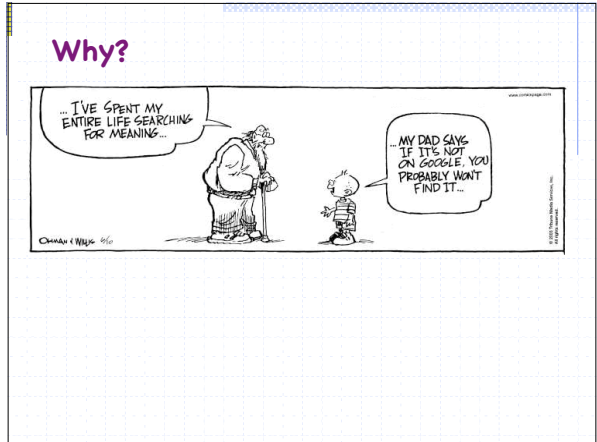


Pay Attention Search Engine!
My web site is important!

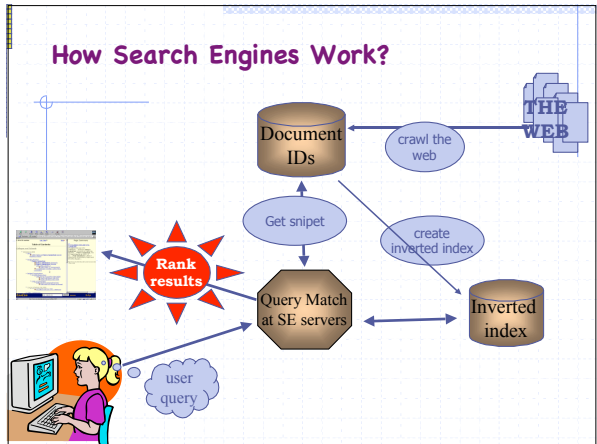
Panagiotis Takis Metaxas
Computer Science Department
Wellesley College

WAAM 2007



10 Questions

- ◆ How Search Engines work?
- ◆ What do SE's see and do not see?
- ◆ How a query is matched against the index?
- ◆ Search Results vs. Paid Placement
- ◆ What kind of searchers are there?
- ◆ What do SEO's do?
- ◆ How do you make your page SE-friendly?
- ◆ What is Web Spam and how is it discovered?
- ◆ How Web Spam shapes the development of SE's?
- ◆ Is there a future for SEOs?



What do SE's see and do not see?

- ◆ Text
 - Page contents
 - Title
 - Hlinks
 - Anchor text
 - URLs
 - ◆ Images
 - ALT tag
 - File name
 - File format
 - ◆ See but ignore
 - Meta-tags
- Do not see:
- ◆ The image itself
 - ◆ Text on image
 - ◆ Flash
 - ◆ Video
 - ◆ Javascript
 - ◆ Deep Web
 - Which is much greater

How a query is matched?

- ◆ Each word (- stop words)
- ◆ Rare words have priority
- ◆ Word density (< 10%)
- ◆ Word sequence
- ◆ Priority in matches given to words in:
 - Titles
 - URLs
 - Low-depth paths
 - Starting sentence of body

Search Results vs. Paid Placement

- ◆ 87% of searchers click on first page only (most only above fold)
 - Of them, at least 60-70% click search results
 - Therefore, 30 - 40% click paid placements
- ◆ Some paid placement easily confused with search results
 - SE's cash big \$\$\$ on acquired trust

What kind of searchers are there?

- ◆ Navigational about 40%
 - Find specific web site
 - Barnes and noble
- ◆ Informational about 40%
 - Find information about something
 - Safety of ADHD drugs
- ◆ Transactional about 20%
 - Do something (buy, sell, etc.)
 - iPod 30MB
- ◆ Who do you want to attract?

What do SEO's do?

- ◆ Search Engine Optimization industry
- ◆ "Help you promote your site"
- ◆ "Ethical"
 - Create well formed SE-friendly pages
 - Increase links to your site
 - Turn visits into actions
- ◆ "Unethical"
 - "black hat" tricks
- ◆ Where do you draw the line?
 - Who is making the call?

How you make a page SE-friendly?

- ◆ Web Graph has a shape:



- ◆ Get your page visited by SE's
- ◆ Choose target keywords so that SE's will notice you
- ◆ Optimize content presentation so that SE's will rank you high
 - I told you what counts, do your homework
- ◆ Place your page in the SCC of the Web Graph

What is Web Spam and how is it discovered?

- ◆ "Forcing ranking higher they deserve" in SE results
- ◆ **Web Spam:**
 - Attempt to **modify** the web (its structure and contents), and thus **influence** search engine results in ways **beneficial** to web spammers
- ◆ How do they do it?
 - Content spam
 - Link spam
 - cloaking
- ◆ SE's do not like it - Why?
 - Money: They represent lost potential revenues (without spam, more advertising revenue)
 - Quality: They want to be able to control value of search results

How Web Spam shapes the development of SE's?

A brief History of Search Engine Evolution

- ◆ **1st Generation (ca 1994):**
 - AltaVista, Excite, Infoseek...
 - Ranking based on **Content**:
 - Pure Information Retrieval
- ◆ **2nd Generation (ca 1996):**
 - Lycos
 - Ranking based on **Content + Structure**
 - Site Popularity
- ◆ **3rd Generation (ca 1998):**
 - Google, Teoma, Yahoo
 - Ranking based on **Content + Structure + Value**
 - Page Reputation
- ◆ **In the Works**
 - Ranking based on "the need behind the query"

Future of Search Engines and SEOs?

- ◆ Bright future for both
- ◆ But too much power at few SE's
- ◆ Need more competitors
- ◆ Need specialized SE's
- ◆ Need enhanced privacy
- ◆ Critical Thinking, Education
 - Realize how do we know what we know
 - "Of course it's true; I saw it on the Internet!"
- ◆ Cyber-social Structures that mimic Societal ones
 - Know why to trust or distrust
 - Who do you trust on a particular subject?
- ◆ A Search Engine per Browser
 - Easier to fool one search engine than millions of users
 - Enable the reader to keep track of her trust network
 - Tools of cyber trust

Thank You!

PMetaxas - wellesley.edu