

Computer Science 1033 – Week 7

SEARCHING

Click on Bart's blackboard to see the Google trends!

"There is no Ctrl button on Chuck Norris's computer. Chuck Norris is always in control."

"Apple pays Chuck Norris 99 cents every time he listens to a song."

Fun with Google

- *The Wizard of Oz*

Slide 2 of 47

Overview of Today's Topics

- Marketing your webpage
- Page Rank Algorithm
- Searching
- Review

Slide 3 of 47

Textbook Readings for this Week

- Websites
 - Online Visibility

Slide 4 of 47

Big Ideas!

- Big Idea 1
 - The quarterback and the nerd!
- Big Idea 2
 - Moving up in the rankings...WE CAN MAKE IT HAPPEN!

Slide 5 of 47

Marketing your Website

- Include the web address for your new site:
 - As part of your signature on outgoing email
 - On all printed material → letterhead, business cards, labels, catalogues, posters, media advertisements
- Try to make your website be in the “first ten listed” when your customers search using search engines for your site.
 - How can you do that, we will give you some tips 😊

Slide 6 of 47

Search Engines

- **Question:** What is the most popular search engine?
- **QUESTION** → What % of the market does the most popular search engine have?
 - [Real Time Info](#)

Slide 7 of 47

How does a Search Engine work?



- 2 Parts:
 - **Part 1:** Finding all the data on the Web and building a database. Similar to Librarians getting new books, cataloguing them and putting them in the library!
 - **Part 2:** Given keywords from a searcher (person looking for a topic), returning the “BEST/MOST APPROPRIATE” pages. Similar to a patron walking into a library, going to the card catalogue and looking for a book.

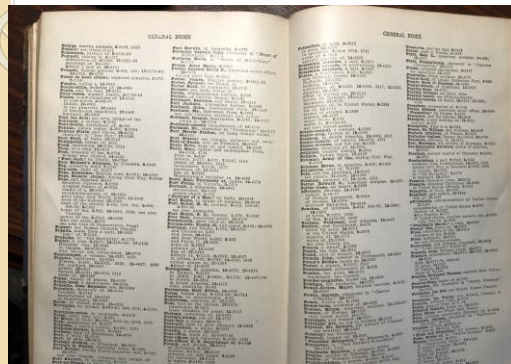
Slide 8 of 47

Part 1: Building the Search Database

- **Web Crawlers** or **Web Spiders** crawl the internet constantly, going from web page to web page via links, looking at all the words on the page, building an index (database):
- Index contains list of alphabetical list of words it finds, where within the page the word was and the links (URL to the page) where it found the words.
- Words are called **keywords**

Slide 9 of 47

Part 1: Index/Database



Part 2: How does the search engine decide which pages to return to the searcher?

- Uses index to decide which pages have the given keywords
- Every engine uses slightly different algorithms to decide the order of displaying the returned pages
- Google uses the “PageRank” algorithm as ONE of the factors to decide what order to present the pages it found to you.

Slide 11 of 47

BG (Before Google...)

- Around 1990, there was no index, so you had to search EVERYTHING (like looking for a needle in a haystack or looking through a book page by page for one word)
- Sample Search from early 1990s BEFORE World Wide Web
 - <http://www.youtube.com/watch?v=-PshzRzxEOE&t=5m45s>

Slide 12 of 47

BG (Before Google...)

- Around 1990 webpages were returned just by the file names of the sites and titles stored in the index but NOT the content, e.g. if you searched for “Dog”, you would get things like:
 - www.uwo.ca/dogs.html
 - www.uoft.ca/mydogisgreat.html OR
 - You might find a URL with a file that had was titled: “The Black Dog” BUT some sites only matched titles if the match was exact (e.g. would only find paper above if user had search for “The Black Dog”)
 - Also, wouldn't find your page if it wasn't in the index

Slide 13 of 47

BG (Before Google..)

- Then came the **World Wide Web**, ~1991
 - Developed by **Tim Berners Lee** in Geneva Switzerland



BG (Before Google)

- Around 1994 some software allowed the user to submit pages they wanted included in the index
- 1994 Yahoo began indexing webpages that included descriptions with
 - Had a searchable index
 - Commercial sites
- Also in 1994 the index began to include more than one webpage (not just one)



BG (Before Google)

- Around 1995
 - Unlimited bandwidth for the first time
 - First natural language queries, E.g “dog, 1992” rather than “Select * from index where year=1992 and pet='dog'”
 - Advanced searching techniques such as:
 - Ca* would return webpages with return Canada, Cats, Calico
 - Add your URL to the index and it gets added within 24 hours
 - Results are still just returned mostly randomly or with basic techniques (e.g. the page that contains the word “dogs” the most would be returned first)

Slide 16 of 47

- Then in 1996 Larry Page and Sergey Brin start working on the “BackRub” algorithm. The beginnings of:
- Da... da... da...

Slide 17 of 47


What is Google's PageRank Algorithm?

- Algorithm gives each webpage returned from the keyword search a weight between 0 and 1.
- The higher the weight given to the page, the more likely it is that this page will be displayed first to you.
- **Question: How is the weighting done?**


Slide 18 of 47

How do we, as humans, decide who or what is "important" ?


A1




A2



B1



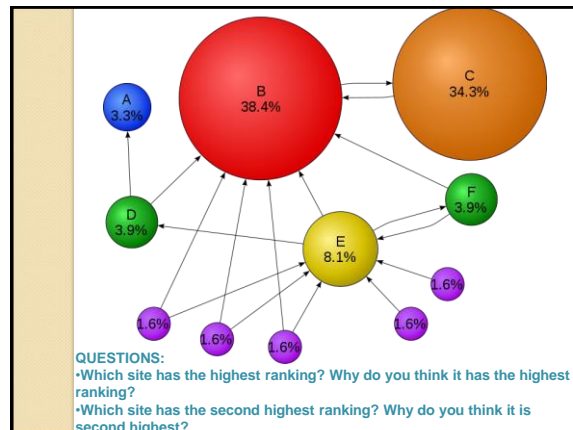
B2



Guy A or Guy B ?
Does the guy get to tell us he is important or do we decide?
Q1. Do you know Guy A
Q2. Do you know Guy B
Activity: Google this: who are some handsome actors in Hollywood

What is the name of A?
What is the name of B?

Slide 19 of 47



How PageRank Works – A simple world with only FOUR pages on the entire Internet!

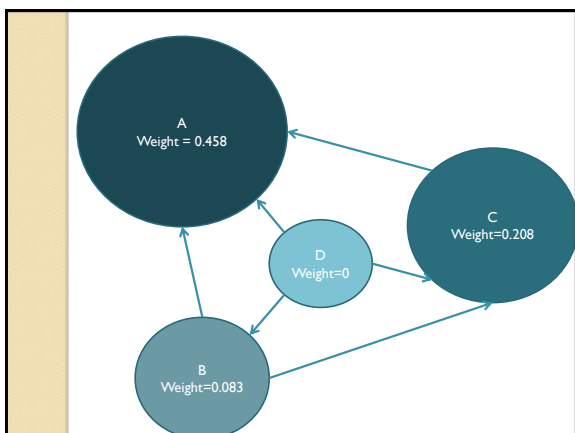
- First, assume we only have 4 pages, Page A, Page B, Page C and Page D on the Internet to simplify this.
- Each page is given a weight of 0.25 (1 divided by 4)
- Scenario 1:**
 - Pages B, C, and D all have a link to Page A. (Thus page A must be very useful because everyone is pointing at it).
 - Then pages B, C and D are each giving their 0.25 rank to A, so A gets a ranking of **0.75**
- Scenario 2:**
 - Page B links to A and C (0.25 divided between 2 pages)
 - Page C just links to A (all of 0.25 goes to A)
 - Page D links to A, B and C (0.25 divide between 3 pages)
 - The Weight of A is now:
 - 0.25/2 (Bs ranking) + 0.25 (Cs ranking) + 0.25/3 (Ds ranking)
 - 0.125 + 0.25 + 0.083 = **0.458**
- THUS, pages with lots of links pointing at them, must be important so they get the highest weight/ranking**

Slide 21 of 47

Let's Try It Out

- Scenario 2:**
 - Page B links to A and C (0.25 divided between 2 pages)
 - Page C just links to A (all of 0.25 goes to A)
 - Page D links to A, B and C (0.25 divide between 3 pages)
 - The Weight of A is now:
 - 0.25/2 (Bs ranking) + 0.25 (Cs ranking) + 0.25/3 (Ds ranking)
 - 0.125 + 0.25 + 0.083 = **0.458**
- Question: What ranking will pages B, C and D have?**

Slide 22 of 47



Summary of Page Rank

- Page Rank evaluates 2 key factors:
 - How many links are there to a web page
 - What is the quality of the linking sites (although a high ranked page with LOTS of links on it may pass you less because it is spread too thinly)
- Page Rank does not take into account the content of the page (thus frequent content updates don't improve Page Rank necessarily)
- Page Rank ranks web pages NOT web sites
- Each inbound link is important in the overall total except for banned site, they don't count.
- Each Page Rank level is progressively harder to reach, it is thought to be calculated on a logarithmic scale.

Slide 24 of 47

How to tick off Google ☺

- Because pages with lots of other pages that point at them get the highest weight, “CLEVER” people have tried making up lots of fake websites that all point to their website, thus boosting their ranking
- If Google catches you doing this, they take you OUT OF THE INDEX/DATABASE!

Slide 25 of 47

How do I get my page higher up on the results page that Google returns?

- Called Search Engine Optimization (SEO)
- 3 Components you should look at when you are trying to make your site more visible are:
 - Text
 - Links
 - Popularity (we just talked about it → pagerank)

Slide 26 of 47

Number 1: Text Component

- Users will type a **set of words or phrases** into the a search engine box to find what they want.
- It is crucial you know what sets of words or phrases, also called **keywords**, that users will type to find your products or services.
 - **STEP 1: IDENTIFY THE CRUCIAL KEYWORDS**
- Once you know the words, make sure they are in the areas that a search engine will consider to be the most important places.
 - **STEP 2: PUT THE KEYWORDS IN THE BEST LOCATIONS**
- Search Engines ALSO use the placement of keywords to weight the webpage when deciding which pages to return in the top positions (as well as PageRank algorithm weighting)

Slide 27 of 47

Identify Crucial Keywords

- **Five-Second Usability Test:**
 - Show people your site for 5 second, then have them write down everything they remember about the webpage. If they did not write down any of your keywords, then you need to redo your site!
- **QUESTION – What do you think the site is about?**
- Brainstorm to find keywords, use your marketing material and see what pops out.
- Think about how people might combine your keywords
- **HINT:** keep updating your keywords:
 - **Question: When do users commonly search for the word allergies?**

Slide 28 of 47

Put

CS1033 - Multimedia and Communications
Assignment #3 (15%)
Website Design & Dreamweaver

Due: Friday, November 14, 2014
Assignment Opened October 14, 2014

PROJECT:

Stellar Stationery has asked you to create a website for their business to showcase their collections of invitations and stationery for weddings, selection of customized decorative papers that will act as an online retail store to payment aspect of it). Take a look at the website they have designed for you to see how they have represented their products online to a spectacular! <p>

CRITERIA:

1. **Header:** for your website made with Photoshop.
 - The color scheme must be EXACTLY 300 pixels wide.
 - The color scheme must be EXACTLY 300 pixels wide.
 - The color scheme must be EXACTLY 300 pixels wide.
 - The color scheme must be EXACTLY 300 pixels wide.
 - The color scheme must be EXACTLY 300 pixels wide.
2. **Content for the web pages:** (total content)
 - The document must be in a .htm file format. The content to be used for creating the web pages must be in a .htm file format. The content to be used for creating the web pages must be in a .htm file format. The content to be used for creating the web pages must be in a .htm file format.
 - The content must be in a .htm file format. The content to be used for creating the web pages must be in a .htm file format. The content to be used for creating the web pages must be in a .htm file format.
 - The content must be in a .htm file format. The content to be used for creating the web pages must be in a .htm file format. The content to be used for creating the web pages must be in a .htm file format.
3. **Number of Web pages:** You need to create 3 web pages in total. These pages need to be created in a .htm file format. The content to be used for creating the web pages must be in a .htm file format. The content to be used for creating the web pages must be in a .htm file format.

computer science in canada

Google search results for "computer science in canada".

Title-Tag is considered the MOST important text by all search engines

Title tag is what is shown as the link for all returned results

Title tag is text shown where people bookmark your site

Can be 100 characters

www.appboard.com - May 15 best computer scienc...
May 15 2015 - 1 job
www.appboard.com - May 15 best computer scienc...
May 15 2015 - 1 job
www.appboard.com - May 15 best computer scienc...
May 15 2015 - 1 job

Body of the Webpage

- Search engines think keywords found near the top of the page make the page more relevant and thus give those pages a higher ranking (**Keyword Prominence**)
- **Keyword Proximity** also gives a higher ranking
 - **Question:** If you are searching for 2 or more keywords, like "civil" and "war", how do you think search engines ranks pages?
- **Keyword Density** also gives a higher ranking
 - **Question:** What do you think keyword density is?

Slide 31 of 47

Assume we are search for:
"Chuck" and "Norris"

Keyword Prominence
Keyword Proximity
Keyword Density

Another trick that people try to beat Google – Keyword Stuffing

- A simple site about cats:
http://www.csd.uwo.ca/~lreid/cs1033/keywords_tuffing.html
- **Question:** How did I do keyword stuffing with the above page? (I did 2 things, can you find them?)
- Another example –how is this stuffed? → [YouTube video](#) (view source)
- Moral of the story: you USUALLY can't outwit Google, just use good content to win over your visitors!

Slide 33 of 47

Number 2: Link Component

- Make sure that your links are easy to navigate so that web spiders can search through your content easily and index ALL your content.
- NOTE: It is hard for spiders to crawl through:
 - Image maps
 - JavaScript
 - Database Driven Web Pages (ones with ?, &, etc in the URL that are created, CGI pages can be problematic)
 - Links in Flash documents
 - Poorly written html (don't use MS Word, etc to write the html tags!)
- Make sure you have keywords on your links...

Slide 34 of 47

Links

- Text in and around links is used by search engines, for example:
 - **Instead of this**

```
<a href=http://www.yarnsandwools.com>Learn more</a> about our wools and yarns
```
 - **Do this:**

```
<a href=http://www.yarnsandwools.com>Learn more about our wools and yarns</a>
```
 - **Or this:**

```
Learn more about our <a href=http://www.yarnsandwools.com>wools and yarns</a>
```

Number 3: Popularity Component

- Popularity is broken into two parts:
 - **Link Popularity**
 - Not just getting many links to your site, the links to your site must be high quality links! REMEMBER PAGERANK!
 - **Click-through popularity**
 - Measures how many times your site is clicked on
 - Measures how often a user returns to your site
 - Measures how long a user stays at your site
 - **Question:** Click-through popularity is not used by some search engines to give a weight to a page, WHY?
- [Review 1](#)
- [Review 2](#)

Slide 36 of 47

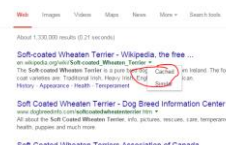
Some tips to increase your popularity weighting

- Check your competitors web pages, try to figure out why their site is ranked higher than yours.
 - For example:
 - Go to <http://www.google.ca>
 - Search for Horseback Riding London Ontario
 - Suppose I work for Circle R Ranch, why did my site not show up first?

Slide 37 of 47

How can you get Google to find you faster?

- Submit your site to here: <http://www.google.com/submityourcontent/#>
- **Question:** When you do a search in google, what happens if you click on cached? For example, search for *Soft Coated Wheaten Terriers Dogs*, then click on the cache link → <http://www.google.ca>
 - Hover over the green arrow next to the link:



This Page or...

That Page?

- | | |
|---|---|
| • https://we.love.cats.ca | • http://we.love.cats.ca |
| • ThisPage | • OrThisPage |
| • ThisPage | • OrThisPage? |
| • ThisPage | • OrThisPage! |
| • ThisPage | • OrThisPage! |
| • ThisPage | • OrThisPage! |
| • ThisPage | • OrThisPage! |
| • ThisPage | • OrThisPage. |
| • ThisPage | • OrThisPage? |

Google Quiz – Can you find which one will rank higher?

A few more tips...

- Improve your page speed (Google checks how fast a page loads – check your code/files)
- Optimize for voice search (that is how people search now)
- Optimize for mobile searches (over 60% of searches are now mobile)
- Have a great user experience – Google checks:
 - How long did the user stay on your page
 - How many times do they visit your site
 - How many pages on your site do they visit
 - What is your **bounce rate** (% of users who leave your site after viewing only 1 page)

Slide 40 of 47

A few more tricks...

- Quotes → Australian Maple Syrup
- Minus Operator → orange -fruit
- Wildcard → “how to * a cake”
- FileType → filetype:pdf learn italian
- Before and After → "Conan O'Brien" before:2000-01-01
- ImageSize → boat imagesize:100X100
- Site → oprah site:indigo.ca

Slide 41 of 47

Search Insights

- <https://www.google.ca/trends/>
 - Search for “Costumes”, “Flowers”, “Chocolates”
 - How it works → <https://www.youtube.com/watch?v=EOdN0CFIN68>
- Last year's trends: <https://trends.google.ca/trends/yis/2019/CA/>
- **Question:** How about the top keywords for today? <http://www.google.ca/trends/hottrends>

Slide 42 of 47

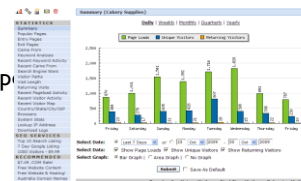
Usage Statistic: What are they?

- It is important to know what your visitors like and dislike about your website so you can improve it!
- Analyze your stats from your IPS provider or from companies on line
- **Question:** What kind of stats do you think you could get about a website?

Slide 43 of 47

Usage Statistics: An Example

- Entry page
- Exit page
- Here is a real report
 - Total Hits: 3,357
 - Total Files: 1,441
 - Total Pages: 413
 - Total Visits: 337
- Hits: Each file (a html file, a graphic file,) sent to a client
- Pages: A webpage
- **Question:** Which will always be bigger: hits or pages?



Slide 44 of 47

Hit Counters

- I installed a hit counter from <http://statcounter.com> into this page: <http://www.csd.uwo.ca/~lreid/cs033/TestCounter.html>



Slide 45 of 47

What if you do NOT want people to find your website?

- https://support.google.com/webmasters/answer/93710?visit_id=1-636564907849131550-2586980481&rd=1

<meta> tag

To prevent most search engine web crawlers from indexing a page on your site, place the following meta tag into the <head> section of your page:

```
<meta name="robots" content="noindex">
```

To prevent only Google web crawlers from indexing a page:

```
<meta name="googlebot" content="noindex">
```

Slide 46 of 47

Google

Some more facts from www.wikipedia.org

- Their search engine was originally nicknamed "BackRub" because the system checked [backlinks](#) to estimate the importance of a site.
- Both Brin and Page were against using advertising pop-ups in a search engine, or an "advertising funded search engines" model, and they wrote a research paper in 1998 on the topic while still students. However, they soon changed their minds and early on allowed simple text ads.
- 99% of Google's revenue is derived from its advertising programs.
 - 2006 earnings from ads → 10.5 Billion, from other revenue 112 million
 - 2017 earnings → 110 Billion (86% from ads)
- [Google usually near the top of lists of the best places to work](#)

Slide 47 of 47

Let's Review

- <https://www.youtube.com/watch?v=0eKVizvYSUQ>
- http://www.youtube.com/watch?v=NWLh_yh5_v0

Slide 48 of 47