Introduction to Internet Technology and Web Programming

Computer Science 103
Boston University
Vahid Azadeh-Ranjbar

These notes are based on the lecture notes provided by Professor Susan Worst.

Lecture 13-1

BU CS 103, Spring 2019

What's Left in CS 103

S	M	T	W	T	F	S
21	22	23	24	25	26	27
28	29	30	1	2	3	4
5	6	7	8	9	10	11

Final Project Presentation
Date 04/29 and 05/01

Final Project Website
Due 05/01

Final exam:
Date 05/08 at 6:00 – 8:00 pm

Open Labs
Date: 04/25 and 5/2

What will be Covered Today?

❖ JS (Wrap up)
❖ Visual Disabilities: Color Blindness
❖ Search engine optimization
❖ Audio Disability

JavaScript: Summary

❖ We learned several JavaScript utilities written by others and made available to us through free licenses.

❖ Also, we have done a mini-project (Calculator) by using JS/CSS/HTML from scratch.

Programming Languages for the Web

❖ As we know, web programming languages come in two types:
  ➤ Client-side:
    ✓ Lives on the user's computer (the "client")
    ✓ Incorporated into web browsers
    ✓ **JavaScript** is the only common one right now
  ➤ Server-side:
    ✓ Lives on the web server (e.g., DreamHost's computers)
    ✓ Examples: PHP, ASP.Net, ColdFusion

What Can JavaScript Do?

❖ Check data submitted via forms
❖ Create web page widgets such as maps, search engines, charts
❖ Create pop-up windows
❖ Make photos in slideshows move
❖ Update part of a page such as current time
❖ Save data in the user's browser in a cookie
❖ Send data from a web page to **Google Analytics**

JavaScript Basics

❖ JavaScript is not related to Java (except that they are both programming languages).
❖ JavaScript is incorporated into HTML pages using the **script tag**
❖ The script tag can either enclose JavaScript code or pull in an external file
❖ The script tag must have a closing tag
❖ The script tag can appear in the head or the body of an HTML page
❖ The order of script tags is important.
❖ External JavaScript files are ordinary text files (like HTML and CSS)
❖ Filenames MUST end in .js

JavaScript Disadvantage #1

❖ Users can turn JavaScript off in their browsers
❖ Example from Safari Preferences
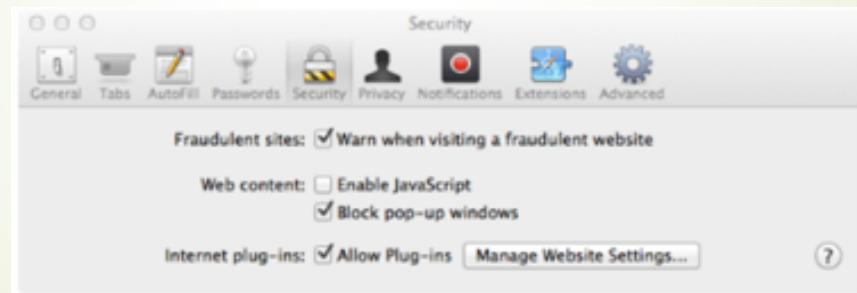
JavaScript Disadvantage #2

❖ Different web browsers support JavaScript differently

❖ However, support is much more uniform than it used to be

❖ The **jQuery library** helps overcome differences. Programmers interact with jQuery, and it does the heavy lifting.

❖ See the W3Schools.com jQuery Tutorial for information on jQuery **selectors**, **events**, and **methods.**

File Minification

❖ Notice that some of the JavaScript files used by Lightbox have ".min." in them E.g. **lightbox.min.js**

❖ Min is short for "minify"

➢ Makes JavaScript (or CSS) files shorter...

➢ ... and harder to read

❖ If you want to actually read minified files, Try the "Online Javascript Beautifier" http://jsbeautifier.org/

Working with Files and Folders

❖ **href** attributes of link tag and "a" tag, and **src** attributes of script tag and img tag, refer to other files

❖ These attributes can start with:

➢ **http://** , **https://**, or **//** - for files on the web host (not the local computer)

➢ **Filename** only – for files in the same folder

➢ **Folder name / filename** – for files in a subfolder

➢ **.. / folder name / filename** – for files in a parallel folder

Visual Disabilities: Color Blindness

➢ There is no assistive technology to help color-blind people to access the web.

➢ So, we need to add some features in out web design to help them.

Most people who are color blind use what assistive technology to access the web?

A. None
B. Special glasses that correct color vision
C. Screen magnifier
D. Screen reader
E. Screen Color Editor

Most people who are color blind…

A. See empty space where there is color
B. See only in black and white
C. Have trouble distinguishing between certain colors
D. None of the above

To make web pages accessible to people who are color blind…

A. Don't use color on web pages
B. Don't use color for anything important
C. Avoid use of red and green
D. Make sure color is not the only means of communicating important information

Don't Rely on Color Alone

❖ To help color-blind people we need to "**make sure that colors are not the only method of conveying important information**".

❖ Important information includes:
  ➢ headings
  ➢ hyperlinks
  ➢ The buttons/links you should click

Some Good Examples

❖ These examples convey important information using color AND other methods:
  ➢ Traffic light
  ➢ W3C HTML validator results page
  ➢ WebAIM hyperlinks in body of text

❖ Note: hyperlinks in menus, such as in a navigation bar, generally do not need to be underlined. The main concern is hyperlinks in the main text of a page. Hyperlinks need to be identified by a change in addition to color.

Facebook uses blue because Mark Zuckerberg is color blind.

A. True
B. False

Facebook uses blue because Mark Zuckerberg is color blind.

According to The New Yorker, **Zuckerberg** is red-green **colorblind**, which means the **color** he can see best is **blue**. That also happens to be the **color** that dominates the **Facebook** website and mobile app. September 20, 2010

Search Engine Optimization (SEO)

❖ "A methodology of strategies, techniques and tactics used to increase the amount of visitors to a website by obtaining a high-ranking placement in the search result page of a search engine (such as Google Search Engine)."

—Webopedia

Search Engine Optimization

❖ To optimize for search engines, it is helpful to understand how the search process works

1. How do search engines find web pages?
2. How do search engines know what a web page is about?
3. What determines how search engines present web pages in the results?

How Do Search Engines Find a Web Page? (**Discovery Step**)

❖ In the **Discovery** process, small computer programs (called Bots) which belong to search engines crawl the web, jumping from link to link and exploring each page.

➢ Bots from Google, Bing, and many international search engines are constantly crawling the web.

Helping Search Engines Find Your Pages

1. Sign up an account in **the Google Search Console** at http://www.google.com/webmasters/tools/

   ❖This allows Google to find your site.

2. Provide a **sitemap** to tell search engines about ALL of your pages.

3. Make a **robots.txt** file.

What Is a Sitemap?

A Sitemap is an **XML file** that

➢ lists URLs for a site along with additional metadata about each URL

➢ so that search engines can more intelligently crawl the site.

—sitemaps.org http://www.sitemaps.org/

More on Sitemaps

❖ Sitemaps are XML files

❖ Write with text editor

❖ Starts by declaring the version of XML being used

❖ one **<urlset>** tag per sitemap

❖ one **<url>** tag per file

❖ the URL of each page is placed at **<loc>** tag

❖ Other elements are optional

Sitemap Example

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
<url>
    <loc>http://www.domain.com /</loc>
    <lastmod>2008-01-01</lastmod>
    <changefreq>weekly</changefreq>
    <priority>0.8</priority>
</url>
 <url>
    <loc>http://www.domain.com/catalog?item=vacation_new_zealand</loc>
    <lastmod>2008-12-23</lastmod>
    <changefreq>weekly</changefreq>
</url>
</urlset>
```

Sitemap Example

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
<url>
    <loc>http://www.domain.com /</loc>
    <lastmod>2008-01-01</lastmod>
    <changefreq>weekly</changefreq>
    <priority>0.8</priority>
</url>
 <url>
    <loc>http://www.domain.com/catalog?item=vacation_new_zealand</loc>
    <lastmod>2008-12-23</lastmod>
    <changefreq>weekly</changefreq>
</url>
</urlset>
```

Sitemap Example

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
 <url>
     <loc>http://www.domain.com /</loc>
     <lastmod>2008-01-01</lastmod>
     <changefreq>weekly</changefreq>
     <priority>0.8</priority>
</url>
 <url>
     <loc>http://www.domain.com/catalog?item=vacation_new_zealand</loc>
     <lastmod>2008-12-23</lastmod>
     <changefreq>weekly</changefreq>
</url>
</urlset>
```

Sitemap Example

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
 <url>
     <loc>http://www.domain.com /</loc>
     <lastmod>2008-01-01</lastmod>
     <changefreq>weekly</changefreq>
     <priority>0.8</priority>
</url>
 <url>
     <loc>http://www.domain.com/catalog?item=vacation_new_zealand</loc>
     <lastmod>2008-12-23</lastmod>
     <changefreq>weekly</changefreq>
</url>
</urlset>
```

Once You Have a Sitemap, What Do You Do with It?

❖ Transfer it to your web host
- ➢ Not part of your navigation menu
- ➢ Not meant for humans to see
- ➢ Sitemap is supposed to help Bots

❖ Submit its URL to the Google Search Console

❖ Also, include the URL in your **robots.txt** file

How do you tell a search engine NOT to look at your page? (using **robots.txt**)

❖ The robots.txt file allows you to give direction to search engines (bots) by specifying folders or files that should not be indexed.

❖ What kinds of files should might those be?
- ➢ Pages under construction
- ➢ test pages
- ➢ Forms

Robots.txt Basics

❖ A text file named "robots.txt"

❖ It has to be placed at the top-level folder.

 ➤ Found at the address **http://mysite.com/robots.txt**

  (substitute your domain name for mysite.com)

❖ Syntax:

Word in robots.txt file	meaning
Disallow:	Indicates what folders or files search engines should ignore
# at the beginning of a line	A comment
User-agent:	Type of browser or crawler

Robots.txt Files

❖ Two questions to answer when you want to make Robots.txt:

 ➤ **What** should not be indexed?

 ➤ **Which** search engines should be excluded?

 ➤ An example:

```
User-agent: *
Disallow: /cs103/
Disallow: /blog/
```

❖ Reference: http://www.javascriptkit.com/howto/robots.shtml

❖ List of user agents: http://www.robotstxt.org/db.html

Robots.txt and Sitemaps

❖ robots.txt must include a link to the sitemap:

Sitemap: *url of sitemap*

> User-agent: Googlebot
> Disallow: /cs103/
> Disallow: /blog/
> **Sitemap:** *cs-people.bu.edu/vranjbar/sitemap.xml*

❖ The primary purpose of robots.txt is to tell search engines what NOT to look at.

Three areas in Search Engine Optimization

1) How Do Search Engines Find a Web Page?

2) How Web Pages are indexed?

3) How Web Pages are ranked in a specific topic?

2) How Web Pages are indexed?
(How Do Search Engines Know What a Web Page Is About?)

❖ An index is a list of topics and the pages on which they are found.

❖ How do search engines know what topics your web page covers? In order...

  ➢ <title> tag
  ➢ h1 heading
  ➢ Other headings (h2, h3, etc.)
  ➢ <strong> and <em>
  ➢ Alt attributes of images
  ➢ Your URL, including the file name
  ➢ *Link text (also called "anchor text") on **other web pages** that link to your page*

How Big Are These Indexes?

❖ The website **WorldWideWebSize.com** calculates estimate of the size of Google and Bing's indexes.

❖ As of November 2018:
  ➢ Google: 47 billion web pages
  ➢ Bing: 4.5 billion web pages

More on Link Text for indexing

❖ Search engines use the text that other pages use as link text to determine what the linked webpage is about

❖ Example: If enough web pages contain the text

   **best school ever**\

❖ It may lead BU to come up in search results for "best school ever"

❖ Attempts to manipulate indexing in this way are called "**Google bombs**" and are considered "**Unethical**" indexing techniques.

"Unethical" Indexing Techniques

➢ To fool web search engines, cybercriminals use popular but irrelevant words into pages so that pages would come up in unrelated searches.

➢ These words will be placed into pages by hidden ways

  ❖miniscule text font

  ❖white text on a white background

  ❖misleading **meta description tag**

Meta Description tags to help search engines?

❖ In the early years of the web, the <meta> tags for "description" were used by search engines for indexing

❖ Because of widespread abuse, search engines no longer use these tags for indexing

❖ However …

How Search Engines Present a Site in Search Results?
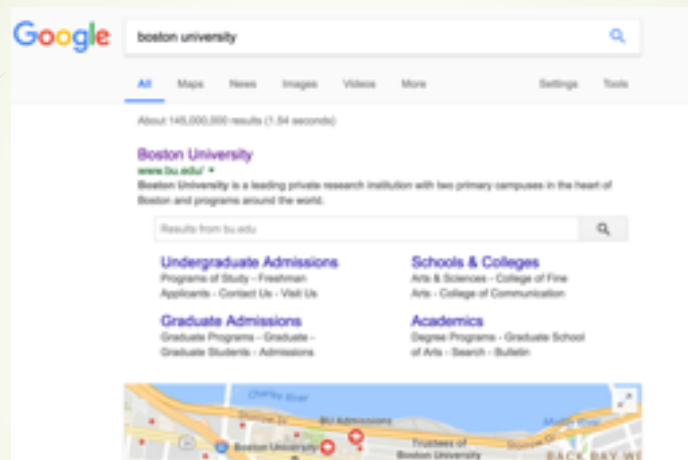
❖ Google's search result page includes both a **title** and a **description**.
❖ The title comes from the **<title>** element of the web page.
❖ The description come from **Meta Description Tag**.
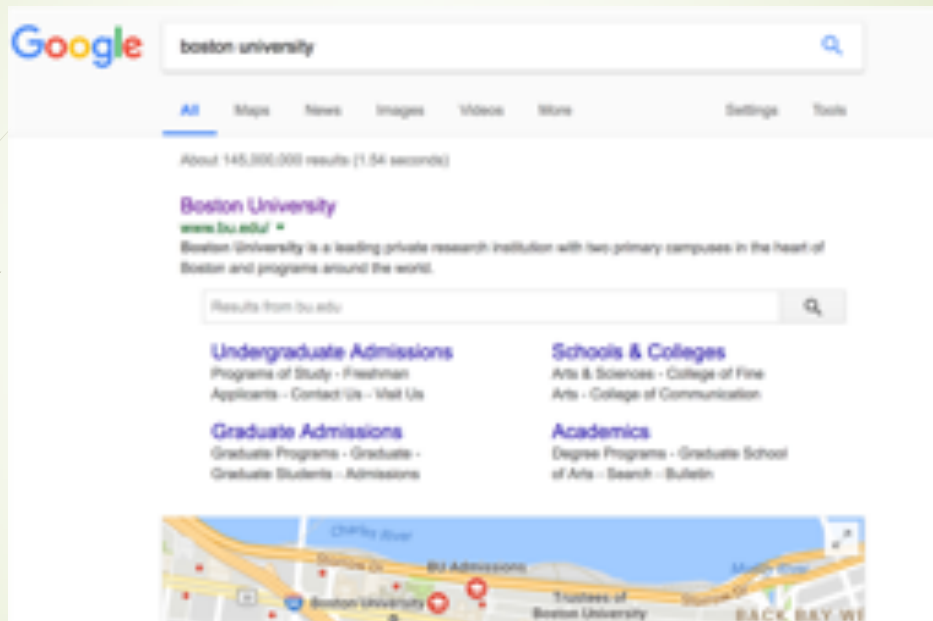
The Meta Description Tag

<meta name="description" content="description of your website here" >

Example from www.bu.edu:

<meta name="description" content="Boston University is a leading private research institution with two primary campuses in the heart of Boston and programs around the world." />

Three areas in Search Engine Optimization

1) How Do Search Engines Find a Web Page?

2) How Web Pages are indexed?

3) How Web Pages are ranked in a specific topic (index)?

---

3) How Web Pages are ranked in a specific topic?
(What Determines the Order of Search Results?)

- ❖ Nowadays, more than 200 factors are considered to rank the web pages.
- ❖ Most of these factors are trade secrets to prevent a business obtaining an economic advantage over competitors or customers
- ❖ The algorithm that first put Google on the map, and made search useful was **PageRank™**
- ❖ Named for **Larry Page**, co-founder of Google.
- ❖ PageRank™ is still used in deciding the order of search results, along with many (200 or more) other factors.

What Other Factors?

❖ If your website supports secure connections (SSL), it will be ranked higher:

https://security.googleblog.com/2014/08/https-as-ranking-signal_6.html

❖ If your site has a responsive design, it will be ranked higher in mobile searches:

https://support.google.com/adsense/answer/6196932?hl=en

See Google's guide to search engine optimization:

http://static.googleusercontent.com/media/www.google.com/en/us/webmasters/docs/search-engine-optimization-starter-guide.pdf

Key Facts about PageRank™

❖ The algorithm that first put Google on the map, and made search useful

❖ **Based on the hyperlink structure of the Internet**

❖ Still used in deciding the order of search results, along with many (200 or more) other factors.

❖ Most other factors are trade secrets
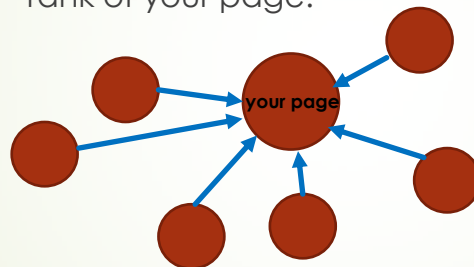
❖ Named for **Larry Page**, co-founder of Google.

3) How Web Pages are ranked in a specific topic?
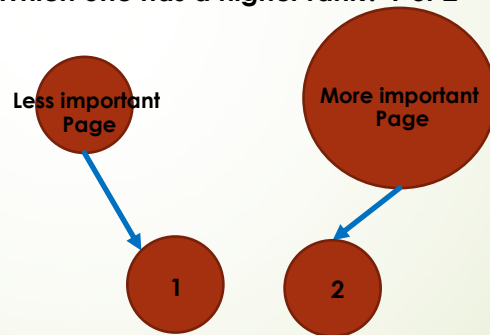**(What Determines the Order of Search Results?)**

Four Ideas of PageRank™ 1

1. The more pages that link to your page, the higher the rank of your page.

Four Ideas of PageRank™ 2

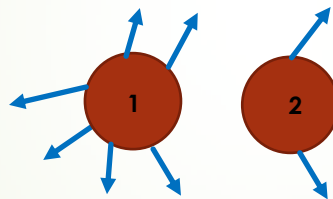2. Links from "important" pages count more than links from less important pages.

Which one has a higher rank? 1 or 2

Less important Page

More important Page

Answer: 2

1

2

Four Ideas of PageRank™ 3

3. Links from pages with few links count more powerful than links from pages with many links.

1

2

3

Getting a link from which one helps you more if the importance of pages 1 and 2 is the same?

Which one has more powerful hyperlinks? 1 or 2

Answer: 2

Four Ideas of PageRank™ 4

4. Hits on a web page does not come from someone clicking on a link on a web page.

Pulling It All Together

❖ This formula shows how the four ideas translate into the PageRank™ algorithm.

The PageRank™ Formula

For each node $x \in V$:

$$PR_i(x) = (1-d) + d \times \sum_{y \in InN(x)} \left(\frac{PR_{(i-1)}(y)}{Outdegree(y)} \right)$$

PR = PageRank™
x = a web page
V = the World Wide Web
d = the damping factor (see next page)
InN(x) = the set of pages that have links to page x
Outdegree(y) = the number of links leading from page y

The PageRank™ Formula

➢ **Damping Factor:** if Google notice that there is an unusual improve in the ranking of a webpage, it will increase the damping factor for that webpage to decrease the improvement.

❖ One reason can be preventing unethical SEO

Ethical vs Unethical SEO

❖ Search engine optimization techniques are broadly classified as "white hat" (good, ethical) and "black hat" (frowned upon, likely to be punished by Google).

"Black Hat" Manipulation of PageRank (Unethical SEO)

❖ In fall 2010, department store JC Penney hired an unethical SEO firm to boost their search engine rankings

❖ Links to JC Penney pages were embedded in hundreds of low-quality, unrelated websites (for a fee)

❖ JC Penney ranked #1 in dozens of categories for several months

❖ Google took a manual corrective action, and ratings fell

—"The Dirty Li[le Secrets of Search"
David Segal, *New York Times*, February 12, 2011 http://nyti.ms/1AlXm7g

"White Hat" Techniques to Improve Page Rank

❖ Produce high quality content so that people will want to link to your website.

❖ Market your website so that more people will find out about it and link to it.

❖ Avoid "black hat" techniques such as paying for links.

auditory disabilities

❖ What barriers do people who are deaf face in accessing the web?
  ➢ Audio and Video Contents

❖ What should web developers provide to make that content accessible?
  ➢ Captions and Transcripts

Captions and Transcripts

❖ **Captions** are "text versions of the spoken word presented within multimedia."

> --WebAIM: Captions Transcripts, and Audio Descriptions

❖ **Transcripts** are textual versions of the content presented outside multimedia (usually on a separate page or document)

❖ Example:

> ➢ https://www.youtube.com/watch?v=6Xg3MeRXSI4

❖ For audio-only content, a transcript is sufficient. For video, both captions and transcripts are needed.

About the Final Project

Credits Page

❖ As a requirement for your final project website, you must make a separate page called credits page and transfer all of the credits to that page.

Changing the WordPress URL

❖ Go to the final project folder on BlackBoard and follow the instruction.

The Footer and Copy Right Statement

❖ People can see the following items in the footer of pages of your website:

1. Copy right statements for your website:
   ©2019 Vahid Azadeh Ranjbar. All Rights Reserved.
2. Social media icons
3. A hyperlink to your credits page
4. …

For your information:

https://www.orbitmedia.com/blog/website-footer-design-best-practices/

Other Project Questions?

One on One Meetings

➢ Do not forget to take a look at BB for your one on one meeting.

➢ 10 minutes is a very short meeting. Make sure that, you have something to show the staffs before going to the meeting.

Interactive features

➢ Take advantage of Show and Tell Day presentations for interactive features of your final project website.

✓ They are posted on BB