Warm up Questions:
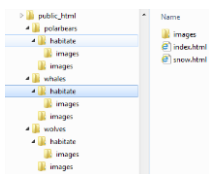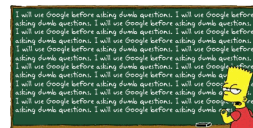
- **Question**: Assuming I have this → directory structure for my website on the domain *arcticanimals.org*
  1. What is the shortest URL I have to type to see the habitat home page for whales?
  2. What URL would I type to see the wolves home page?

Slide 1 of 70

Computer Science 1033 – Week 8
WEB SITE DESIGN & SEARCHING

Click on Bart's blackboard to see the Google ad that was on during the 2010 Super Bowl

*"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

- Type in the following terms to http://www.google.com and then for each term, hit the "I'm feeling lucky" button:
  ◦ elgoog
  ◦ Google gravity – underwater sphere
  ◦ Chuck Norris google easter egg
  ◦ http://lmgtfy.com/
- Something else cool

Slide 3 of 70

Overview of Today's Topics

- Announcements
- Using a table for simple layout
- Publishing your webpage
- Page Rank Algorithm
- Searching
- Review

Slide 4 of 70

Textbook Readings for this Week

- Websites
  ◦ Online Visibility

Slide 5 of 70

Announcements

- Web Assignment due Friday (March 13th)
- Final Exam
  ◦ Thursday, April 23rd at 2:00 pm
  ◦ All multiple choice – 2 hour time period
  ◦ Bring:
    • Pencil (soft) and eraser
    • Student card
  ◦ Do NOT bring: calculator, iPod, hat, etc..

Slide 6 of 70

1

Hints for Web Assignment

- **Remember:**
  - ◦ Titles
  - ◦ Headers
  - ◦ File names, spaces, lowercase
  - ◦ Images
    - • Alt, Title
    - • Size within page
    - • On each page
  - ◦ Banner
  - ◦ No Scrolling horizontal/vertically to see buttons
  - ◦ Consistency
  - ◦ Buttons → look, ease of use
  - ◦ Back to top
  - ◦ References page layout
  - ◦ Text colours, contrast
  - ◦ Broken links
    - • Check from a different computer!
  - ◦ Underlining
  - ◦ Colours
    - • Link Colours
  - ◦ Paragraphs
  - ◦ Padding
  - ◦ Followed the instructions (make the anchor links, etc..)

Slide 8 of 70

How to do Web Assignment

- Colours → look at: http://kuler.adobe.com/
- https://www.design-seeds.com/
- Marks for:
- Nice Banner
- Good Colour Scheme
- Alt/title tag on banner/images
- No scrolling horizontally EVER at 1000 pixels
- Consistency in the layout of the pages and the buttons/NICE LAYOUT ON EVERY PAGE
- NO Underlining

Slide 9 of 70

- Roll over buttons get 1 or 2 marks more
- PROPERTY TITLES! Website – Web Page
  - ◦ Antonio's Restaurant – Home
  - ◦ Antonio's Restaurant – References
- Must use table to lay things out
- Headings on pages
- Use an image for each page/good size
- Layout paragraphs well (don't squish on edge)
- Bold/highlight headings

Announcements

- Web Assignment
  - ◦ IT will take you longer than you think, give yourself LOTS of time to finish it and to hand it in (Hand it in on before Wednesday, the labs will be VERY busy on Thursday and Friday).
  - ◦ Hints:
    - • Remember your titles
    - • Think about layout, consistency, ease of use!
    - • Follow the specs
    - • Use a table for a clean layout (put the banner in the top row and content in the bottom row)
    - • Common Mistakes → http://www.csd.uwo.ca/~lreid/cs1033/assignment3/SamplesOfCommonMistakes/

Slide 11 of 70

Announcements

- Web Assignment Hints
  - • THINGS THAT CAN GO HORRIBLY WRONG ☹
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student1/
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student2
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student3/assign3/howtohelp.html (need to scroll for buttons, bullets, banner)
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student9/major/ (consistency)
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student19/
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student20 (edges)
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student14 (scrolling and banner on next page)
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student22
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student18/Contact%20Us.html (file names and button movement)
  - • THINGS THAT CAN GO RIGHT ☺
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student8/assign3/registration.html
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student22/assign3/
    - • http://cs1033.gaul.csd.uwo.ca/~lreid2/other/student15/index.html

Slide 12 of 70

- Paragraphs → Centering → DON'T
- Hyphens on lists
- Emails links
- Back to top, anchors
- Think about link colours
- Link lengths
  - ◦ http://www.csd.uwo.ca/images/CSD_long_photo_16.jpg vs.
  - ◦ Middlesex
- Common Mistake: renaming files/folder…this will TOTALLY SCREW up your website if you do not do it correctly!

Slide 13 of 70

Let's try it out

- Watch me start Web Assignment!

HTML Tables and BlueGriffon

...ns
...called a
...links, text,

...(called a

...e a simple

...ne of:

Tables expressed in terms of **Percentage**

- Here is an example: http://www.csd.uwo.ca/~lreid/outreach/iveysurveyTables.htm
- % of the browser, not the entire screen
- Another example: http://www.csd.uwo.ca/~lreid/cs1033/ExamplesForLectures/tablesexamples.html

Table expressed in terms of **Pixels**

- Resolution affects the way a page is displayed.
- In general, most people do NOT have their resolution below 1024 by 768. Thus if we make our page 1000 pixels wide EVERYONE should be able to view it. If we make our page 1200 pixels wide, some people (the ones whose resolution is still 1024 by 768) will have to scroll horizontally.

Resolution 1280 by 1024:
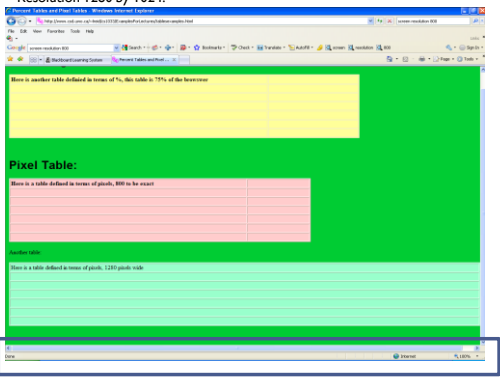
Resolution 800 by 600:
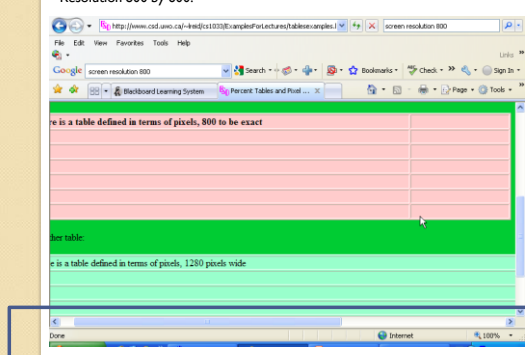
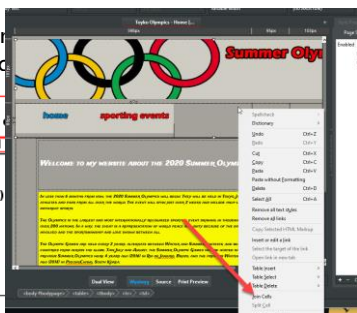## Some hints for using tables to organize your overall appearance:

- Have the bor
- Make sure yo

Table 1 (1 row x 1)
table 2 (1 row x 1

Table 3 (1 row x 1 col)

Slide 20 of 70

Publishing your Website

- Move your website from the machine you built the site on to the web server
- **Question**: What program do we use to perform the move?
- NOTE: Not all web servers allow certain FTP software to connect to them because of security reasons.

Slide 21 of 70

Publishing your Website

- **Question**: There are 4 pieces of information you need to in or the
- Qu Ch
- Ne
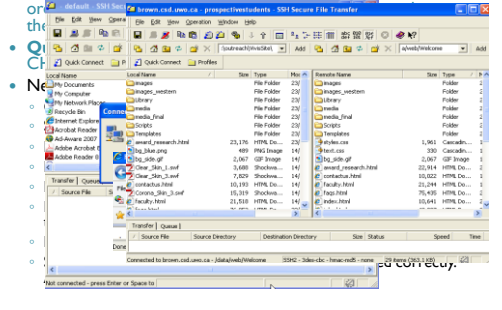
Slide 22 of 70

Hint: Use Google Fonts if you want cool fonts in your web page!

- http://www.google.com/fonts
- Click on red + button and then the black bar at the bottom that says *Families Selected* and copy and paste the code you get at the google site into your code below the <head> tag (the code will look like this →
  <link href="http://fonts.googleapis.co
- In BlueGriffon, view t below the <head> tag
  ◦ Save your file
  ◦ Highlight text
  ◦ Go to (no class) near t
  ◦ Go to style properties, ur name
  ◦ Click on GENERAL and press + and pick new font
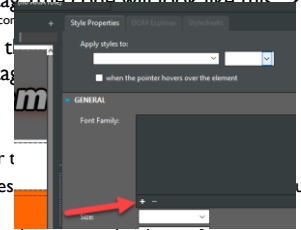
Slide 23 of 70

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 24 of 70

Break

- Check out some Poster Assignment:

Slide 25 of 70

Search Engines

- **Question**: What is the most popular search engine?
- **Question**: What search engine is coming in second? Hint: The Colbert Report
- **QUESTION** → What % of the market does the most popular search engine have?
  - Real Time Info

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.

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**

Part 1: Index/Database

- Index is stored in a really big database
- Looks sort of like this:

keyword	Index: Page and position within page			
aardvark	3 (body, 2nd word, 8th word)	5(body, 12th word)	20	
africa	2 (title)	7 (…	19	27	30
anteater	2 (title, body 1st word)	3 (…	5	7	

Index Number	Website found on
1	http://www.antartic.org/index.html
2	http://www.worldaidforafrica.org/info.html
3	http://www.worldwiderefuge.org/animals.html
4	http://www.csd.uwo.ca/~lreid/resume.html
5	http://www.AnimalsRUs.com

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.

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

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 32 of 70

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 beg[...] webpages that i[...] descriptions wit[...]
  - Had a searchable[...]
  - Commercial site[...]
- Also in 1994 the[...] began to include[...] webpage (not ju[...]

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 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 as the first result)

Slide 35 of 70

- Then in 1996 Larry Page and Sergey Brin start working on the "BackRub" algorithm. The beginnings of:
- Da… da[...]

Slide 36 of 70

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 37 of 70

6

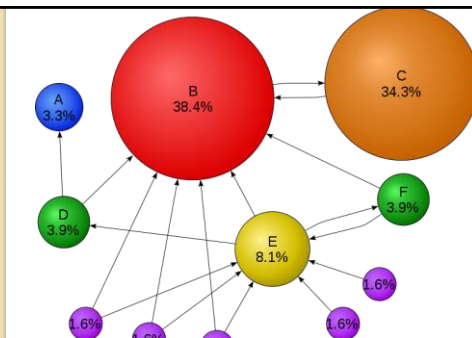## How do we, as humans, decide who or what is "important" ?
CLICKER QUESTION

A1   B1

A2   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 38 of 70

QUESTIONS:
•Which site has the highest ranking? Why do you think it has the highest ranking?
•Which site has the second highest ranking? Why do you think it is second highest?

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**
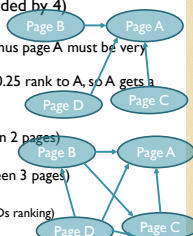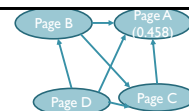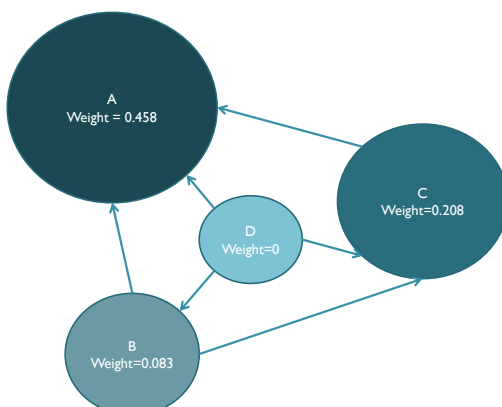
Slide 40 of 70

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 41 of 70

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 43 of 70

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!

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)

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)

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**?

- **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 when people bookmark your site

- Can be 100 characters but keep most important stuff in the first 63 characters

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*)
- **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 50 of 70

Assume we are search for:
Chuck, Norris
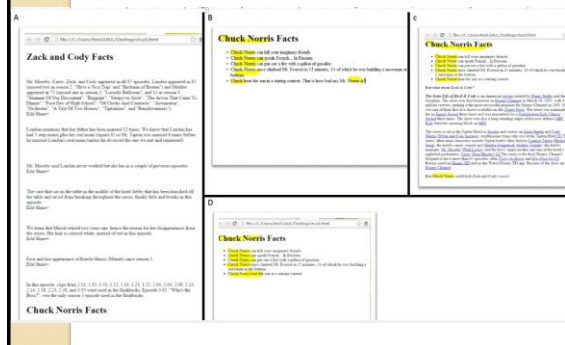
Another clever trick that people try to beat Google – Keyword Stuffing

- A simple site about cats: http://www.csd.uwo.ca/~lreid/cs1033/keywordstuffing.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 52 of 70

Body of the Webpage

- Text in and around links is used by search engines, for example:
  - **Instead of this**
    Learn more about our wools and yarns
  - **Do this:**
    Learn more about our wools and yarns
  - **Or this:**
    Learn more about our *wools and yarns*

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!)

Slide 54 of 70

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**?
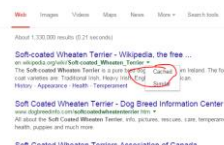- Review

Slide 55 of 70

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 56 of 70

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:

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

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?
ThisPage	OrThisPage?

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 59 of 70

Search Insights

- **https://www.google.ca/trends/**
  - **Search for "Costumes", "Flowers", "Chocolates"**
  - **How it works →** **http://www.youtube.com/watch?v=4uNrhACTv_c**

- **Question**: How about the top keywords for today? http://www.google.ca/trends/hottrends

Slide 60 of 70

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 61 of 70

Usage Statistics: An Example

- Entry page
- Exit page
- Here is a real rep[...]
  - 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**?
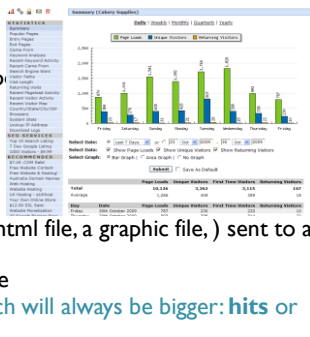
Slide 62 of 70

Hit Counters

- I installed a hit counter from http://statcounter.com into this page:

http://www.csd.uwo.ca/~lreid/cs033/TestCounter.html

Slide 63 of 70

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 64 of 70

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 65 of 70

Let's Review

- http://www.youtube.com/watch?v=NW1h_yh5_v0
- https://www.youtube.com/watch?v=0eKVizvYSUQ

Slide 66 of 70

Review

- **Question**: What is the PageRank algorithm?
- **Question**: Name the most popular search directory web site?
- **Question**: Name the most popular search engine website? What % of the market does it currently have?
- **Question:** What are the 3 components you must address to get your page a higher ranking?
- **Question**: Who started Google?
- **Question**: Who invented URLs, the World Wide Web and HTML?

Slide 67 of 70