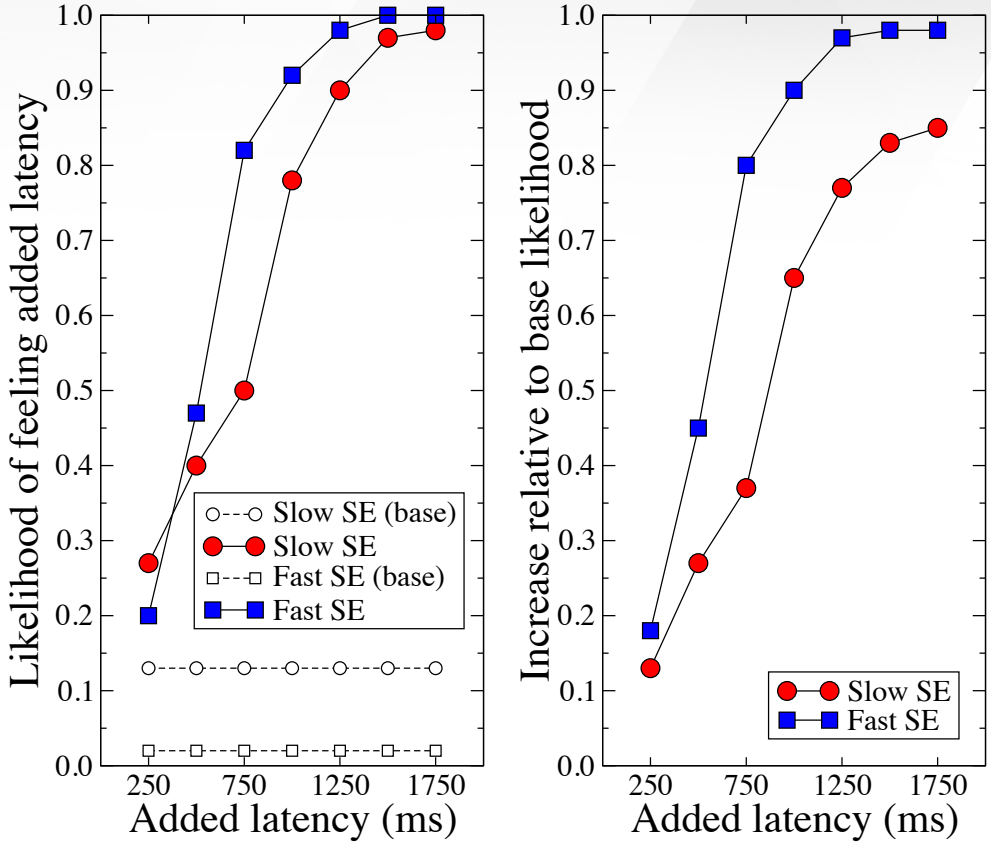
- Two independent variables
  - Search latency (0 – 2750ms)
  - Search site speed (slow, fast)
- 12 participants (female=6, male=6)
  - Studying (33.3%)
  - Studying while working (54.3%)
  - Full-time employees (16.6%)

Task 1: Procedure

- Participants submitted 40 navigational queries
- After submitting each query, they were asked to report if the response of the search site was “slow” or “normal”
- For each query we **increased** latency by a fixed amount (0 – 1750ms), using a step of 250ms
- Each latency value (e.g., 0, 250, 500) was introduced 5 times, in a random order

Task 1: Results

- Delays <500ms are not easily noticeable
- Delays >1000ms are noticed with high likelihood



YAHOO!

Task 2: Procedure

- Participant submitted 50 navigational queries
- After each query submission they provided an estimation of the search latency in milliseconds
- For each query we **increased** latency by a fixed amount (500 – 2750ms), using a step of 250ms
- Each latency value (e.g., 0, 250, 500) was introduced 5 times, in a random order
- A number of training queries was submitted without any added delay

Task 2: Results

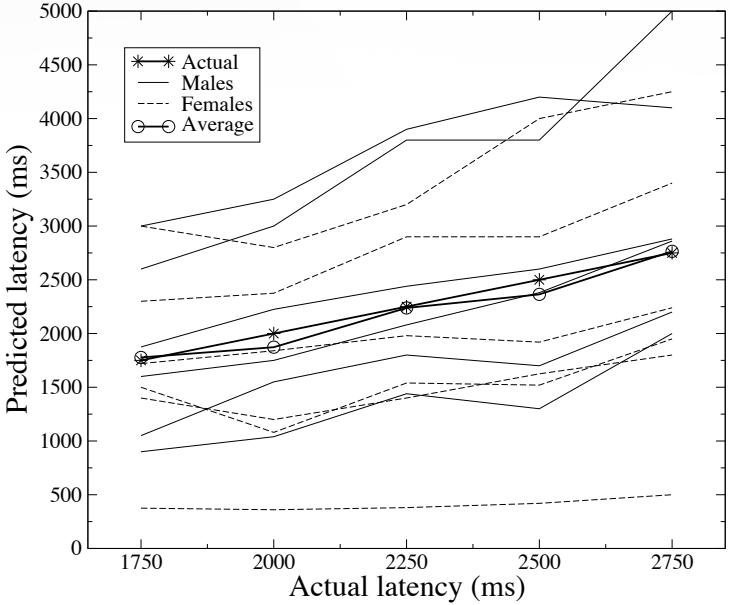


Fig. 1: Slow search engine

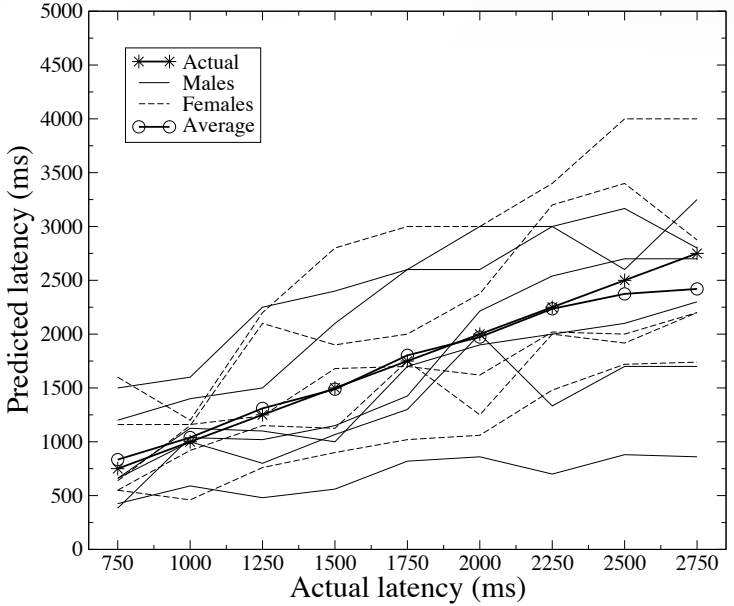


Fig. 2: Fast search engine

YAHOO!

Study 2:
Impact of Latency on Search Experience

Experimental Design

- Investigate the effects of response latency on the user engagement and satisfaction
- Two independent variables
  - Search latency (0, 750, 1250, 1750)
  - Search site speed (slow, fast)
- Search latency was set to desired amount using a custom-made javascript deployed through the Greasemonkey extension
- 20 participants (female=10, male=10)

Procedure

- Participants performed four search tasks
  - Evaluate the performance of four different backend search systems
  - Submit as many navigational queries from a list of 200 randomly sampled web domains
  - For each query they were asked to locate the target URL among the first ten results of the SERP
- Training queries were used to allow participants to familiarize themselves with the “default” search site speed

Questionnaires

- User Engagement Scale (UES)
  - Positive affect (PAS)
  - Negative affect (NAS)
  - Perceived usability
  - Felt involvement and focused attention
- IBM's Computer System Usability Questionnaire (CSUQ)
  - System usefulness (SYSUSE)
- Custom statements
  - Custom-1: "This search site was fast in responding to my queries"
  - Custom-2: "This search site helped me accomplish my task in a reasonable amount of time"
  - Custom-3: "I feel satisfied with the retrieved results"

Descriptive Statistics (M) for UE and SYSUSE

	SE_{slow} latency				SE_{fast} latency			
	0ms	750ms	1250ms	1750ms	0ms	750ms	1250ms	1750ms
Post-Task Positive Affect	16.20	14.50	15.50	15.20	20.50	19.00	20.80	19.30
Post-Task Negative Affect	7.00	6.80	7.60	6.90	6.80	7.40	7.40	7.20
Frustration	3.20	3.10	2.90	3.30	2.80	3.00	3.50	2.60
Focused Attention	22.80	22.90	19.90	22.20	27.90	26.60	23.90	29.50
SYSUS	32.80	28.90	29.80	27.90	35.20	31.30	29.80	33.20

- Positive **bias** towards SE_{fast}
- SE_{fast} participants were more **deeply engaged**
- SE_{fast} participants' usability perception was more **tolerant** to delays

Correlation Analysis of Beliefs and Reported Scales

Beliefs	postPAS	postNAS	FA	CSUQ-SYSUS	custom-1	custom-2	custom-3
SE_{slow} will respond fast to my queries	**.455****	.041	**0.702****	.267	.177	.177	.082
SE_{slow} will provide relevant results	.262	-.083	**.720****	**.411****	.278	.263	.232
SE_{fast} will respond fast to my queries	-.051**	.245	**.341***	**.591****	**.330***	**.443****	**.624****
SE_{fast} will provide relevant results	-.272	.133	-.133	**.378***	.212	.259	**.390***

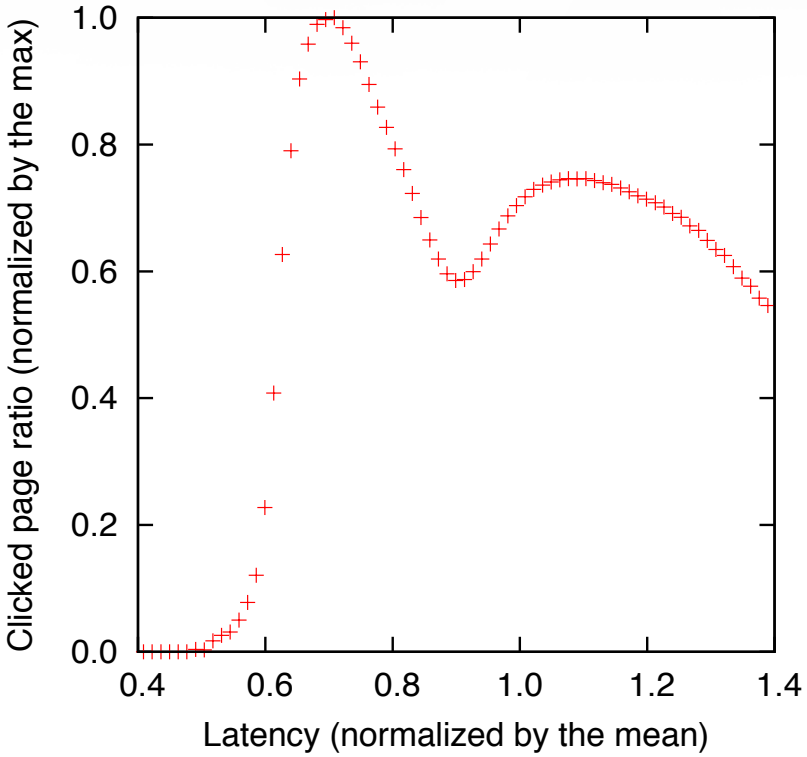
*. Correlation is significant at the .05 level (2-tailed). **. Correlation is significant at the .01 level (2-tailed)

Query Log Analysis

Query Log and Engagement Metric

- Random sample of 30m web search queries obtained from Yahoo
- We use the end-to-end (user perceived) latency values
- To control for differences due to geolocation or device, we select queries issued:
  - Within the US
  - To a particular search data center
  - from desktop computers
- We quantify engagement using the clicked page ratio metric

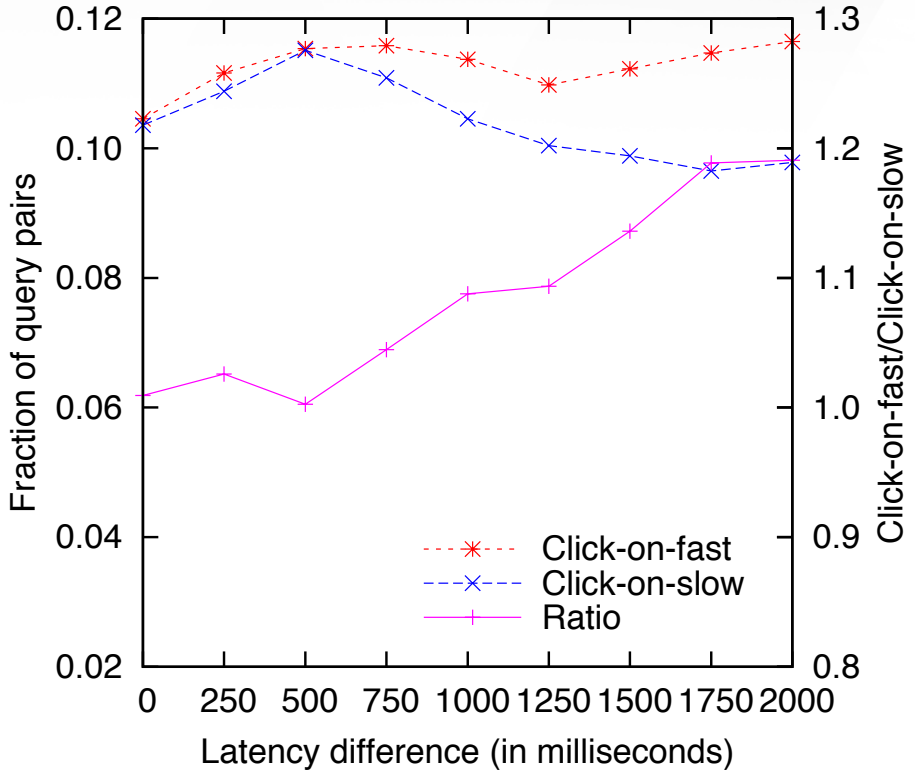
Variation of Clicked Page Ratio Metric



YAHOO!

Eliminating the Effect of Content

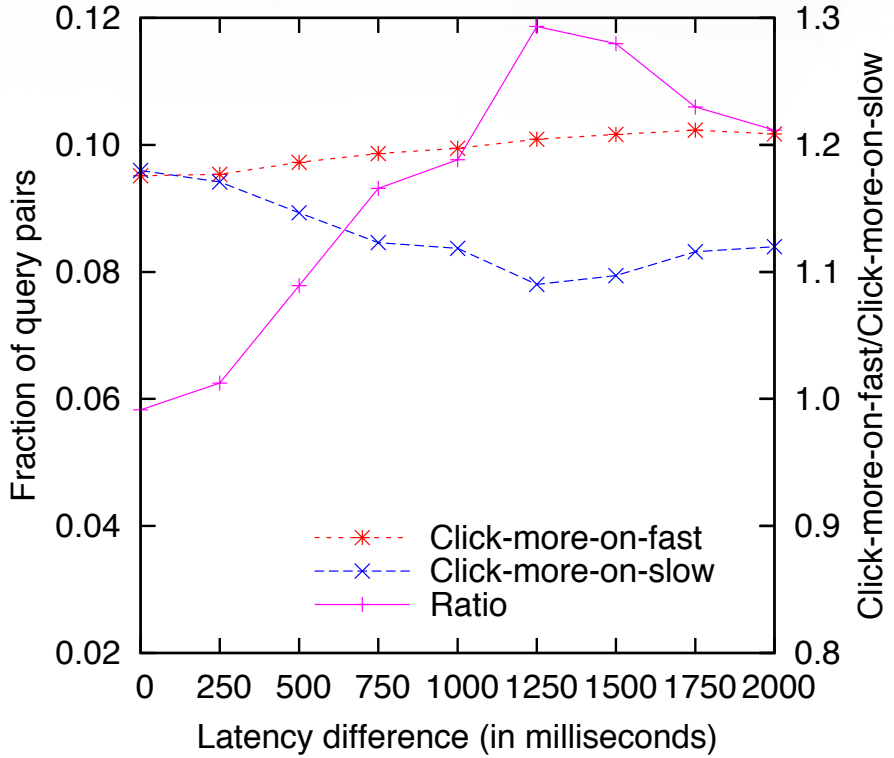
- 500ms of latency difference is the **critical point** beyond which users are more likely to click on a result retrieved with lower latency



YAHOO!

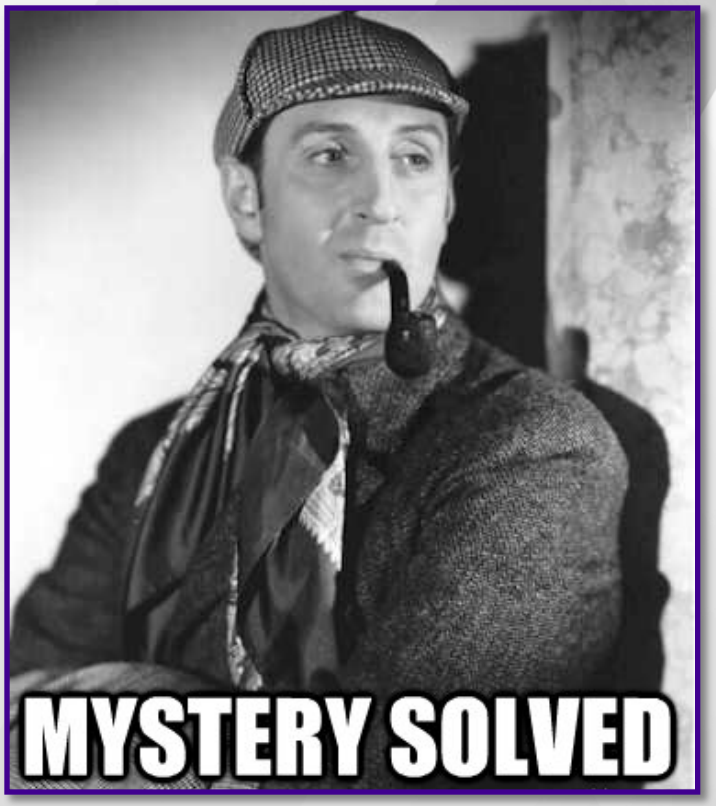
Eliminating the Effect of Content

- Clicking on more results becomes **preferable** to submitting new queries when the latency difference exceeds a certain threshold (1250ms)



YAHOO!

Conclusions



Conclusions

- Up to a point (500ms) added response time delays **are not noticeable** by the users
- After a certain threshold (1000ms) the users **can feel** the added delay with very high likelihood
- Perception of search latency varies considerably across the population!
- The tendency to **overestimate** or **underestimate** system performance **biases** users' interpretations of search interactions and **system usability**

Conclusions

- Given two content-wise identical result pages, users are more likely to click on the result page that is served with **lower latency**
- 500ms of latency difference is the **critical point** beyond which users are more likely to click on a result retrieved with lower latency
- Clicking on more results becomes **preferable** to submitting new queries when the latency difference exceeds a certain threshold (1250ms)

Thank you for your attention!

arapakis@yahoo-inc.com

iarapakis

http://www.slideshare.net/iarapakis/sigir2014