PLTW COMPUTER SCIENCE

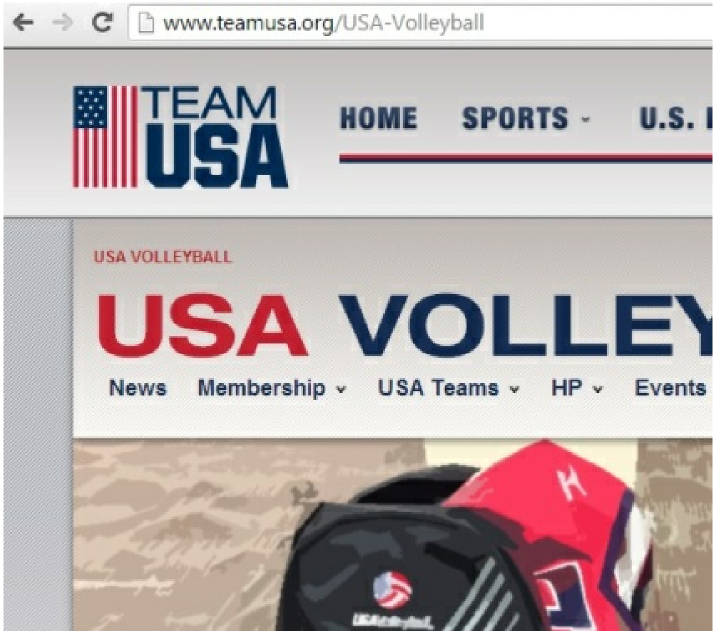
Activity 2.1.2

Your Favorite Web Page: Websites, Browsers, and Search Engines

Introduction

What is your favorite web page? Why do you like it so much? There are a number of factors that contribute to the creation of an effective website. What are they?

Even if you have created a wonderful website, how are people going to find it or even find out about it?



Materials

- Writing utensil
- Computer with Internet access and Firefox with Firebug installed

Resources

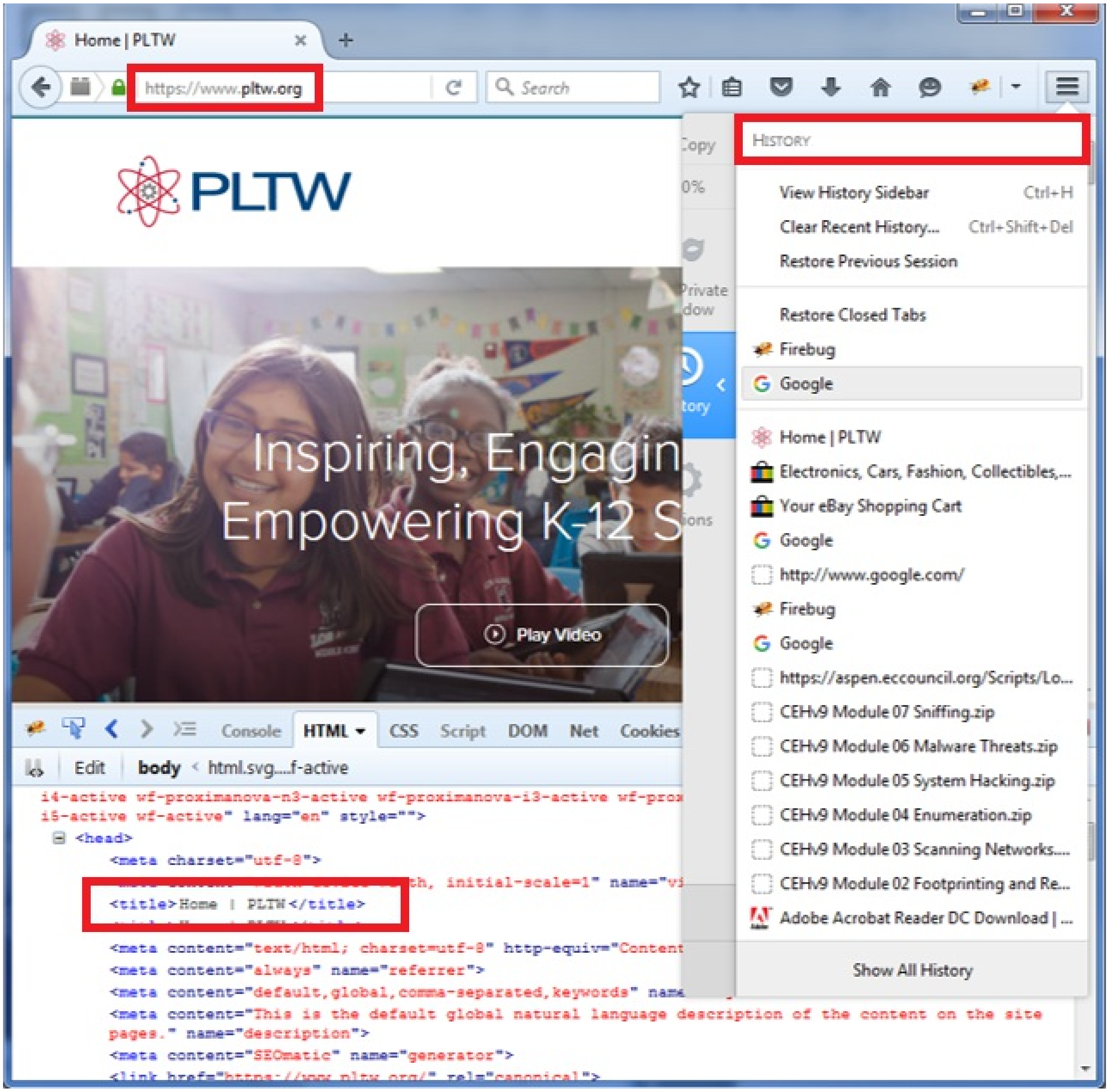
2.1.2 Parsing URLS.docx

Procedure

Part I: Your Favorite Website

1. Form groups of three or four as directed by your teacher. Meet or greet each of the other members of your group, practicing your personal and professional skills.
2. The “name” of a web page might be a URL, a filename, or `<title>` elements that are used by broswer tabsand bookmarks. When you bookmark a web page, it shows up in your

brower's history. The content of a web page can also suggest a name for that page. Write down a name for your favorite website here.

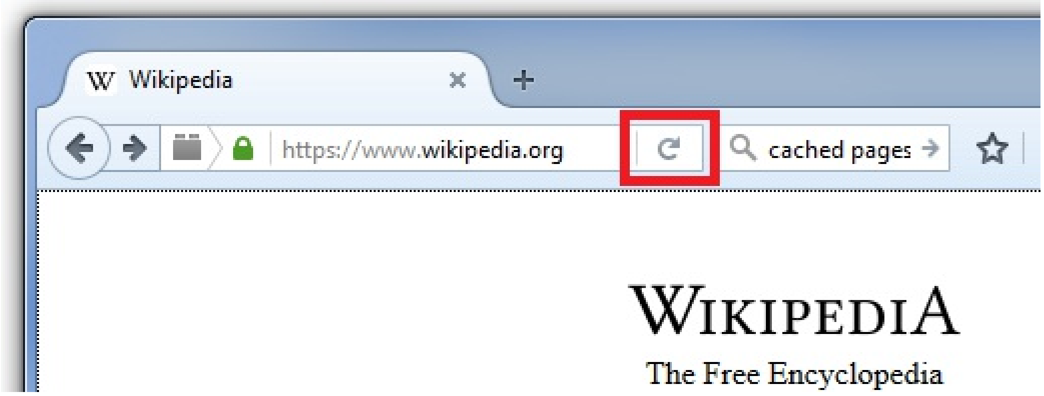

Home | PLTW
https://www.pltw.org
Search
PLTW
Inspiring, Engagin
Empowering K-12 S
Play Video
History
View History Sidebar Ctrl+H
Clear Recent History... Ctrl+Shift+Del
Restore Previous Session
Restore Closed Tabs
Firebug
Google
Home | PLTW
Electronics, Cars, Fashion, Collectibles,...
Your eBay Shopping Cart
Google
http://www.google.com/
Firebug
Google
https://aspen.eccouncil.org/Scripts/Lo...
CEHv9 Module 07 Sniffing.zip
CEHv9 Module 06 Malware Threats.zip
CEHv9 Module 05 System Hacking.zip
CEHv9 Module 04 Enumeration.zip
CEHv9 Module 03 Scanning Networks....
CEHv9 Module 02 Footprinting and Re...
Adobe Acrobat Reader DC Download | ...
Show All History
Console
HTML
CSS
Script
DOM
Net
Cookies
Edit
body < html.svg....f-active

```
i4-active wf-proximanova-n3-active wf-proximanova-i3-active wf-prox
i5-active wf-active" lang="en" style="">
⊟ <head>
    <meta charset="utf-8">
                           th, initial-scale=1" name="vi
    <title>Home | PLTW</title>

    <meta content="text/html; charset=utf-8" http-equiv="Content
    <meta content="always" name="referrer">
    <meta content="default,global,comma-separated,keywords" name
    <meta content="This is the default global natural language description of the content on the site
    pages." name="description">
    <meta content="SEOmatic" name="generator">
    <link href="https://www.pltw.org/" rel="canonical">
```

3. Load your favorite website in the browser. The browser, like Firefox or Chrome, is a client application running on the computer at which you are working, which is the client machine. The browser exchanges TCP/IP packets with the server to open a connection, request the web page, and get the content. Most packets on the Internet are TCP/IP packets and follow two protocols.

   If a user has visitied that website from the computer you are now using, you might be seeing a cached page. A cached page is a page previously loaded from the web server and saved on the local client so that it can be viewed again quickly without getting a fresh copy from the server. You can ensure that the browser loads a fresh page by using the Refresh button

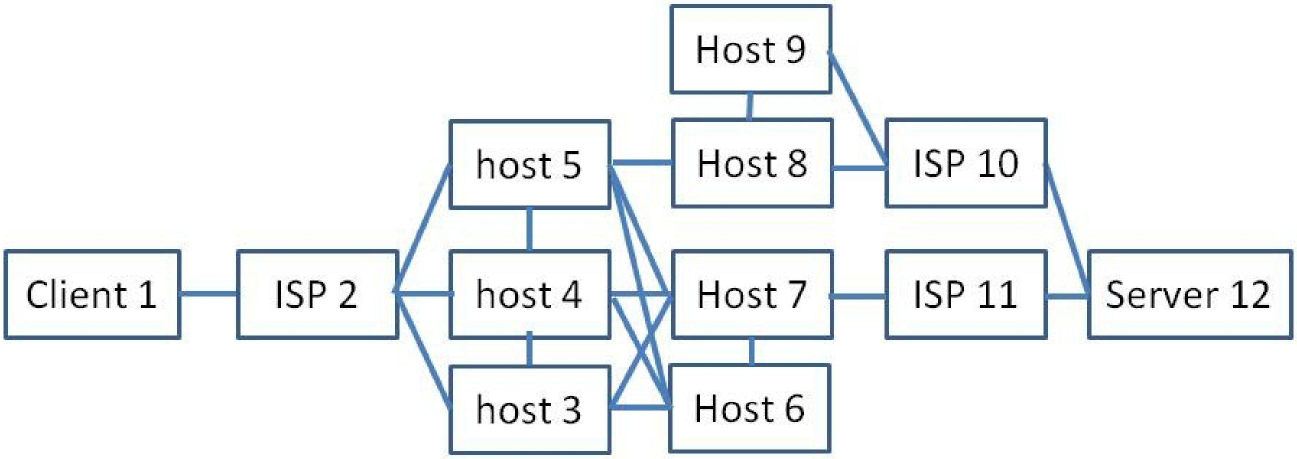


Does the website load all at once or in pieces?

4. Think back to your experience sending packets across the classroom. Why is it advantageous for a client and server to exchange information in packets? What is a disadvantage of using packets?
5. The web page that was just transmitted to you was sent using packets. Why would this page load all at once when cached as opposed to in pieces when you hit "refresh" ?
6. Evaluate the website for usability using the principles of HCI discussed in the last lesson.
   - Structure: The interface should be organized, putting related elements together.
   - Simple: Common tasks should be easy.
   - Visible: Information and options should be easy to find, without the distraction of unnecessary information.
   - Feedback: User should be informed of actions, changes in state, and errors.
   - Tolerance: Mistakes should be easy to undo and reasonable input should be interpreted.
   - Reuse: Design should be consistent across components.

7. For what audience was this site intended and how effective is it at tailoring to that audience? What makes it effective or ineffective?
8. How reliable, believable, or authoritative is the information presented on this site and how do you know?
9. Consider how you would access this site if you were color blind, vision impaired, blind, or deaf. How much would it change your experience? In what ways and to what extent?
10. Take turns presenting to the group about the website you, the presenter, analyzed. Direct your web browser to each person's site in the group before that person presents their site. Each presentation should be two minutes and address each of the following qualities: audience, reliability, and accessibility.
11. Choose one of the other websites that was shared in your group and compare and contrast its qualities with the qualities that yours exhibits.
12. Web pages, like other information on the web, are sent from server to client as packets. Thinking back to the last activity, would you still be able to view a web page hosted on the left side of the room from the right side of the room if one of the routers in the middle went out of service?
13. The physical structure of the Internet has a lot of redundancy, meaning that multiple systems that accomplish the same thing.

    The diagram below includes a client (host #1) that is being sent a page from the server (host #12). If host #2 goes down, the packets will not be successfully transported. Similarly, if BOTH host #11 AND the link from #10 to #12 go down, packets will not be transported.



14. Why do you think that the routes traveled by packets on the Internet are usually least redundant near the beginning and end of their journey?

Part II: Request and Response

When you enter a web address into your browser, three sets of exchanges occur:

- Your computer uses domain name servers to get the IP address of the web server.
- Your computer exchanges SYN-ACKs with the webserver to open a connection(more on that in a later activity).
- Your computer sends an http request. Then the webserver sends an http response and closes the connection.

Understanding request and response can help you find what you're looking for, analyze content for reliability, and protect the security and privacy of your information.

The URL (Uniform Resource Locator) of a website is how you direct your browser to a specific site. Shown below is an example of a URL within the address bar of a web browser.

A URL starts with a scheme such as file or http. The scheme is usually a protocol for exchanging information such as http, https, or ftp, followed by a colon and two slashes "://". After the scheme or protocol, the URL has a domain name or IP address, then a slash, and finally a full path to the resource that we want to view in the browser. The supplemental resource *2.1.2a ParsingURLs* contains a table to help you understand the parts of a URL and provides rows for you to diagram other URLs. These examples are referenced in the following paragraphs.

In the first example, the protocol is `https` and the domain is `mail.example.com`. The path to the resource is news/current/index.html.

After the question mark is a set of parameters passed to the website by your browser. Parameters are a lot like arguments in *Python*® programming language. The list of parameters begins with the question mark, and each parameter is separated from the others by an ampersand (&). For example, the parameter x in example one is assigned the value 35.

Some URLs, like the second example, include a colon followed by a number between the host name and the slash preceding the resource path. This number indicates the **port** on the server that is being accessed. Port numbers are used by software and hardware to **filter** packets. A filter only lets some packets through. Unencrypted web traffic typically uses port 80, and the the :80 can be omitted from the URL.

15. Use the resource worksheet *2.1.2a ParsingURLs* to parse and diagram URLs. As described in the previous step, two URLs that have been diagrammed in the worksheet. Use the remaining rows to parse and diagram the following URL along with any others your teacher provides.

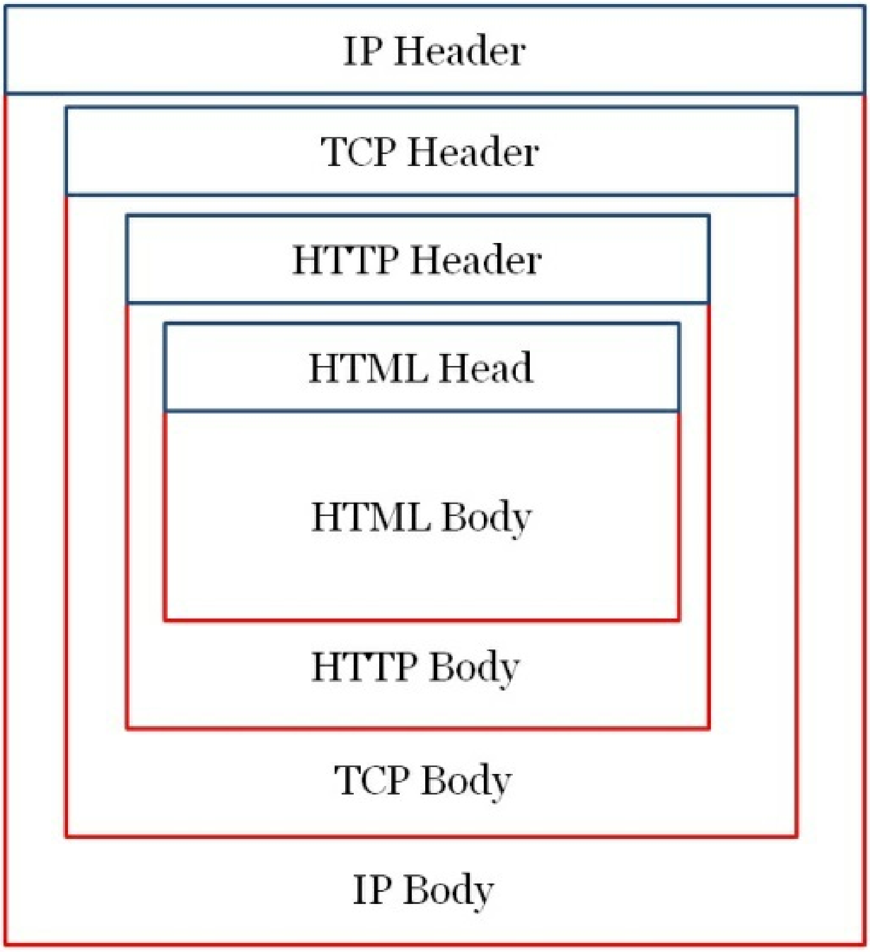
```
http://samplehs.pltwcs.org/students/bkiGag3/sample.php?f=2
```

16. Consider the following URL. Match the components of the URL in the left column with the component names in the right column.

```
https://free.coolsite.com:443/index/page1.html?x=1&y=3
```

https	a. Port
free.cool.site	b. path
443	c. parameter #1 key
index/	d. parameter #1 value
page1.html	e. parameter #2 key
x	f. parameter #2 value
1	g. resource name
y	h. host name
3	i. scheme or protocol

17. It is likely that you use http as your connection protocol most of the time. This protocol specifies what information the client and server will place in the body of the TCP packet and how it will be formatted. The TCP packet is in turn contained within the body of an IP packet and the HTTP packet contains as its payload the actual HTML code for a website. The image shown below is a diagram of this nested packet structure.

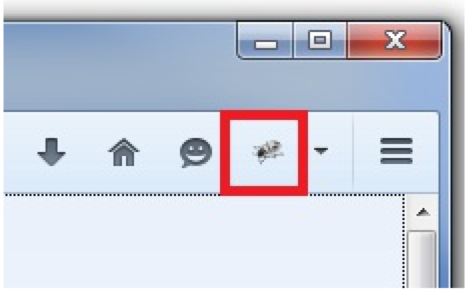


Why do you think it might take many packets to assemble a web page within your browser?

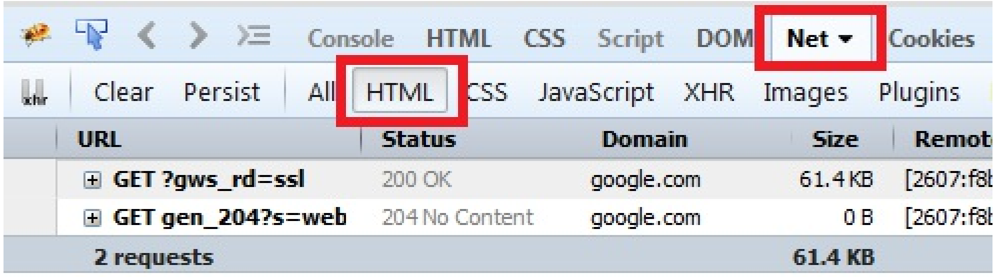
Part III. Your Digital Footprint

What enduring bits are left behind by your activities on the Web? Requesting a web page by its domain name causes chatter among name servers. What other network traffic is generated when you browse the Web? In this section, we look at cookies. What information about you gets stored, and by whom? This question reaches into issues of democracy, power, privacy, and more. In this section we consider cookies, a way that information about you is stored by web servers on your own machine.

18. A cookie is a small amount of data stored on the client machine. In order to view your cookies for one web page, turn on Firebug in Firefox by clicking its icon in the navigation bar shown below. Firebug is not built into Firefox; it is an Add-on. Add-ons are separate pieces of software that interact with another to extend its functionality. In this activity Firebug serves to allow us to examine websites in greater detail and provides some additional functionality beyond what is available in traditional browser-based tools like the “View Source” option that we will use later in this activity.



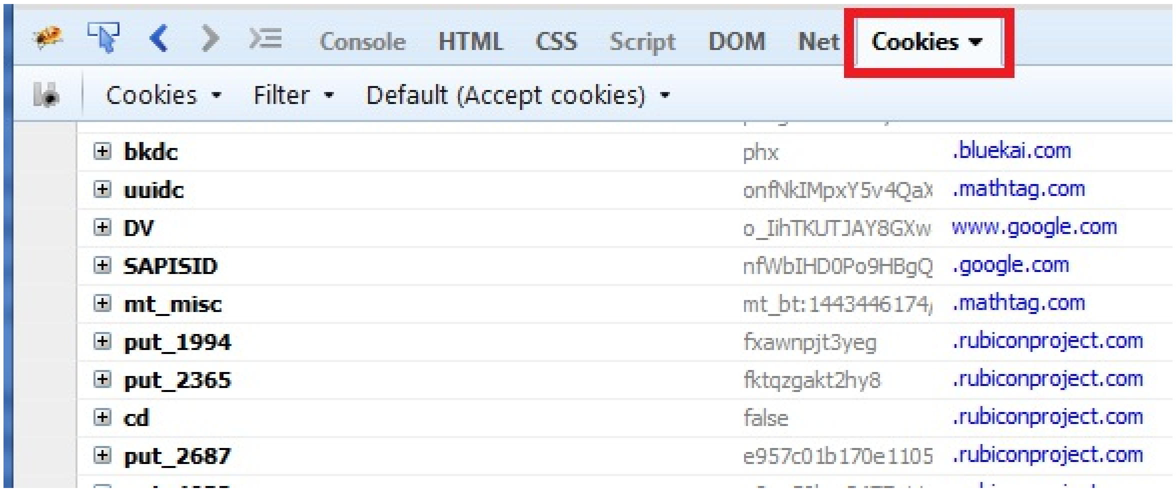
Direct your browser to `google.com`. Switch to the “Net” tab in Firebug and then select “HTML” as shown below.



URL	Status	Domain	Size	Remot
⊞ GET ?gws_rd=ssl	200 OK	google.com	61.4 KB	[2607:f8t
⊞ GET gen_204?s=web	204 No Content	google.com	0 B	[2607:f8t
2 requests			61.4 KB	

Now direct your browser to `ebay.com` and search for something you might be interested in buying. Using the Firebug window, notice the number of cookies and their domains. Much of the HTML that the browser shows is provided by sources other than `ebay.com`. Name two of these other domains.

19. Switch to the “Cookies” tab in Firebug as shown below.



⊞ bkdc	phx	.bluekai.com
⊞ uuidc	onfNkIMpxY5v4QaX	.mathtag.com
⊞ DV	o_IihTKUTJAY8GXw	www.google.com
⊞ SAPISID	nfWbIHD0Po9HBgQ	.google.com
⊞ mt_misc	mt_bt:1443446174,	.mathtag.com
⊞ put_1994	fxawnpjt3yeg	.rubiconproject.com
⊞ put_2365	fktqzgakt2hy8	.rubiconproject.com
⊞ cd	false	.rubiconproject.com
⊞ put_2687	e957c01b170e1105	.rubiconproject.com

Web programmers use cookies to do things like remember your preferences, history, shopping cart, or login status. Each set of cookies is associated with a domain name and with a directory path on that host. So a cookie can be specific to one web page or can be shared across a web site. Cookies cannot be accessed by a server outside their domain, so a page at `mail.google.com` can access a cookie associated with `google.com` but not one associated with `ebay.com.` The cookies tab in Firebug lets you view the details of each cookie created or accessed when you visit a site. It provides the following information:

- Cookie Name

- The value of the cookie, which is the actual information being accessed by the site
- The domain that is accessing the cookie
- The raw size of the cookie
- The date and time it expires
- A flag called "httponly" which improves security by limiting access to cookie data
- A security flag to indicate the cookie data must protected by using a secret code (encrypted)

Choose one of the cookies that you can see and record its name, domain, size, and expiration date/time.

cookie name	domain	size	expires

20. Now search for something on ebay and examine how the number and variety of cookies changes. Before your search, most if not all of the cookies should have belonged to the ebay domain. To whom do they belong now? List several different domains.

21. These cookies being accessed by domains other than `ebay.com` are called **third-party cookies**. Use the Internet to find out more about the domains that are accessing your cookies when you searched for something on ebay. Who are the entities using these domains and what do they do? Choose at least one to examine at this level of detail and then share what you found with 3-5 other students.

Part IV. What Web Pages are Made Of

22. Web pages are written in Hypertext Markup Language (**HTML**). The HTML can include other languages, like Cascading Style Sheets (**CSS**) and **JavaScript**. You will learn a more about these langauges in a later activity.

    Web pages are rendered by the client machine's browser. Different browsers render the page differently. They are configurable by the user. So neither the server nor the client can know how a web page should be displayed without information from the other. Look at the source code that was sent to the client to display this web page. Navigate back to ebay.com if you've left it. Select **Firefox** > **WebDeveloper** > **Page Source** (Ctrl-U). Firefox should create a new window filled with text. The text is the source code for the page. Use the Firefox “Find” tool (Ctrl-F) to search the text for word “meta”. How many lines do you find that contain this word?

23. Switch from the Cookies tab to the HTML tab in Firebug and click on the + sign in the box next to the characters “<head>” in order to view the scripts and metadata for this site as shown below.

Refer to your downloadable resources for this material. Interactive content may not be available in the PDF edition of this course.

This displays the source code of the web page. For now we are interested only in the lines beginning with “<meta ...>” . Look at each of these lines to find the one that contains a list of key words separated by commas. These key words used to be used heavily by search engines to help determine their search results. Watch the following video and describe why Google now chooses not to use these key words in determining their search results: **http://www.youtube.com/watch?v=jK7IPbnmvVU**

24. Web site designers can use many strategies to raise a web page's ranking in search engine results. Submit “search engine optimization” as a web search query to your favorite search engine. Examine the domain name of each search result's link. Note whether the search engine reports some links as *sponsored links* which are paid advertising. Skim the Wikipedia page to familiarize yourself with *search engine optimization*. Describe how companies can influence search results.
25. Navigate to your favorite web page and use Firebug to examine it. Did you find any unexpected third-party cookies. If so, for what domains?
26. We will now use Firebug to revisit the question of accessibility for your website. One important design concern when developing a website is the ability of text-to-speech readers to process images. To this end, `img` tags in HTML have an “alt” property which can be used to provide alternate text to be displayed in the event that the image cannot be loaded or the user cannot see the image. On your favorite webpage (or another if yours has no images), use the search feature on the Firebug/ HTML panel show below. Enter "img" to find an image on your web page. Mouse over parts of the HTML until the image appears highlighted in the browser. Observe the `img` tag to see if it has an `alt` field with the alternate text. What site did you visit and what was the alternate text on the image, if any? Was the alt text sufficient to give a user a good idea of what the image is?

Refer to your downloadable resources for this material. Interactive content may not be available in the PDF edition of this course.

Part V: Create Search Queries

Search engines are one of the most powerful ways to quickly access the data you want on the web. Search engines typically use **web crawlers** or spiders, autonomous softbots that examine the pages on the web by passing from link to link, aggregating all of the information that they discover into a massive database called a **web index**. Search engines then use their own algorithms to determine relevance of the sites that they have indexed to a given query and return results back to the user. The words that you type into the search field of a search engine form the **query**.

27. Use this information and optionally the video found at **http://www.google.com/intl/en_us/insidesearch/howsearchworks/algorithms.html** to answer the question “What role do algorithms play in determining search results?” .
28. After watching the following video, use the information you gained to try and answer why some web pages might not be returned as search results even if they are relevant to your query. **http://www.youtube.com/watch?v=BNHR6IQJGZs**
29. Even today one of the most important skills you can learn is how to effectively format your search queries. Higher quality queries get you better results in less time, leaving you more time for everything else. They help you conduct research, find images or videos, and locate

communities for discussion of various topics. Some of the best tips for improving your searches are incredibly basic: be as specific as you can, consider the type of vocabulary that is likely to appear on sites that you want to see, and keep your queries simple. You can get a lot trickier if you have specific criteria for your search. Use this page provided by Google to answer the following questions:
http://www.google.com/intl/en_us/insidesearch/tipstricks/all.html

- You want to search for Mustangs but not the car, just the horse. What search operator would you use to ignore sites about cars?
- You want to find PowerPoint slides about a given topic. How can you get results that contain only .ppt compatible documents?
- You saw a really interesting article about whales on cnn.com, but it's been a few days and you can't find it by searching for whales anymore. How can you narrow down your search and only get results from cnn.com?

30. Search is a powerful tool for many reasons. Analysis of search query data has revealed that by using Google Flu Trends, researchers were able to predict the outbreak of flu cases more than a week in advance of the CDC (Centers for Disease Control). Google's trend search feature shows a visualization of the data that they have collected about what people search for and when. The data is crowd sourced: people all over the world are the ones doing the searching. All Google has to do is record what they searched for, when, and then analyze the data.

    Note: Google limits the number of requests to Google Trends per IP address. Refrain from submitting extras until everyone in your class has finished.

    Navigate to `google.com/trends` and search for “flu”. When did people start really worrying about Swine Flu?

31. What else might you be able to predict using a visualization like Google trends? Conduct a search and note any patterns that you found here.
32. Describe a strategy for using Google trends to buy your best friend the most popular Christmas gift.
33. Google is certainly a popular search engine though by no means the only one. In your group of 3 or 4, brainstorm a total of 5 queries and decide upon 3 search engines to test on those queries. Examine the first 10 results in each engine and record your comments and analysis here.

Engine Names	www.google.com	www.google.cn	www.bing.com
Queries			
1			
2			
3			
4			
5			

34. Which search engine's algorithm do you think performed the best, and why?
35. Compare the image results for the query “democracy” on google.com vs. `google.com/hk`. Governments control or influence network infrastructure and content. Describe your observations.
36. Owners of individual domains get to decide what content is published on their websites. Why might this autonomy be important to the development of the Internet? How or why does autonomy scale?

Conclusion

1. We barely even looked at any code in this activity. Why do you think that knowing about the protocols and features of the web might be useful to you as:

    - An informed citizen:
    - A professional developer:

2. How do various factors influence the results that different search engines produce?
3. How can crowd sourced data about search trends help predict the future?
4. From the standpoint of the governing bodies of .com, why is it important that owners of individual domains maintain authoritative records of their subdomains and manage the content that is published on their sites? (Hint: think about what the alternative would be.)
5. If the most direct route for packets representing a webpage that you have requested from their server to you is broken (hardware goes down for some reason) what happens to the packets?
6. What makes for a high quality website?
7. What is the basic function of the cookie and where is it stored?