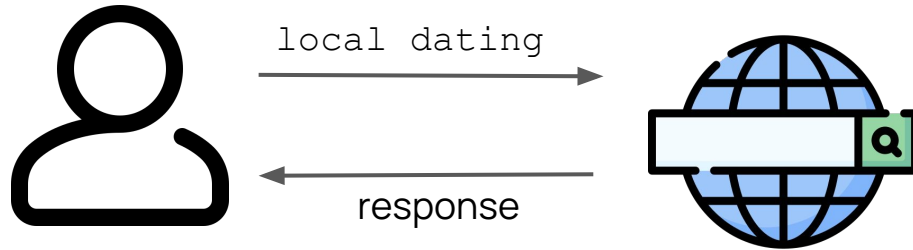




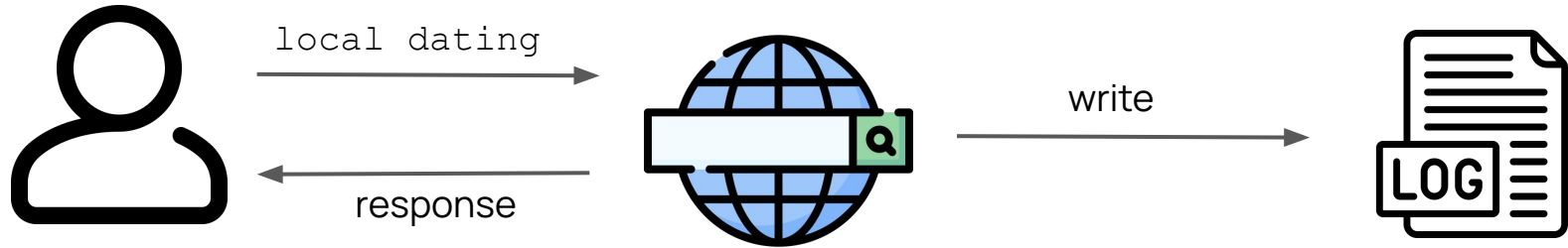
The Performance of Human Query Obfuscation - A Gamified Approach

Bachelor's Thesis Defence by Nicola Lea Libera on 26.11.2021
Bauhaus-Universität Weimar

Search Engines - A Threat to Your Privacy?



Search Engines - A Threat to Your Privacy?



Search Engines - A Thread to Your Privacy?

Google is giving data to police based on search keywords, court docs show



Search Engines - A Thread to Your Privacy?

Google is giving data to police based on search keywords, court docs show



“A face is exposed for AOL searcher no. 4417749”

From New York Times, August 9, 2006

In August 2006, Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

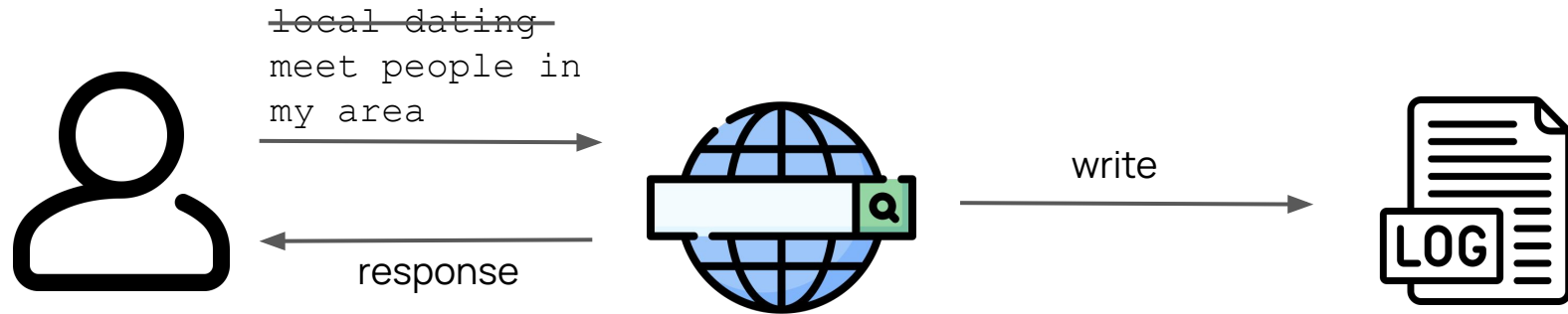
No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn", and several people with the last name Arnold.

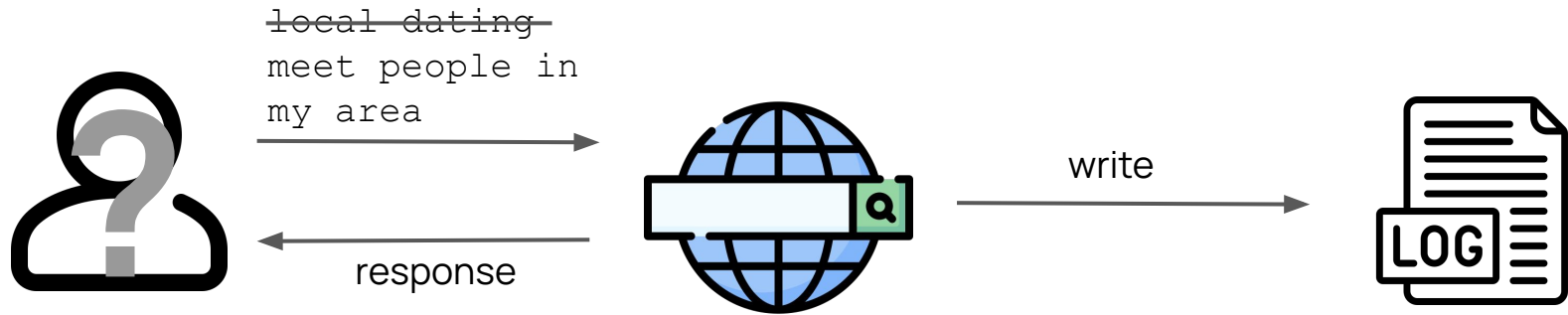
It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, GA



Query Obfuscation



Query Obfuscation



Research Question

Are humans able to effectively obfuscate private information needs?





A Query Obfuscation Game

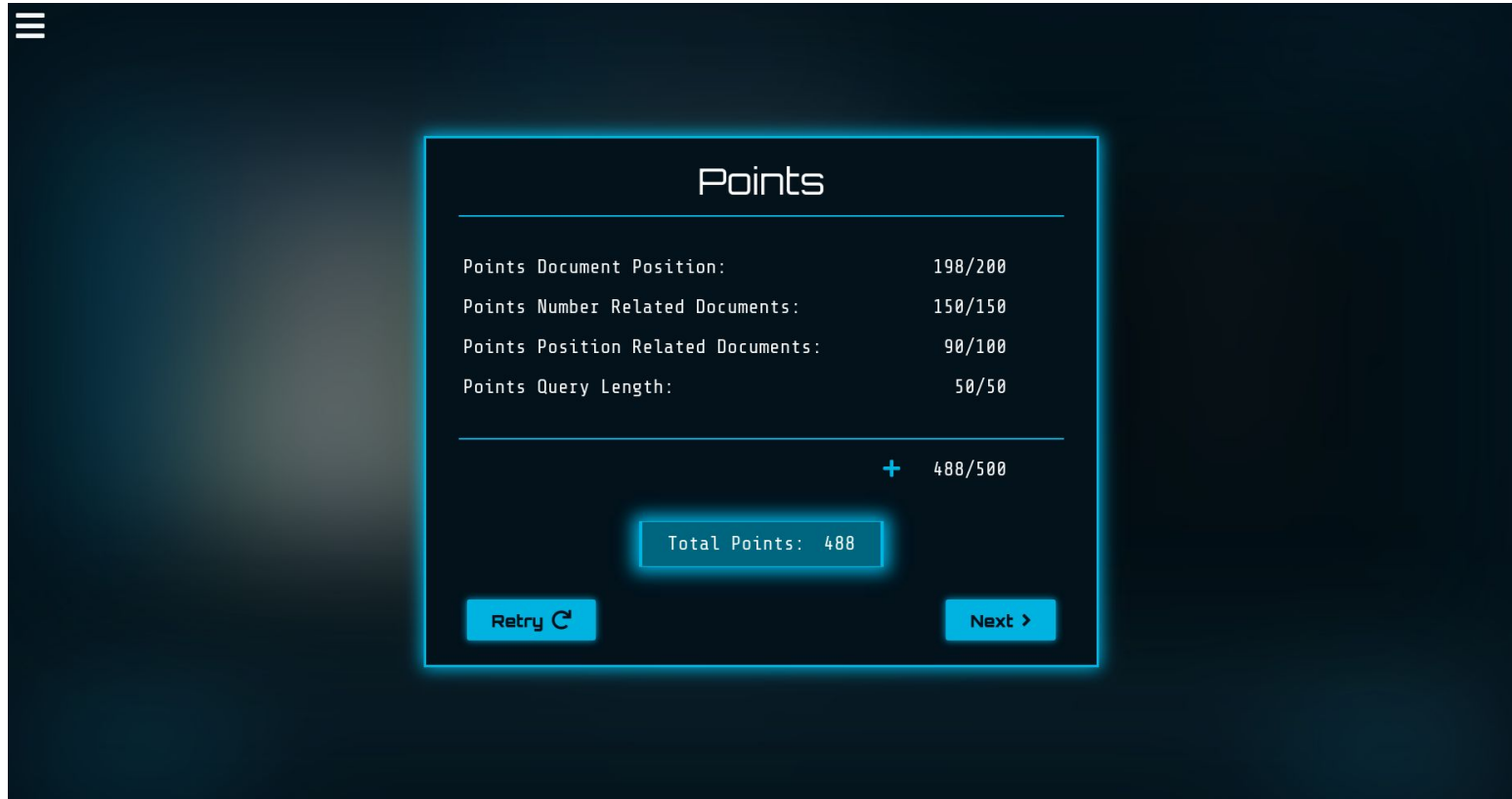
Game Interface



Game Interface



Game Interface

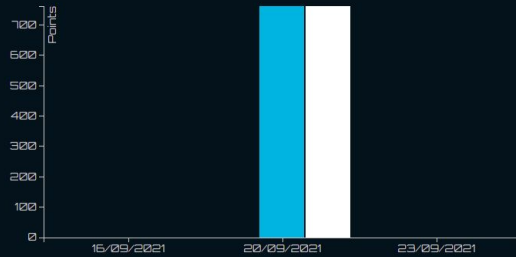


Game Interface

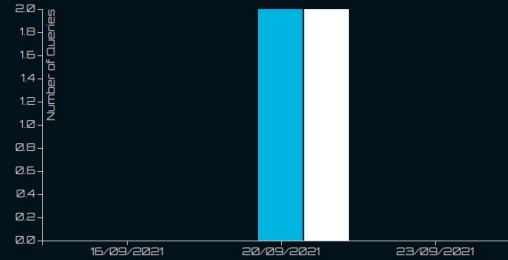
Statistics

Weekly Statistic

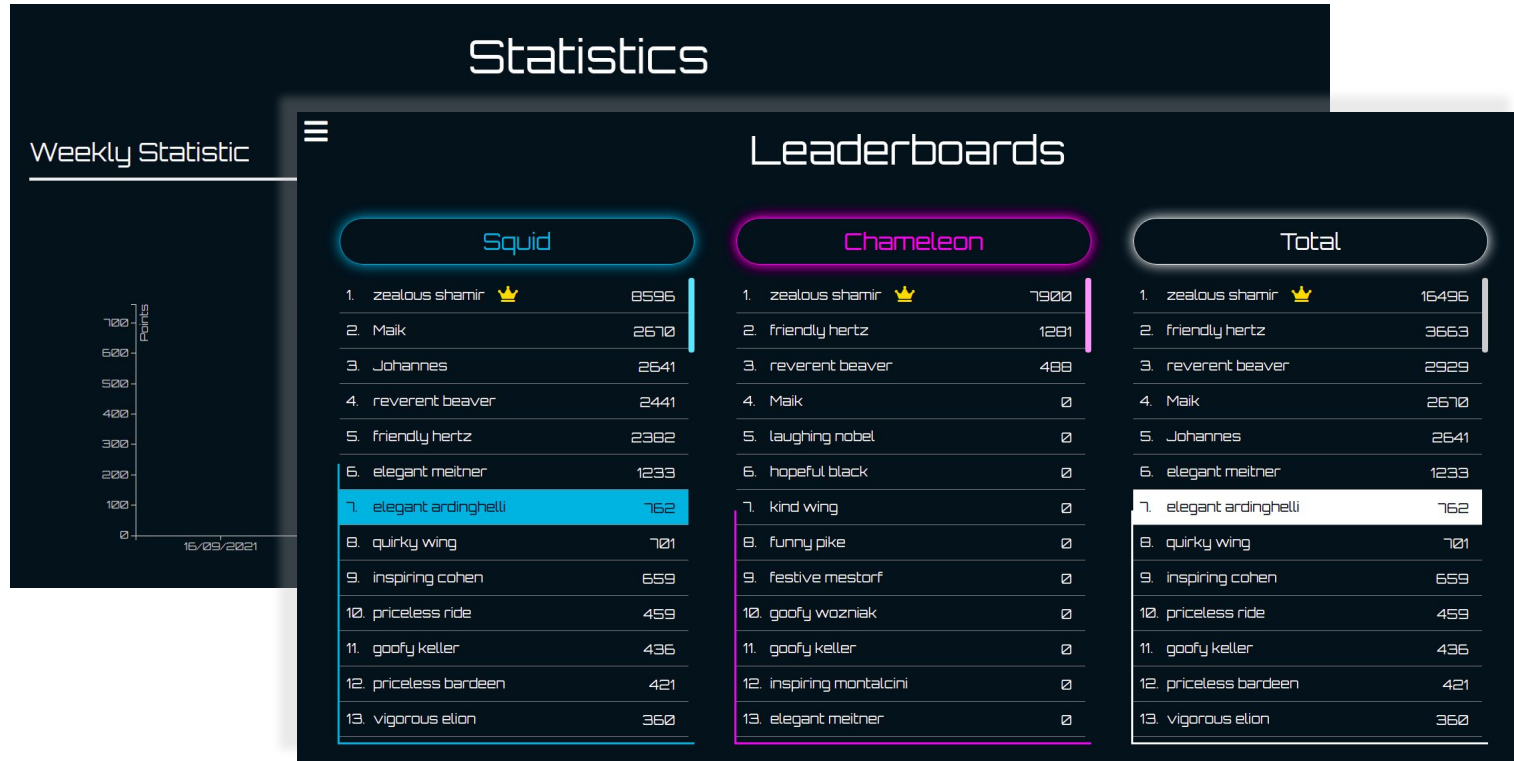
Points



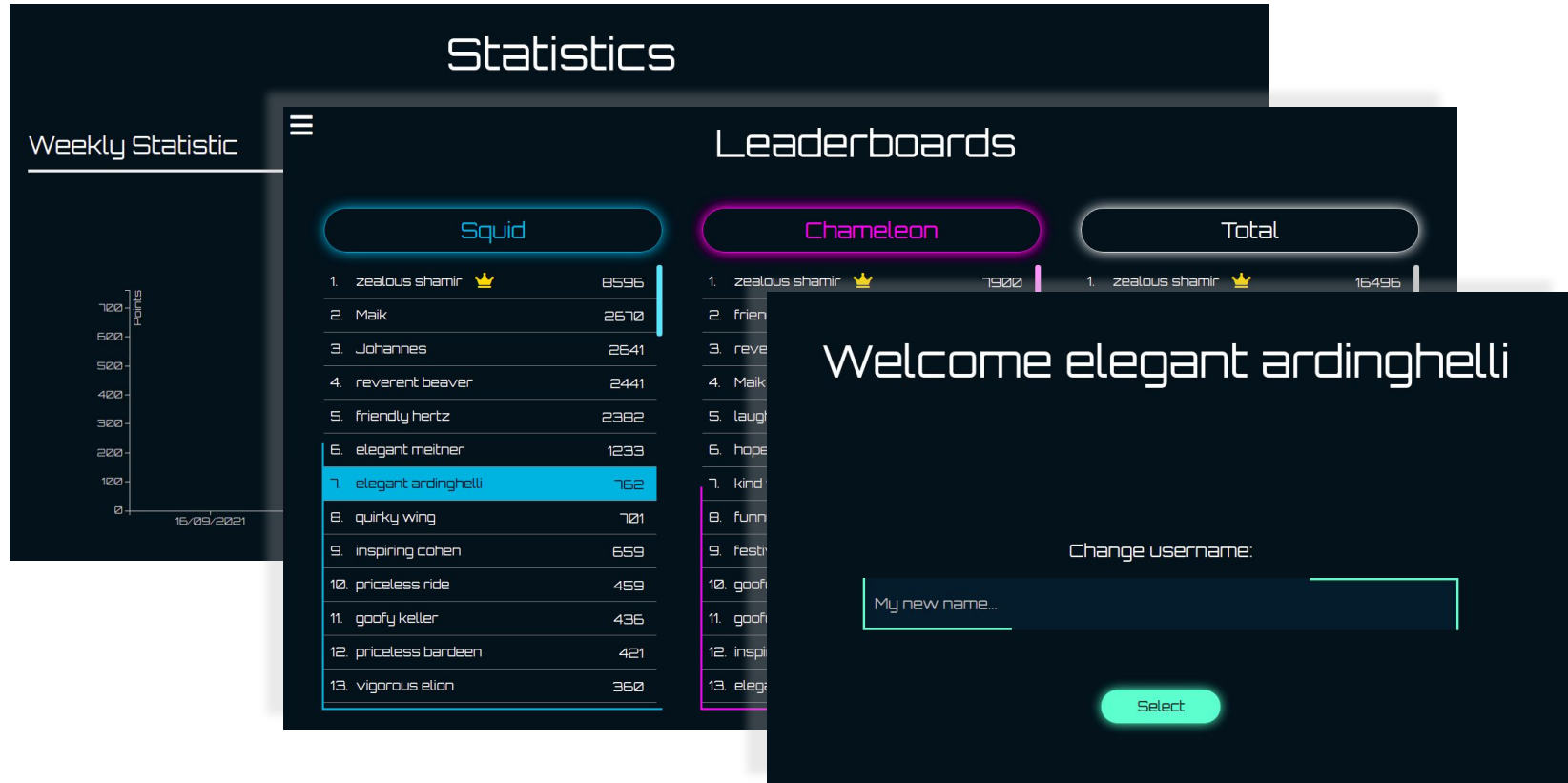
Number of obfuscated Queries



Game Interface



Game Interface



Implementation


- 200 sensitive information needs in six categories

Implementation

- 200 sensitive information needs in six categories
 - Selection and rendering of relevant page from ClueWeb
 - Corpus: ClueWeb as in previous work [Arampatzis, Inf. Retr. J.'15]
 - Search: ChatNoir

Implementation

- 200 sensitive information needs in six categories
 - Selection and rendering of relevant page from ClueWeb
 - Corpus: ClueWeb as in previous work [Arampatzis, Inf. Retr. J.'15]
 - Search: ChatNoir
 - Goal:

Category	Sensitive Query	Relevant Web Page
Personal	local dating	

Visual Enriching of Sensitive Documents



Match2be Logo

Choose a Username:

I'm a:
 seeking a

Your City:

Your State:

Email:

Note: Please use valid email

Confirm Email:

Password:

Must be at least 5 characters

Confirm Password:

Submit

Visual Enriching of Sensitive Documents



Match2be Logo

Choose a Username:

I'm a:
Man
seeking a
Woman

Your City:

Your State:

Email:

Note: Please use valid email

Confirm Email:

Password:

Must be at least 5 characters

Confirm Password:

Match²be
Social Dating Network

Get Your Free Dating Profile

Choose a Username:

I'm a: Man seeking a Woman

Your City:

Your State:

Email:

Note: Please use valid email

Confirm Email:

Password:

Must be at least 5 characters

Confirm Password:

[Join](#) | [Contact Us](#) | [Site Tour](#) | [News](#) | [Dating Advice](#) | [Affiliates](#) | [Advertise with Us](#) | [Jobs](#) | [Dating Coaches](#) | [Privacy Statement](#) | [Terms of Use](#)
2011 Match2be.com All rights reserved.

Visual Enriching of Sensitive Documents



+

INTERNET ARCHIVE
WayBackMachine

INTERNET ARCHIVE
WayBackMachine

Match²be
Social Dating Network

Get Your Free Dating Profile

Choose a Username:

I'm a: **Man** seeking a **Woman**

Your City:

Your State:

Email:
Note: Please use valid email

Confirm Email:

Password:
Must be at least 5 characters

Confirm Password:

[Join](#) | [Contact us](#) | [Site Tour](#) | [News](#) | [Dating Advice](#) | [Affiliates](#) | [Advertise with Us](#) | [Jobs](#) | [Dating Coaches](#) | [Privacy Statement](#) | [Terms of Use](#)
2011 Match2be.com All rights reserved.

Match²be
Social Dating Network

Get Your Free Dating Profile

Choose a Username:

I'm a: **Man** seeking a **Woman**

Your City:

Your State:

Email:
Note: Please use valid email

Confirm Email:

Password:
Must be at least 5 characters

Confirm Password:

[Join](#) | [Contact us](#) | [Site Tour](#) | [News](#) | [Dating Advice](#) | [Affiliates](#) | [Advertise with Us](#) | [Jobs](#) | [Dating Coaches](#) | [Privacy Statement](#) | [Terms of Use](#)
2011 Match2be.com All rights reserved.

Logging and Preprocessing Data

- Logging to save obfuscated queries
- Preprocessing collected data on obfuscated queries
 - Position of target document,
 - Time to formulate query,
 - Used keywords,
 - ...

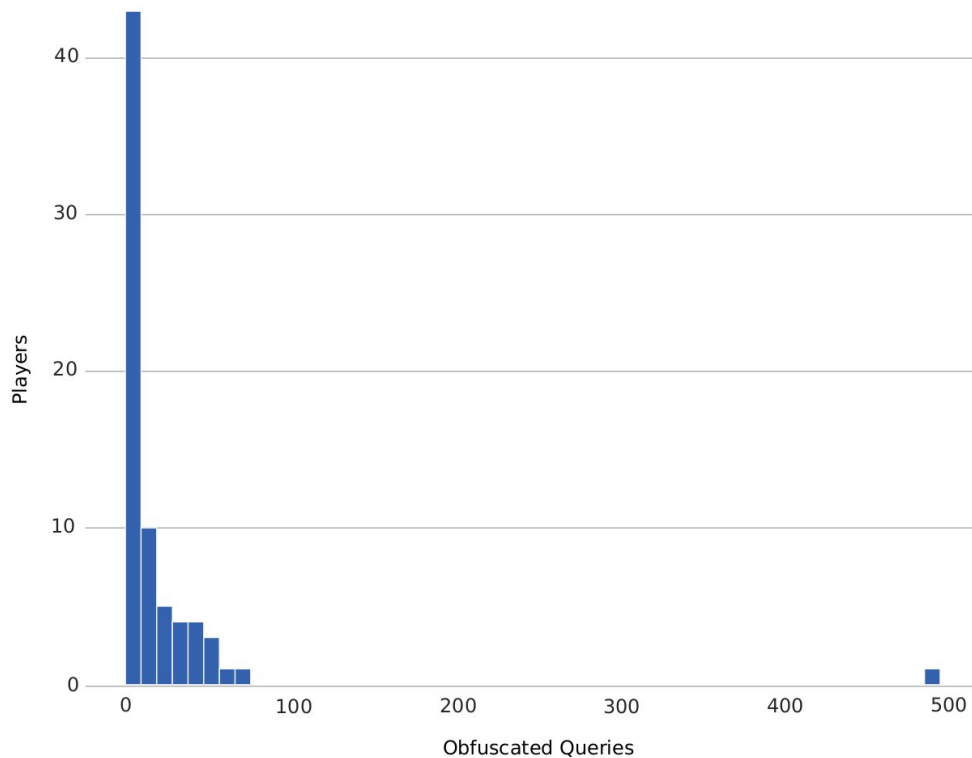


A close-up photograph of a person's hand typing on a computer keyboard. The keyboard has a dark, possibly black or dark blue, keycap set with vibrant blue and red backlighting. The scene is dimly lit, with the primary light source being the keyboard's illumination, which casts a cool blue glow over the hand and the surrounding area. The background is out of focus, showing more of the keyboard and some indistinct light sources. The word "Evaluation" is superimposed in the center of the image in a clean, white, sans-serif typeface. The text is framed by a thin white line that forms a partial rectangle, with the top and right sides being more prominent than the bottom and left sides.

Evaluation

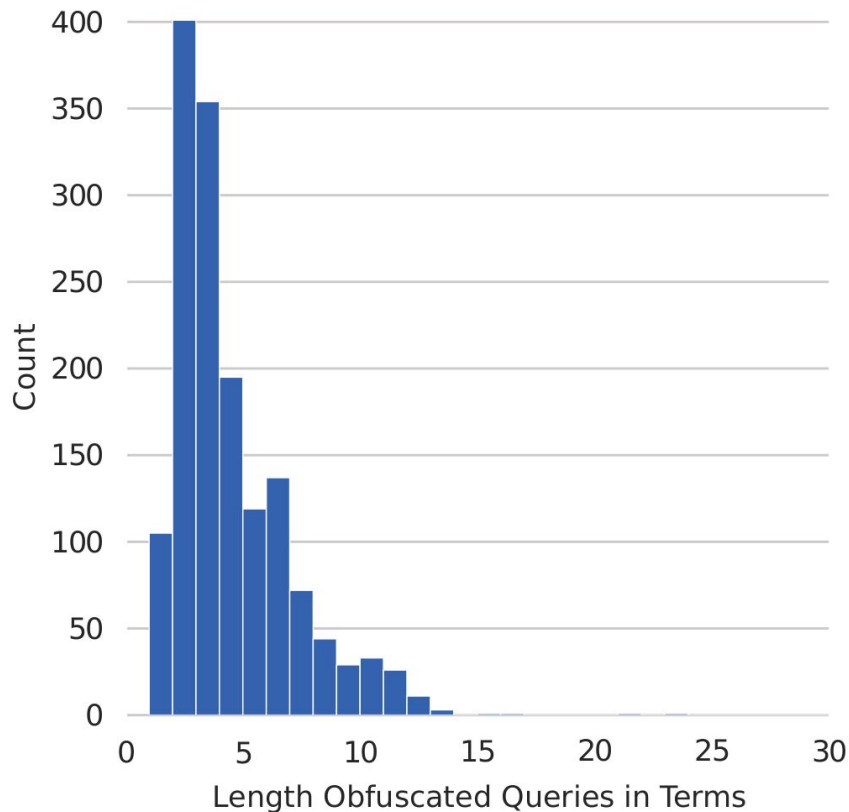
Players

- Thesis:
 - Players: 72
 - Age: 19-64
 - Queries: 1.534
- Current status:
 - Players: 89



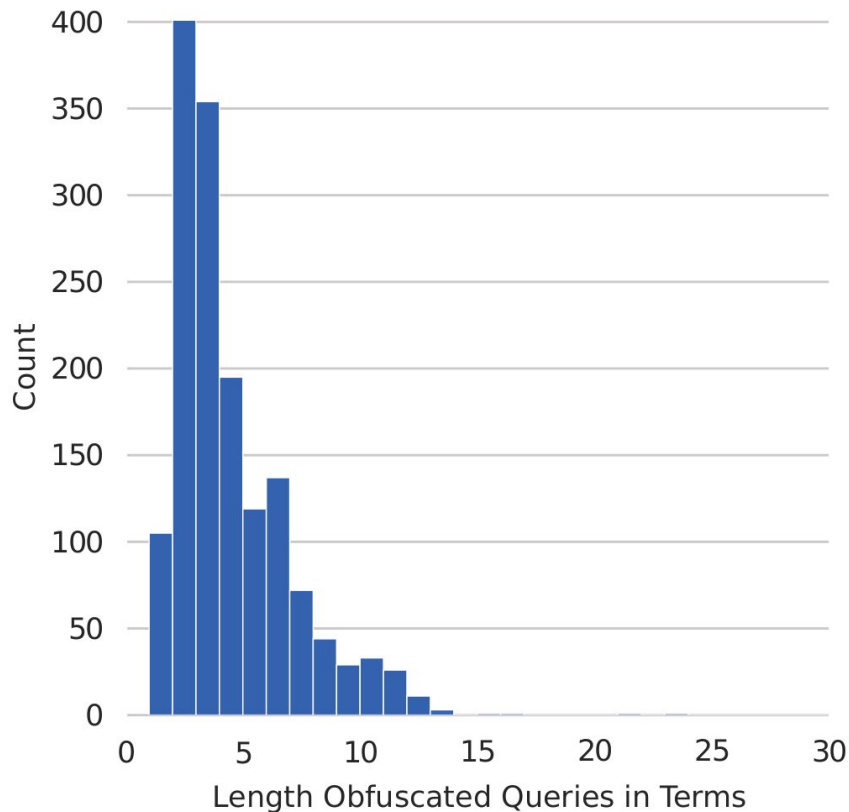
Query Length

- Average length of obfuscated queries: 4.04 terms
- Normal range: 2 - 4 terms [Arampatzis, SIGIR'08]
- Normal average length: 3.31 - 3.5 terms [Azzopardi, CIKM'11, Chowdhury, ICSIS'06]



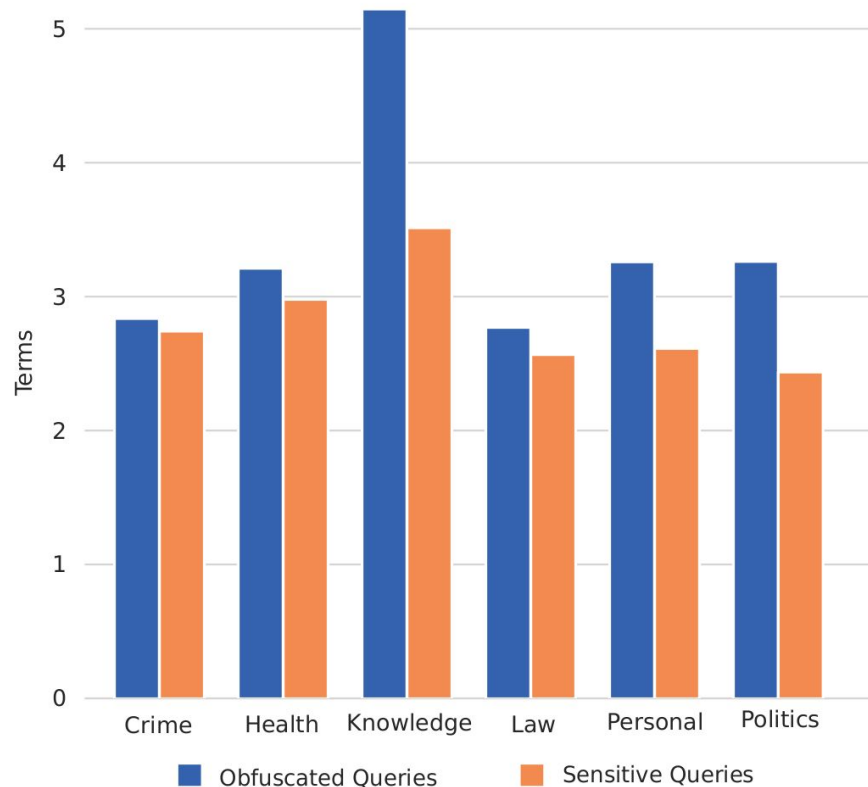
Query Length

- Average length of obfuscated queries: 4.04 terms
 - Normal range: 2 - 4 terms [Arampatzis, SIGIR'08]
 - Normal average length: 3.31 - 3.5 terms [Azzopardi, CIKM'11, Chowdhury, ICSIS'06]
- Obfuscated queries slightly longer than normal queries



Query Length

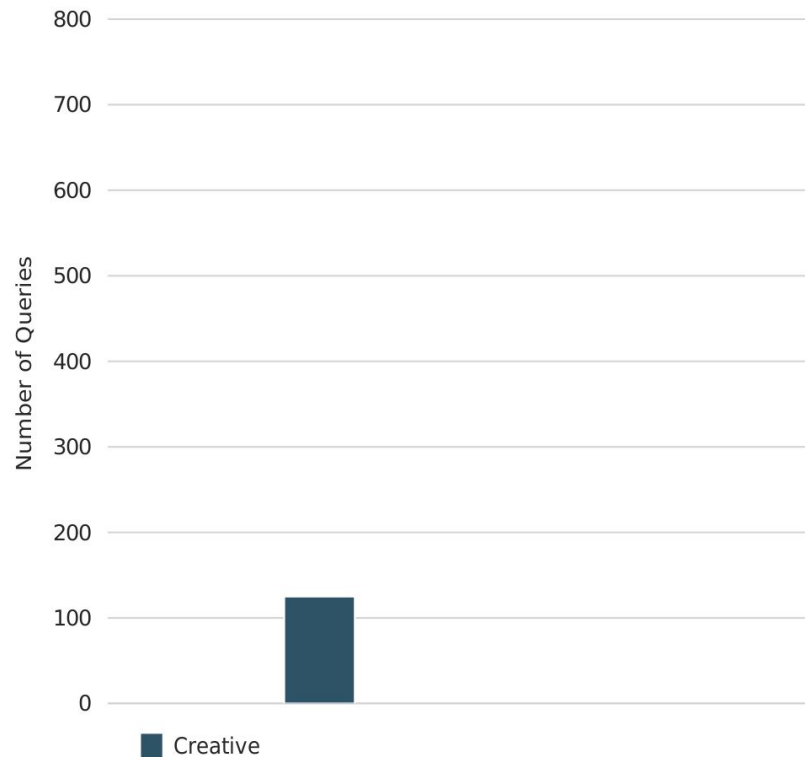
- Average length of obfuscated queries: 4.04 terms
 - Normal range: 2 - 4 terms [Arampatzis, SIGIR'08]
 - Normal average length: 3.31 - 3.5 terms [Azzopardi, CIKM'11, Chowdhury, ICSIS'06]
- Obfuscated queries slightly longer than normal queries



Five Query Types

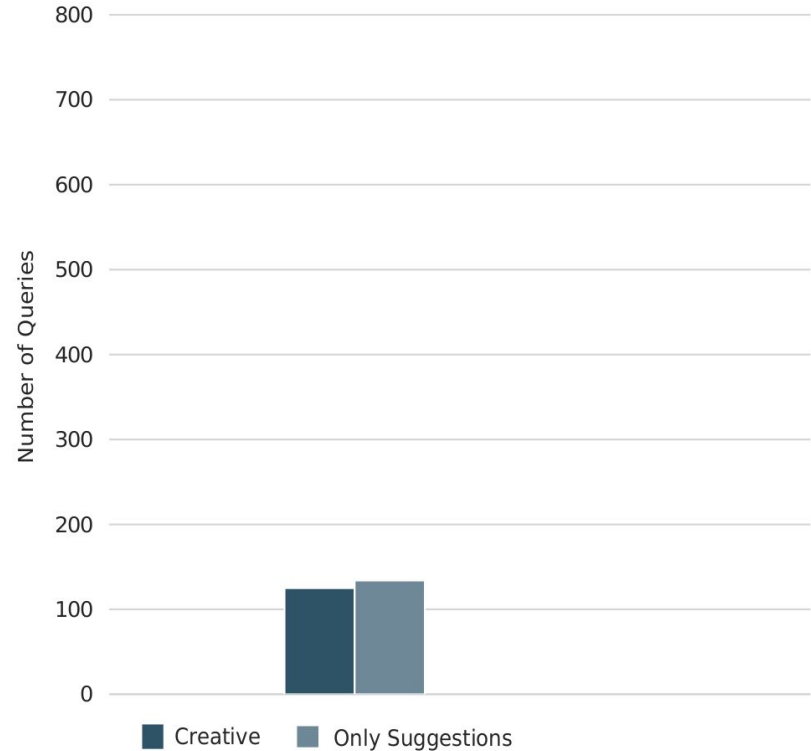
Five Query Types

- Creative
→ neither keyword from provided list nor the web page



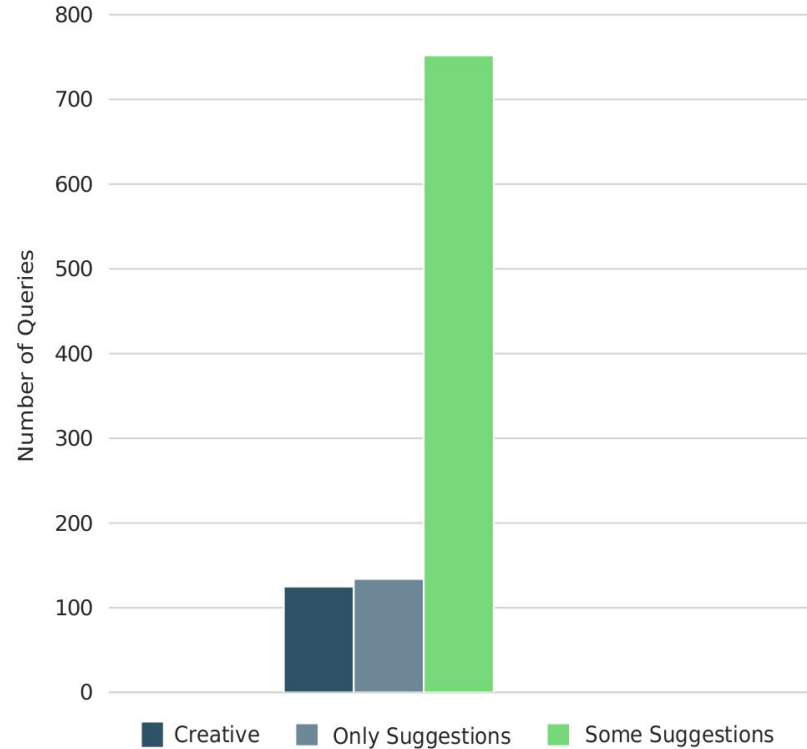
Five Query Types

- Creative
→ neither keyword from provided list nor the web page
- Only Suggestions
→ all terms from keyword list



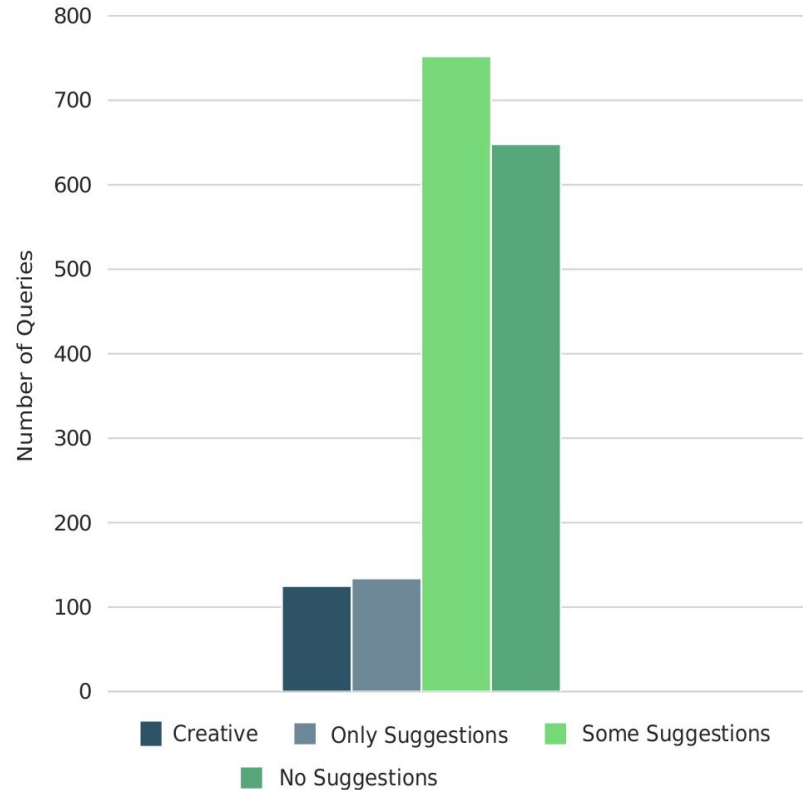
Five Query Types

- Creative
→ neither keyword from provided list nor the web page
- Only Suggestions
→ all terms from keyword list
- Some Suggestions
→ at least one term from keyword list and one other



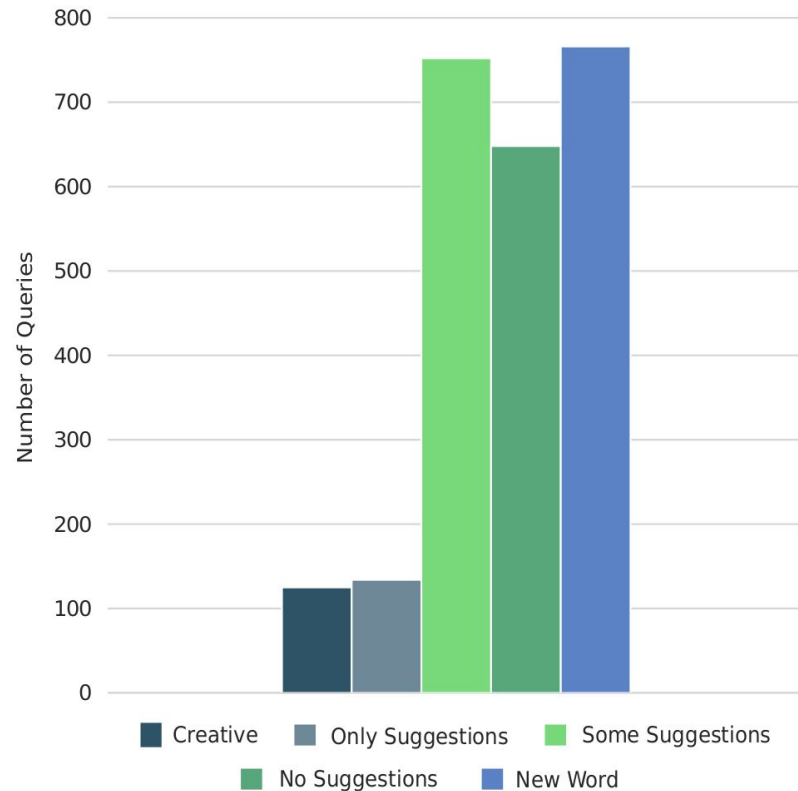
Five Query Types

- Creative
→ neither keyword from provided list nor the web page
- Only Suggestions
→ all terms from keyword list
- Some Suggestions
→ at least one term from keyword list and one other
- No Suggestions
→ not one word from keyword list



Five Query Types

- Creative
→ neither keyword from provided list nor the web page
- Only Suggestions
→ all terms from keyword list
- Some Suggestions
→ at least one term from keyword list and one other
- No Suggestions
→ not one word from keyword list
- New Word
→ at least one word not from the web page



Efficiency & Effectiveness

Driven by	Query Type	Time
Suggestions	All Suggestions	40.50s
	Some Suggestions	42.45s
Creativity	No Suggestions	44.39s
	New Word	46.27s

Efficiency & Effectiveness

			Our Sensitive Queries
Driven by	Query Type	Time	Mean Reciprocal Rank
Suggestions	All Suggestions	40.50s	0.093
	Some Suggestions	42.45s	0.046
Creativity	No Suggestions	44.39s	0.029
	New Word	46.27s	0.002

Efficiency & Effectiveness

Driven by	Query Type	Time	Our Sensitive Queries	
			Mean Reciprocal Rank	Documents Original Query*
Suggestions	All Suggestions	40.50s	0.093	5.223
	Some Suggestions	42.45s	0.046	4.667
Creativity	No Suggestions	44.39s	0.029	2.935
	New Word	46.27s	0.002	1.517

*Number of documents one would also retrieve with the original query

Efficiency & Effectiveness

Driven by	Query Type	Time	Our Sensitive Queries	
			Mean Reciprocal Rank	Documents Original Query*
Suggestions	All Suggestions	40.50s	0.093	5.223
	Some Suggestions	42.45s	0.046	4.667
Creativity	No Suggestions	44.39s	0.029	2.935
	New Word	46.27s	0.002	1.517
—	Automatic	—	0.088	9.229

*Number of documents one would also retrieve with the original query

Efficiency & Effectiveness

Driven by	Query Type	Time	Sensitive Web Track Queries	
			Mean Reciprocal Rank	Relevant Documents
Suggestions	All Suggestions	40.50s	0.010	3.094
	Some Suggestions	42.45s	0.013	3.632
Creativity	No Suggestions	44.39s	0.015	1.783
	New Word	46.27s	0.002	1.235
—	Automatic	—	0.014	2.872

Conclusion

People are not able to effectively obfuscate sensitive information needs on their own.



A close-up, artistic photograph of a hand typing on a computer keyboard. The keyboard has a dark, possibly black or dark blue, keycap set with vibrant blue and red backlighting. The hand is positioned in the lower-left to center area of the frame, with fingers curved over the keys. The background is heavily blurred, showing more of the keyboard and some indistinct light sources, creating a bokeh effect. The overall color palette is dominated by deep blues and purples, with the red and blue key lights providing a strong contrast. A white rectangular frame is superimposed over the image, with the text centered within it.

Thank you
for your attention

A close-up, artistic photograph of a hand typing on a computer keyboard. The keyboard has a dark, possibly black or dark blue, keycap set with vibrant blue and red backlighting. The hand is positioned in the lower-left to center area of the frame, with fingers pressing down on the keys. The background is dark and out of focus, showing more of the keyboard and some ambient light. A white rectangular frame is superimposed over the image, with the URL text centered within it.

<https://demo.webis.de/city-of-disguise/>

Sources

- Icons from <https://www.flaticon.com/>
- Cartoons from <https://www.pinterest.de>
- Avi Arampatzis, George Drosatos, Pavlos S. Efraimidis. “Versatile Query Scrambling for Private Web Search”. In: Information Retrieval Journal. Springer, 2015, pp. 331–358.
- Avi Arampatzis, Jaap Kamps. “A Study of Query Length”. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Singapore, Singapore, 2008, pp. 811–812.
- Carly O’Neil, James Purvis, Leif Azzopardi. “Fu-finder: A Game for Studying Querying Behaviours”. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management. Glasgow, Scotland, UK, 2011, pp. 2561–2564.
- Greg Pass, Abdur Chowdhury, Cayley Torgeson. “A Picture of Search”. In: Proceedings of the 1st International Conference on Scalable Information Systems. Hong Kong, China, 2006.