

LBSC 690: Subject summary

William Webber

April 27, 2012

1 Week 1: Computers and Data

Digital versus Analogue Distinguish digital and analogue representation. What happens to each on repeated copies? Is there a common analogue representation for all data? Is there a common digital representation for all data? What is the fundamental representation of data in digital system? What is “digital convergence”, and what drives it? What do we mean by a “universal digital computer”?

Binary representation What is the essence of a binary representation? How many distinct states are there in a binary representation? Why this many, not more, not less? Give examples of two-state implementations in the physical world; e.g., how are two states represented on a CD? on an HDD? on a network wire? Be able to convert between binary and decimal; to do addition in binary. What is a logic gate? What do we mean by the “word size” of a computer? What is driving the move from 32-bit to 64-bit servers and (increasingly) desktops? What does it mean to say that computers use a “fixed-width” binary representation of numbers? What happens if we have an 8-bit wide binary number 11111111, and we add 1 to it? Be able to convert between binary and hexadecimal (see Week 2).

Size and speed How many bytes is a megabyte? gigabyte? terabyte? Roughly what power of 2 is one thousand? one million? one billion? If a link transfers 10Mbps, how many bytes does it send in a minute? Be able to calculate download time for an object, given object size and network speed. When we say that a CPU runs a “2GHz”, what does that mean? What does Moore’s law state? What is the “memory mountain” and why is it becoming a more serious problem? Why does buying more RAM for your computer sometimes make it faster, sometimes not?

2 Week 2: Networking and the Internet

Ethernet What fundamental logical elements are we sending across a network medium when we do (digital) network? What is modulation and demodulation? What type of network is ethernet (size, transport medium) used for? How are machines addressed in Ethernet? How does an ethernet frame make it from the sending machine to the

destination machine? What stops one machine on an ethernet network from listening to another machine's frames? What about on a wireless network? How can we use one redundant bit to detect a one-bit error?

Internet protocol (IP) What is the purpose of the internet protocol? How are machines addressed in IP (v4)? How many bits are used in IPv4 addresses? How many distinct addresses are there? Is this enough? What is the main difference between IPv4 and IPv6, and why are we (slowly) moving to IPv6? What is meant by a "private" IP address? What is meant by the joke t-shirt "There's no place like 127.0.0.1"? Do IP addresses have to be unique on the internet? How does a packet find its way from one machine to another? When an IP packet has to go across an ethernet network, how is this done?

Transport control protocol (TCP) What functionality does TCP add on top of IP? What are TCP ports and port numbers for?

Hyper-text transfer protocol (HTTP) What does it mean to say that HTTP is an "application-level protocol"? What applications speak HTTP to each other? What is a URL, and what are its different components?

The protocol stack What is meant by a "protocol stack" and why do we need one (rather than, say, one universal protocol)? What is the purpose of each level of five-level protocol stack?

Domain name service (DNS) What does DNS do? Are all domain names held in a single server? If not, how does a computer look up a domain name? What are the parts of a domain name? What does a domain name say about a website that is hosted at the domain name? Distinguish between physical location, IP address, and domain name of a computer. How is legal jurisdiction over a website determined?

Tools What does traceroute do? whois? DNS lookup? How can you find out who is responsible for a web site?

3 Week 3: Structured documents and (X)HTML

Hyper-text markup language (HTML): concepts Does HTML specify precise layout and appearance? nothing about layout or appearance? partial information about layout and appearance? (Consider what the browser does when we resize the window.) What is the basic syntax of HTML?

HTML: details What are the two top-level components of an HTML document? Be familiar with meaning, usage of `<html>`, `<head>`, `<title>` `<body>`, `<p>`, `
`, `<h1>`, ``, ``, ``, ``, `<i>`. Be familiar with HTML tables,

links, images. Be able to produce a well-formed HTML page, from scratch, that uses the above elements.

Beyond XML What is a “markup language”? What does WYSIWYG stand for? How does editing an HTML page in source view differ from editing in WYSIWYG view? What is the difference between logical, structural, and presentation markup? What layer would you say HTML mostly fits at? What is XML? What is XHTML, and how does it differ from HTML?

4 Week 4: Separating content, structure, and style

Motivation What are the advantages of separating content, structure, and style of each web page in a multi-page website? (Think of personnel; maintenance; consistency.) What technological means can we use to achieve this separation?

CSS: concepts How does CSS help in the separation of tasks described above? How can you best use CSS to apply one style across a whole site? How does CSS help deploy a website on multiple media? How are the `id` and `class` attributes used for CSS, and how do they differ from each other? What is the difference between “logical” and “stylistic” class or id names, and why is the former preferable to the latter? What are the `<div>` and `` tags, and why are they particular used with CSS? What is the basic CSS syntax?

CSS: details Be familiar with `background`, `color`, `font`, `font-style`, `font-family`, `margin`, `padding`, `border-width`, `border-style` properties. Be able to create and apply an external style sheet to style an HTML page.

Template of layout Why might you want to separate the layout and content of a multi-page web site? What are server-side includes, and how can you use them to separate layout and template? Why are SSI’s called “server-side”? What other techniques can be used to separate layout and content? What is the connection between this separation and accessibility?

5 Week 5: Structured data and databases

Data types What does it mean to say that a bit string has a “type”, and how does this affect our interpretation of it? What are some types that a bit string could have? What is an integer? What is the difference between a signed and an unsigned integer? How are non-integral numeric values represented? What is a “character”?

Character sets What is a character set? What does it map between? What is ASCII, and what language does it cover? Can you think of a character that you might use in written (or typeset) English that is not covered by ASCII? How many bits are used

in ASCII? Why and how to Latin character sets extend ASCII? Give an example of a character that might be in a Latin character set but not in ASCII? Why may a document written in one Latin character set not display properly in an application configured for another Latin character set (or for ASCII only)? How do language-specific Asian character sets extend ASCII? What is, and what is the goal of, UNICODE? What is the distinction between UNICODE as a character set and the different encodings of the character set? What is UTF-8? Why might an application support UNICODE and yet some characters in a UNICODE document not display properly on the screen?

Database representation What is structured data? What is the simple form of structured data that a single relational database table represents (and what form of structured data is a single relational database table not able to represent)? What is an “entity”? How does a database table, its rows and columns, relate to “records” and “fields”, and how do these relate to “entities”, “properties”, and “values”? What is the “schema” of a database table?

Linking entities What is an index in a database, and why would you want them? What is a key? What is a primary key? How does binary search work? How do (and why should) you separate compound or repeated information into separate tables? How do you link records in different tables together?

Database design Give examples of a one-to-one, a one-to-many, and a many-to-many relation. Why can’t many-to-many relations be directly represented in relational databases? How do we represent them instead? Be able to draw a database design as an entity-relationship diagram (ERD), with cardinality of relations, primary and foreign keys, and fields with names and types.

XML data What are the advantages and disadvantages of XML as a data representation (compared to a relational representation)? Why is XML particularly useful for data transfer and exchange?

6 Week 6: RDBMS and SQL

Relational database management system (RDBMS) What are the functions of an RDBMS? What is the distinction (often ignored) between an RDBMS and a database? What is the difference between an embedded, a desktop, and a server DBMS? What are the particular features of a server DBMS? Where does it run? What is phpMyAdmin, and how is it distinct from MySQL? Describe the DBMS place in a three-tier web application.

Structured Query Language (SQL) What is the function of SQL? Is it intended for use by end-users? administrators? programmers? Distinguish between the data definition, data manipulation, and data query parts of SQL. What is a databases’s “data

dictionary”? Be able to write simple SELECT queries, given a data schema. Be able to join tables via primary key to foreign key link using SELECT queries.

7 Week 7: Programming and Javascript

Programming languages What language does the computer speak? Why don't humans just write their programs in this language? What is the function of a higher-level language (HLL)? What is the difference between an interpreted and a compiled language?

Javascript: concepts Where and for what purpose is Javascript predominantly used? Why are browsers able to run Javascript programs? Be able to embed a Javascript program in an HTML page.

Javascript: details Be familiar with `textttdocument.write()`, `alert()`, `prompt()`. Be able to understand, modify, or write a simple Javascript program, similar to ones given in slides.

8 Week 8: Developing Web Applications

Javascript in the browser What is event-driven programming? Why do we use such programming on the browser? What are some user events that we might capture? How do we capture these events in HTML? What is the document object model (DOM)? How can we use the DOM in Javascript to get access to a particular HTML element? How do we mark an HTML element to make it easy to access from Javascript? Be able to understand, modify, write simple Javascript code that responded to user action and updated page, similar to ones shown in notes or given in homework.

Three-tier architecture of a webapp What are the three tiers of a three-tier webapp architecture? What technologies might you use to implement each tier? What is the responsibility of the logic tier? Where does each tier sit in a network architecture? (server).

Server-side programming Why do we need server-side programming to implement the logic tier? Why does it sit on the server side? What simple example of a server-side programming language did we see in an earlier lecture? Be able to distinguish in a simple piece of HTML with PHP, Javascript, and ``, distinction between what is executed on the server side, what is executed on the browser, and what is included by the browser. How does PHP transfer data from the data tier to the presentation tier? How does PHP transfer instructions from the presentation tier to the data tier? Be able to read, understand basic thrust of PHP code example similar to those in notes (e.g. “this code is fetching student records for students enrolled before 2000, and outputting them one row per record to the browser); you *won't* be asked to write or modify PHP or to understand it in detail.

Server-backed interactive pages What is the “request–response” model? What is the object that a traditional web request is responded to with? How does AJAX change this? What is the goal of AJAX? Give an example of a web page that uses AJAX, and explain why it would be awkward or impossible to do provide the same functionality without AJAX.

9 Week 9: Blogs, Wikis, and CMSs

Managing content What are the options for managing content on a web site? What is wrong with just editing directly into static HTML? What advantage does using a content management system (CMS) give us? Distinguish between general CMS systems; wikis; and blogs.

CMS proper A CMS separates content from two other components of a web site; what are these two components? What separation of responsibilities does a CMS support? Why does a CMS limit authorial styling freedom? What are the three modes that a CMS might allow a content author to author content in, and what are the advantages of each? What is the difference between an author-side, static-output CMS and a server-side, dynamic-output one? What server-side support do you typically need for a server-side CMS?

Wikis What is distinctive about Wikis as a CMS platform? Is any user of the internet able to edit Wikipedia? What evidence is there for or against the reliability of Wikipedia as an encyclopaedia?

Blogs What is distinctive about a blog as a CMS platform? What are the setup options if you want to run your own blog? What is an RSS “feed” and how does it work?

10 Week 10: Multimedia and Web Integration

Graphics What is the difference between raster and vector graphics? How are images represented in raster graphics? What is color depth? What is a color channel? As well as RGB, some image formats support four channels: RGBA; what does “A” stand for and what does the channel it represents do? How are images represented in vector graphics? How are complex curves represented? Can you take a vector-graphic photo? Why or why not? What are the relative advantages of raster and vector graphics?

Digital representations of images, video, sound How does a digital camera represent an (analogue) image? What do “megapixels” in a camera mean? Does increasing the number of megapixels in a camera always mean increasing the quality of the photos it takes? What are the two dimensions captured in a digital recording of (analogue) sound? What is the (simplest) digital representation of a video? In addition to the three dimensions of a still image (height, width, color depth), what is the fourth dimension

added by video, and how do we measure “fidelity” in this dimension? If you see a monitor rated as doing “1280x800 pixels at 60Hz”, what does the “60Hz” refer to? What happens to the quality of an image, recording, or video if we reduce the digital fidelity of it? Why does continuously increasing the digital fidelity of an image, recording, or video (say, continually doubling the sample rate or frame rate) not continuously improve the quality of the reproduction when viewed or listened to by humans? When might we want higher fidelity than humans can perceive?

Compression Why compress data? What is the basic principle behind (lossless) compression? What is the difference between lossy and lossless compression? Why would you use lossy compression? Why can you get away with lossy compression? What is the basic principle behind lossy compression? What sort of data can and can’t you use lossy compression on?

Website integration of web services Embeddable web services allow server-like capabilities without you having to install anything on your server; how? Using `<iframe>` is the simplest way of embedding; how does this work? What technologies do richer embeds use? What advantages do they offer? How do such rich embeds talk to their server? Give an example of an embeddable web service widget.

11 Week 11: Characterizing and Searching the Web

Boolean search How does Boolean search work? What are the advantages of Boolean search? What are the drawbacks? Why, in particular, is Boolean search not helpful to the user with simple queries and very large datasets?

Relevance ranked search What is the principle behind relevant-ranked search? What do TF and IDF stand for in $tf * idf$, and what is the meaning or use of each component? What is the $tf * idf$ weight of a word in a document, and why do we weight words? State in words what a high weight $tf * idf$ document is. What is the cosine measure? How does it measure similarity between query and document? Why do we want to measure similarity between query and document? The cosine measure measures angle, not linear distance; why?

Text inference How could we measure the similarity between two documents? What is the vocabulary problem that a user with a short query can run into, and how does query expansion try to fix it? Describe how query expansion can be performed using a document ranking. How can user interaction help improve query expansion? How does text classification work?

Web pages and links How do computer scientists commonly model the structure of the web? The structure of the web is sometimes described as a “bow tie”; why? What is a scale free network? Is the web a scale free network? Give some other examples of scale free networks. Why are scale free networks “vulnerable to attack”? What is

a process that leads to the formation of scale-free networks? Why might this process occur on the web?

Web search What is a web crawler, and how is it used by search engines? How can we control what a web crawler does on our web site? What is anchortext, and how is it important in web search? What should we be careful to do when writing our own anchortext? What is inlink count, why could we use it to measure the authority or importance of a web page, and why does it not work well in practice on the open web? How does PageRank work, and why does it improve over inlink count? What is click-through data, and can it be used to improve search quality? Search engines keeps logs of queries people enter and the search results they clicked on; how can such logs be used for spelling correction? for query suggestion?