

San Diego State University
Imperial Valley Campus

MKTG 476
Lars Perner, Instructor

Marketing, Computers, and the Internet
Spring, 2006

SUMMARY OF NEW IN-CLASS INFORMATION FOR FINAL

Web Design Software

Web design programs can be thought of as specialized word processing programs optimized for making web pages. Today, there are several types of standards for displaying web pages. The most commonly used method today continues to be hypertext markup text language (HTML). These files are made from plain text and a simple editor can theoretically be used to design a web page. Software programs such as Word and WordPerfect can save files in HTML format. Programs such as Microsoft FrontPage and Macromedia Dreamweaver are designed specifically for making web pages.

Considerations in design. The web site designer should consider a number of factors in deciding how to design a web site for a given purpose and setting. There are considerable tradeoffs, such as how much to invest in creating the site and whether to emphasize speed or more elegantly designed pages that will take more bandwidth and perform more slowly. Generally, files that use extensive graphics will require more bandwidth as large files need to be retrieved. Complex animations may require large files and may perform less reliably, due in part to problems of compatibility between design programs and browsers.

Fonts. Mono spaced fonts such as Courier devote an equal amount of space to each letter in the alphabet. Proportionally spaced fonts such as Times Roman, Arial, and Helvetica give more space to “longer” letters (e.g., “m” and “w”) and less to shorter letters (e.g., “i” and “l”). *Serif* fonts have sharp edges at various points while *sans serif* fonts lack these sharp edges. Sans serif fonts are often more aesthetically appealing, but it may be more tiring, on paper, to read lengthy text passages written in sans serif. There have been some reports that sans serif may actually be easier to read on screen, but this appears to be a controversial point. Sans serif fonts may be used especially for headlines and the main body of text may be written either in serif or sans serif fonts.

Preserving formatting. The appearance of a web page may differ between browsers and computers. There are subtle differences in how different browsers (and versions of the same browser) show similar information. The size of the monitor and user choices (e.g., whether to display certain information on the left margin of the screen) may also impact what is displayed. Files written in HTML generally require that the font intended be available at the visitor’s computer. If it is not, the visitor’s computer

may substitute another font. Arial and Times Roman are generally available on most computers, but many others are not.

To reduce the likelihood of varying appearances between browsers, it is useful to use standard fonts. In some cases, files may be saved in Adobe Acrobat (PDF) format. Most computers today have an Adobe Acrobat Reader, but these files will generally download more slowly.

File structure. Web site files exist on “server” disk drives much like the hard drives on ordinary computers. (In fact, the server is a computer that has its own directory structure). A universal resource locator (URL) “points” to a particular Internet address (IP address) and files are stored at that address. At the “root” level, there is usually a file called index.htm or index.html. The file may have a slightly different name like “default.htm.” This is usually the file you reach when you type in the web site address. For example, if one enters <http://www.ConsumerPsychologist.com>, one will reach the index file. Other files can be put either in this root directory or in a subdirectory. For example, the path http://www.larsperner.com/past_courses.htm takes the visitor to a file called “past courses” in the root directory. In contrast, <http://www.larsperner.com/mktg476/index.htm> takes the user to the file named “index” in the directory located immediately below the root directory.

The address <http://www.larsperner.com/mktg476/index.htm>, which includes the domain name, is an “absolute” file reference. It alone is sufficient to get the user to the ultimate destination. In contrast, “/mktg476/index.htm” is a “relative” reference that may be used within a given site. This identifies the file as being in the subdirectory “mktg476.” Another file in the “/mktg476” subdirectory would refer to the file as “index.htm.”

Note that many web servers are UNIX or LINUX based. UNIX uses forward slashes to demark directory levels, as opposed to the backward slashes used in Windows.

Web site implementation. To run a web site, it is necessary to have a domain name. It may be possible to use an address that builds on a domain name already owned by an organization or Internet service provider (ISP). Domain names can be registered with various domain name registrars. .com registrations usually cost between \$8-15 per year. Other types of domain registrations—particularly for certain countries—may cost more.

A web site must be hosted. Theoretically, if one has a high speed connection, a site can be hosted on one’s home computer. It is possible to buy hosting for as little as \$50 per year.

Forms. A useful feature in a web site is a “form.” Like a conventional form, a standardized setup allows the user to fill in blanks with information. This information can either be stored by the web site in a database and/or be e-mailed to the web site owner.

Search Engine Optimization

Many Internet users find desired information and sites through search engines such as Google. Research shows that a large proportion of the traffic goes to the first three sites listed, and few people go to sites that appear beyond the first “page” or screen. On Google, the default screen size is ten sites, so being in the top ten is essential.

Because of the importance of search engines, getting a good ranking or coming up early on the list for important keywords is vitally important. Many consultants offer, for large fees, to help improve a site’s ranking.

There are several types of sites that are similar to search engines. Directories involve sites that index information based on human analysis. Yahoo! started out that way, but now most of the information is accessed through search engine features. The Open Directory Project at <http://www.dmoz.org> indexes sites by volunteer human analysts. Some sites contain link collections as part of their sites—e.g., business magazines may have links to business information sites.

Several issues in search engines and directories are important. Some search engines, such as Google, base rankings strictly on merit (although sites are allowed to get preferred paid listings on the right side of the screen). Other search engines allow sites to “bid” to get listed first. Some sites may end up paying as much as a dollar for each surfer who clicks through. If a potential customer is valuable enough, it may be worth paying for enhanced listings. Often, however, it is better to be listed as number two or three since only more serious searchers are likely to go beyond the first site. The first listed site may attract a number of people who click through without much serious inspection of the site.

Some search engines are more specific than others. The goal of Google, Yahoo! and MSN is to contain as many sites as possible. Others may specialize in sites of a specific type to reduce the amount of irrelevant information that may come up.

Search engines often have different types of strategies. Google is very much technology oriented while Yahoo! appears to be more market oriented. Another major goal of Google is speed. Some sites may contain more content of one type than another. For example, AltaVista appears to have more images, as opposed to text pages, indexed.

Search engine rankings. The order in which different sites are listed for a given term is determined by a secret **algorithm** developed by the search engine. An algorithm is a collection of rules put together to identify the most relevant sites. The specific algorithms are highly guarded trade secrets, but most tend to heavily weigh the number of links from other sites to a site and the keywords involved. More credit is given for a link from a highly rated site—thus, having a link from CNN.com would count much more than one from the site of the *Imperial Valley Press*. On any given page, the weight given from a link will depend on the total number of links on that page. Having one of one hundred links will count less than being the only one. One source reports that the weight appears to be proportional so that one

out of one hundred links would carry one percent of the weight of being the sole link, but that may change and/or vary among search engines.

Types of search engines. Some engines, such as Google, are general purpose search engines. Some are specialized. Some are hybrids, containing some directory structure in addition to search engine capabilities. Some “reward” sites such as iwon.com attract people by allowing them to enter a lottery when doing a search. Some sites are aggregator sites—they do not have their own databases but instead combine the results from simultaneous searches on other search engines.

Text optimization. It is important to repeat important words as much as possible subject to credibility. Search engines today are increasingly sophisticated in identifying “spamming” through frivolous repetition of the same words or early use of words that are not relevant to the main content of the site. Words that appear early in the text and on the index page will tend to be weighted more heavily. For some search engines, it may be useful to include common misspellings of a word so that the site will come up when that spelling is used. For aesthetic reasons, many firms may object to having much text on the front page, but text may be put below the graphic elements—e.g., see <http://www.LarsPerner.com> . Some web site owners have attempted to include hidden text so that a search engine would find the desired words while the visitor would see something else. Some web designers, for example, would hide text behind a graphic, make the text in a very small font, and/or make the font color the same, or nearly the same, as the background. Other web site designers have made a “legitimate” site, only to have a command to move the visitor to another site when they go to the searched site. Search engines today are increasingly able to detect this type of abuse, and sites may be penalized as a result.

Early search engines relied heavily on “meta tags” where the web site creator specified what he or she believed to be appropriate keywords, content descriptions, and titles. Because these tags are subject to a lot of abuse, these no longer appear to be significant.

Link optimization. Many web sites engage in “link exchanges”—that is, complementary sites will agree to feature links to each other. It may be useful for a webmaster to ask firms whose content does not compete for a link. Sites should register with the Open Directory Project at <http://www.dmoz.org> since, if a site is classified favorably, this may help rankings.

Domain names. The domain name of a site may be significant in determining ranking. A site at <http://www.Marketing.com> may get a more favorable ranking, with the same information, than a site like <http://www.JoesHosting.com/users/jdoe/>. Google apparently counts a hyphen as a space, so Marketing-Strategy.com would carry both the words marketing an strategy. Using a hyphen is definitely preferable to using an underscore (e.g., Marketing_Strategy.com) or unseparated words (e.g., MarketingStrategy.com).

Security Issues and Privacy

Firms and consumers today face a number of problems:

- **Identity theft.** Fraud and identity theft are not new, but technology has made this a more serious problem today. “Hackers” are able to break into computer systems and retrieve massive amounts of data on individual consumers.
- **Credit card fraud.** Stolen credit card numbers may be used for online transactions before the theft is discovered. Banks today usually assume most of the responsibility for the theft, and it is actually usually safer to order online than to use a credit card in a retail store when clerks actually see the card. Many consumers are not aware of this, however, and many are hesitant to shop online.
- **Loss of privacy.** Some consumers resent attempts by firms to collect information about them even if this if the intention is not dishonest. This is felt by some as an invasion of their private lives.
- **Costs of authentication.** Because of increasing security problems, firms have to invest in increasingly elaborate systems to authenticate transactions. These processes may also slow down the process.

Online information and privacy. A great deal of information about individuals is actually considered “public.” Real estate ownership records, often including the price paid for a property and the size of its mortgage are often publicly available. In the old days when this information had to be accessed manually on paper, this may not have seemed as threatening, but nowadays, it is relatively easy to buy the computerized records. In some states, it may be possible to buy information about driver’s license records and vehicle registrations, and commercial firms will sell reports on individuals indicating current and past residences as well as possible relatives. Some information—such as credit cards—is restricted in principle and consent of the consumer may be required before a third party can access the information.

Online storage of consumer information. Firms often store customer information online. This information may have resulted from online transactions, but information obtained by the firm elsewhere may also be stored online. Customer records may hold information about past purchases and payments and may contain credit card or other numbers. Firms often store information within a network so that this information can be accessed by different individuals within the firm. Networks can often be infiltrated through the Internet, meaning that outside hackers may be able to access the information without authorization.

Certain types of information are especially critical. Social security numbers and birth dates are important pieces of information that may be used for identity theft. Login and password information can be abused, and some individuals may be fearful of having their home addresses made public.

Data interception methods. There are various ways that hackers can get access to confidential information. The “phishing” method usually involves sending an e-mail message to customers indicating that they will need to logon to their accounts to fix some kind of problem. The title part of hyperlink in the message may look quite legitimate, but the actual link may be for some other site. Hyperlinks generally allow

the designer to highlight as a link a description of where it leads—e.g., the hyperlink may say “Further information,” which may be more meaningful than the URL destination. In phishing, this design feature is abused. Note that the hackers do not need for a large percentage of consumers to fall for this ploy—if even one percent responds, this will give access to a great deal of information.

The “pharming” technique is more advanced. Here, a consumer going to a legitimate web site may be redirected elsewhere. This may happen either through a virus infection that alters addresses sent into web browsers or through infiltration of Internet “name servers” that translate domain names into IP addresses. For major web sites, this diversion will likely be discovered quickly and corrected, but a great deal of information be obtained within just a few minutes.

To deal with phishing and pharming, a number of remedies may be used. One involves targeting the illegitimate server with a great deal of bogus information. It may be tempting to use the method of “denial of service attacks”—basically overwhelming the server with repetitive inquiries about the server’s identity—but this is a controversial practice.

Password vulnerabilities. Some users choose relatively obvious passwords—either based on personal information such as their license plates or names of family members. Certain words are also reported to be used frequently. Other users rarely change their passwords. Some systems now require passwords of a certain length or with various characteristics (e.g., the use of non-words, use of both upper and lower case letters, and use of both letters and numbers). Systems may also force users to change passwords periodically.

Servers. Servers are vulnerable both to “vandalistic” practices such as denial of server attacks (slowing the server down or overwhelming it to the extent that it entirely stops functioning and to “hacking”—individuals trying to break into a computer to obtain information.

Viruses. Viruses involve computer code that maliciously alters the user’s computer code. These are often spread when the user causes a program to be executed. This may happen when the user attempts to open an innocent-looking attachment. It may also be possible to penetrate firewalls to directly plant the code into the computer. The “Trojan horse” is a variation on the viruses where the user is given what looks like a legitimate program which in reality contains the virus code. Computer games are often used for this purpose.

Spyware is a special kind of malicious code that will send information from the user’s computer to some third party. This information may be used for fraudulent purposes or for extortion.

Spam involves unsolicited e-mail. This concept has been previously discussed in the context of permission based marketing. Software is available to identify messages that are likely to be spam, but this software is not perfect. Many messages that actually contain viruses may be let through and certain legitimate ones may be incorrectly identified as spam.

Current Topics in Electronic Commerce

Tradeoffs between photo size and quality. Digital cameras today can take photos at very high resolution. Photos taken at high resolution will, however, result in very large files that take time and bandwidth to download. Various ways are used to reduce file sizes. Some file formats result in compression—“squeezing” the picture into a smaller file. This compression can either be reversible or irreversible depending on its method. In reversible compression, redundant information is stored in a more efficient manner. For example, instead of actually having fifty consecutive green dots, something like “50GD” can be recorded. Other types of compression are not completely reversible—some information that contributes less to the overall picture is removed. The JPEG format reduces file sizes relative to more “pure” formats but also in some irreversible decrease in picture quality.

Blogs. Recently, blogs written by individuals have become popular. Some are written for personal expression while others may be written to promote certain products. Some very popular blogs are able to attract a great deal of traffic and can make money through advertising.

Mobile Internet access. Increasingly, the Internet can be accessed by individuals who are away from their homes and offices. Many hotels now offer either wired Ethernet connections or wireless access. Many establishments—such as certain Starbucks locations—now offer wireless access through “hotspots.” Depending on the specific arrangement, users may be able to pay either for limited time use may belong to plans that offer unlimited access at participating locations. T-Mobile, for example, offers unlimited access to all participating hotspots for a fixed monthly fee.

Individuals who carry computers while traveling may need to synchronize contents with their office computers—that is, to make all files available at both locations. This requires criteria to determine which files will take precedence. Frequently, the file with the most recent date will replace files with earlier dates. This, however, may not be what was intended if a modification was made to an old version of the file at a remote location. Therefore, backup of all files prior to synchronization is helpful.

Search engine developments. Microsoft recently launched the “beta”—or trial—version of its new search engine Live.com. So far, Google and Yahoo! have had a large lead in the market share of online searches. Microsoft hopes provide an improved search engine that consumers will prefer. The site is expected to contain more indexed pages than Google currently does.

Search engines today face new challenges as new technologies may allow for better search results. One approach involves the so-called “neural networks” whereby the computer uses software code that effectively changes itself based on experience. Another challenge is the use of human “natural language”—that is, using artificial intelligence to actually “understand” human conversations and expressions rather than just analyzing keywords. Finally, Bayesian updating involves using past experiences to improve future results. For example, if a large number of searchers for a given term select a site listed, say, as number five, this option might be moved up above selections that ranked higher by existing search algorithms but were not chosen as frequently.

New market research opportunities. The increasing number of blogs found on the Internet might be analyzed to gain insight into the experiences of consumers. Although the writers of these blogs are not representative of the consumer population as a whole, they tend to represent “opinion leaders” who run ahead of the rest of the population. More and more online sites offer hosting of photos—e.g., webshots.com. Although many of the photos here are “staged,” these may nevertheless provide an opportunity to see how certain products and branded versions are used.

Proposals by Internet Service Providers for special treatment for some sites for a charge. Some Internet service providers have proposed charging select content providers such as search engines and file sharing services (e.g., sites offering music for downloading) faster access in return for a special access fee. This proposal is being opposed by most content providers and many consumers.

Relational Databases

“Flat” vs. Relational Databases. Ordinary, or “flat” databases are much like spreadsheets that contain a limited assortment of data. An instructor, for example, may have a list of students and their scores. These databases only reflect information about the specific class, however, even though each student may be taking other classes. If we created a database containing records of each grade issued (by course, term, and student name), we could combine this with a database of students, allowing us to retrieve information about all courses taken by each student. This is known as *linking*. The grade and student databases share one or more common variables—e.g., student ID number—which are used to match information from the two databases. The student database could also be linked to a fee payment database listing fees paid by each student.

Variable types. Like spreadsheets, databases can contain different types of variables. Text, or alphanumeric, variables contain letters, numbers, and certain permitted symbols—e.g., period or comma. Numeric variables contain number quantities—e.g., dollars paid or grade point average. Note that certain information that looks numeric should be treated as a text variable. For example, the phone number 760-555-1212, if entered as a number, could result in the two latter numbers being subtracted from the first one. Sometimes, the “zip+4” format is used—e.g., 92231-4050. This, too, should be stored in a text variable. Other types of variables include dates—the data 5/11/06 will be displayed as this date by the computer but stored by the computer as a number—say the number of days since January 1, 1800. This allows the database to make calculations that can be expressed as dates—e.g., if a bill is issued on January 1 and is due 30 days after the date of issue, the conceptual calculation would be something like $\text{date}(01/01/06)+30$, resulting in the date January 31. A number of other variable types exist:

- **Logical.** Here, a variable is said to either “true” or “false.” For example, a bill can be either paid (true) or not paid (false).
- **File.** A record can be set to point to a file. For example, the university may have a database of students with a digital photograph of each student. The filename of the photograph is identified in the database.

- **Memo.** This type of variable allows non-routine text to be stored by each record. For example, if it is normally university policy not to release information about students, a note might be made in one student's record that he or she has given permission for his or her e-mail address to be released if his or her long lost aunt contacts the institution.

Microsoft Access is an example of a relational database. Microsoft FoxPro is another. Because many databases involve a great deal of programming, maintaining compatibility with previously popular software may be practical. dBase III was one of the popular early databases. FoxPro is able to run the programming language used in dBase III, making that application popular for running databases that were started in earlier years.

Reference:

Carolyn Siegel (2006), *Internet Marketing: Foundations and Applications*, Houghton-Mifflin.