Exploiting Social Networks for Internet Search

Alan Mislove[1] Krishna Gummadi[2] Peter Druschel[2]

[1]Max Planck Institute for Software Systems

[2]Rice University

Presented by Danny Tarlow
October 4, 2007

The General Plan

1. Their Big Questions
   - What types of content is Google bad at finding?
   - Can this information be found by exploiting a social network?
2. My Big Questions
   - Would we want to share this content?
   - Do their experiments convince us of anything?
   - What could they have done to make this better?
3. Further Discussion

Outline

Outline

What types of content is Google bad at finding?

They say: Content that is...

- New
- Ambiguous
- Isolated

Outline

Can this information be found by exploiting a social network?

They say:

- Yes

Outline

Outline

Would we want to share this content?

New content

First... can we define this content?

- How do *you* find recently published content?

Would we want to share this content?

New content

First... can we define this content?

- How do *you* find recently published content?
  - I use RSS feeds (Google Reader)

Would we want to share this content?

New content

My simple experiment: How long does it take Google to index?

BBC News Front Page
10 min - 0 0
22 min - 1
2 hrs - 1 1 1 1
4 hrs - 0 1 1 0 1
5 hrs - 1 1
6 hrs - 1 0

Would we want to share this content?

New content

My simple experiment: How long does it take Google to index?

ESPN.com
2 hrs - 0 0 0
11 hrs - 1

Would we want to share this content?

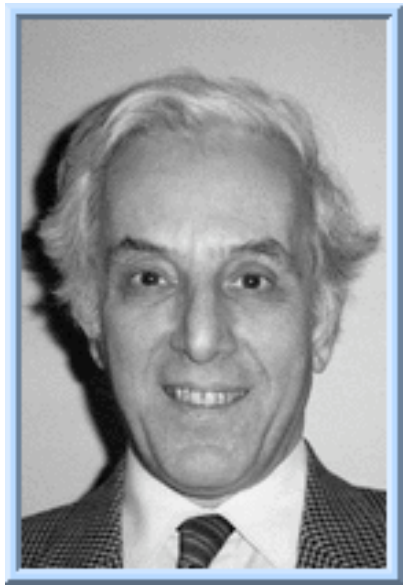
New content

My simple experiment: How long does it take Google to index?

PhD Comics
24 hrs - 1

Would we want to share this content?

Ambiguous content



Michael Jackson
9th result for
“Michael Jackson computer science"
> 500 otherwise

Michael Jackson

Would we want to share this content?

Isolated and poorly linked content

1. Deep web
   - ... /pres0031.html
   - ... /target21.html
   - On personal homepages
2. Dark web
   - http://72. ... .163/status.asp

Why is it isolated?

Outline

Main Result 1

13.3% of URLs viewed were in PeerSpective but not Google

- Were these cross-user views or repeat visits/refreshes??

Main Result 1

13.3% of URLs viewed were in PeerSpective but not Google

- Were these cross-user views or repeat visits/refreshes??

Main Result 2

Augmenting Google results with PeerSpective results yielded 9% more clicks

- Were these cross-user clicks or “bookmarks”?
- Is clicks the right metric to use?
- Bias is a known issue with search results.

Main Result 2

Augmenting Google results with PeerSpective results yielded 9% more clicks

- Were these cross-user clicks or “bookmarks”?
- Is clicks the right metric to use?
- Bias is a known issue with search results.

Main Result 2

Augmenting Google results with PeerSpective results yielded 9% more clicks

- Were these cross-user clicks or “bookmarks”?
- Is clicks the right metric to use?
- Bias is a known issue with search results.

Main Result 2

Augmenting Google results with PeerSpective results yielded 9% more clicks

- Were these cross-user clicks or “bookmarks”?
- Is clicks the right metric to use?
- Bias is a known issue with search results.

Outline

What could they have done to make this better?

My two cents:

- Distinguish between within-user behavior and cross-user behavior
- Run experiments comparing their system to:
  - 5 additional random results
  - 5 additional Google results (results 11-15)
  - 5 results from Google Scholar
- Focus more specifically on the disambiguation problem

What could they have done to make this better?

My two cents:

- Distinguish between within-user behavior and cross-user behavior
- Run experiments comparing their system to:
  - 5 additional random results
  - 5 additional Google results (results 11-15)
  - 5 results from Google Scholar
- Focus more specifically on the disambiguation problem

What could they have done to make this better?

My two cents:

- Distinguish between within-user behavior and cross-user behavior
- Run experiments comparing their system to:
  - 5 additional random results
  - 5 additional Google results (results 11-15)
  - 5 results from Google Scholar
- Focus more specifically on the disambiguation problem

What could they have done to make this better?

My two cents:

- Distinguish between within-user behavior and cross-user behavior
- Run experiments comparing their system to:
  - 5 additional random results
  - 5 additional Google results (results 11-15)
  - 5 results from Google Scholar
- Focus more specifically on the disambiguation problem

What could they have done to make this better?

My two cents:

- Distinguish between within-user behavior and cross-user behavior
- Run experiments comparing their system to:
  - 5 additional random results
  - 5 additional Google results (results 11-15)
  - 5 results from Google Scholar
- Focus more specifically on the disambiguation problem

What could they have done to make this better?

My two cents:

- Distinguish between within-user behavior and cross-user behavior
- Run experiments comparing their system to:
  - 5 additional random results
  - 5 additional Google results (results 11-15)
  - 5 results from Google Scholar
- Focus more specifically on the disambiguation problem

What could they have done to make this better?

My two cents:

- Distinguish between within-user behavior and cross-user behavior
- Run experiments comparing their system to:
  - 5 additional random results
  - 5 additional Google results (results 11-15)
  - 5 results from Google Scholar
- Focus more specifically on the disambiguation problem

What could they have done to make this better?

Your two cents?

What could they have done to make this better?

Your two cents?

Outline

More Questions

Are they conflating “browsing” with “searching”?

- e.g. StumbleUpon

More Questions

Are they conflating “browsing” with “searching”?

- e.g. StumbleUpon


http://www.stumbleupon.com/
Stumble!
I like it!
Send to
Channel:
All
Favorite

More Questions

Do they need a social network, or would a recommendation system work?

- Could Google build this same system by looking at their logs?

More Questions

Do they need a social network, or would a recommendation system work?

- Could Google build this same system by looking at their logs?

More Questions

Is privacy a deal-breaker for real-world deployment?

- How would you design a system that respects privacy?

More Questions

Is privacy a deal-breaker for real-world deployment?

- How would you design a system that respects privacy?

More Questions

Is is possible to spam PeerSpective?

- How would you design a system that is robust to spam?

More Questions

Is is possible to spam PeerSpective?

- How would you design a system that is robust to spam?

More Questions

What are characteristics of a problem that you should exploit social networks to solve?