DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.033—Computer System Engineering Handout 16—March 6, 1997

Design Project #1: A PARTLY READ-ONLY PORTABLE WEB SITE

Introduction. One of the features of 6.033 is that we discuss real systems, both successful and unsuccessful. To get beyond discussion and give you some direct experience with designing systems we also assign two design projects. As in the real world, these projects have (we hope) a simple high-level problem statement but when you get into the system design you will find that there are hard choices to make. And, as in the real world, it is your job to explore, understand, and explain what the best choices are, how to reconcile sometimes conflicting goals, and how to keep the complexity of your design under control.This term, the first of these design projects relates to naming and the World-Wide Web.

The problem. A friend, an Egyptologist, has constructed a web site that consists of about 1,000 distinct web pages holding text, drawings, and photographs of archeological sites--pyramids, monuments, tombs, mummies, and that sort of thing. One of the characteristics of these pages is that they are highly interlinked with cross-references, but they contain no links out to other sites.

Your friend travels to Egypt frequently, with a laptop computer, and would like to carry the content of his web site along when he is in the field. The reason for carrying a copy along is that Internet connectivity in Egypt ranges from weak to non-existent, depending on the location. This year's laptops come with CD readers, and the idea of copying the contents of the web site onto a CD-ROM is very attractive--CD-ROM's are durable even when dropped in the sand, CD writers have recently become cheap and work well, the CD-ROM can easily be placed in any computer in addition to the laptop, and any web browser that happens to be available on the computer provides a convenient user interface. There is even a standard way of writing a complete file system onto a CD-ROM in such a way that it can be mounted on practically any computer system, much like the file system on a floppy disk can be mounted on a PC. A single CD-ROM can hold 650 Megabytes of data, so there is plenty of room on it to hold the 400 Megabytes of stuff currently in the web site.

The problem with this idea is that once written, a CD-ROM is unchangeable. Your friend considers an important virtue of the web organization for his materials to be that he can update individual pages or add new pages as fast as he discovers new things. So there is a need to provide some form of update despite the fact that a CD-ROM is read-only. The option of carrying a CD-ROM writer to Egypt is not attractive, because writing CD-ROM's requires a carefully controlled environment. Your job is to come up with a scheme that your friend can use that has the following features:

- Before leaving on a trip, your friend can copy all of the current content of the web site onto the CD-ROM. (Such a read-only copy is sometimes called a *snapshot*.)
- Your friend, while traveling with the laptop, can add new pages and images, update old ones, delete or rearrange things, and generally make any changes he deems appropriate.
- When internet connectivity is available, he can upload the updates to his home web site, so the rest of the world can keep up with his discoveries. (Your friend is the sole author of this web site, so you don't have to worry about updates coming from multiple sources.)
- As he travels, he can give extra copies of the CD-ROM to his Egyptian colleagues, for use on their own workstations, whether or not portable. He can also give them copies of any updated pages he has on his laptop at the time.

- A standard web browser can be the user interface everywhere: on the laptop even when not connected to the Internet, when your friend is at home, for holders of copies of the CD-ROM, and for the rest of the world. Most important, when your friend browses the local copy on his laptop he will always see the latest version of the materials.

Your job is to analyze the situation based on what you what you have learned in 6.033 and its prerequisites about the World-Wide Web, naming systems, and general computer system design issues--modularity, complexity, and so forth. Then develop a complete design that caters to your friend's requirements.

The design project. Propose a design of something (you get to decide what) that solves the problem, while at the same time introducing a minimum of new problems. You do *NOT* have to implement the design.

Before jumping to the design stage, you should do some more reading:

- Sections A and C of 6.033 reading #19, for a discussion of naming in file systems and naming networks.
- Tanenbaum's book, *Modern Operating Systems*, pp. 287-288, to learn how absolute path names, relative path names, and symbolic links work in the UNIX file system. (You are welcome, but not required, to assume that your friend's laptop runs a version of UNIX)
- There is a whole page of references on the mechanics of web naming that you might find helpful. Two things of particular interest there are:

    1. The discussion of relative addressing in Mark Andreessen's HTML primer. (You can find links to this and the next item, as well as other possibly relevant documentation about the Web, on the 6.033 Design Project 1 Bibliography web page.)

    2. The description of the BASE element in the HTML specification, to see if that offers any help in solving this problem. The BASE element is an HTML construction that looks like this:

        <BASE HREF = "some URL or local directory name">

Another possible source of information is to invoke Alta Vista and look for Web pages that have the word "CD-ROM" or "portable" in them. Unfortunately, there are about 500,000 such Web page for each of those queries. But maybe you can come up with a more helpful query. ("Egyptology" will lead you to some fascinating stuff, but since your friend is on the leading edge of technology you probably won't find this problem already solved there.)

Some things to think about...

- Should the home web site be organized in exactly the same way as the portable copy?
- Can you allow links to be constructed with relative references such as "../xyz/abc.html" or to things in the same directory? Or do you have to restrict the organization and kinds of references found in web links?
- Is the BASE construct of any help?
- Just before each trip to Egypt, your friend will need to write a new CD-ROM containing the latest snapshot of the web site. Are there any special tools that are needed--or that could help--at this stage? More generally, what procedure (and what custom software, if any) is needed for each of

    - Taking a snapshot of the web site,
    - copying a partially-updated laptop copy to another laptop, and

- re-syncing the web site with changes made in the field.

Which of these procedures can be done, for example, with ordinary copying, and which require adjustments to content?

- How can your friend be sure that he has got things set up right? Does he have to click on every (web) link to verify that it goes to the intended target? Is there a tool that can help?
- (This is actually something *not* to think about.) Your friend's web site consists entirely of HTML pages, images, movies, and sound tracks. If you happen to know about more exotic web things such as cgi scripts, shockwave, ActiveX and java applets, etc., he has no current use for them and probably never will, so you shouldn't worry about them.
- Recall that your friend also gives colleagues gift copies of the current CD-ROM. What are the possibilities, and what is the best method of getting updates from the laptop to them? Should the laptop be prepared to do it several ways?
- More on the gift CD's: If recipients have internet access, they will still prefer to use the CD-ROM for pages that haven't changed, but they would like to be able to automatically obtain any available updated pages from the home web site.
- Another possible complication related to the gift CD's is that your friend visits different colleagues on each trip to Egypt, so some of those colleagues will be using a copy of the latest CD-ROM, while others may still be using older ones of various different vintages.

There are at least three fundamentally different ways to approach this problem, each with its own merits and disadvantages. It may be very difficult to completely achieve all of the properties that you consider desirable at the same time. In most system designs, trade-off and compromise is required, so you have to decide how important each desirable property is in relation to the others.

Your report. Your paper should be 8 to 10 pages in length. You should start by explaining to your intended audience the background of the problem in terms that audience can understand. Next, describe what you have decided are desirable properties of a solution. Then give your solution and explain how well it achieves (or fails to achieve) the desirable properties. Throughout the paper you should justify each of your design decisions, especially in relation to alternative decisions that you could have made. You will be more convincing if you say not just why your idea is good, but why it is better than the alternatives. (For example, if another approach would meet all of the objectives perfectly, but the cost would be 100 times higher, then you should mention that as a reason for choosing your less-general but cheaper approach.)

Write for an audience consisting of colleagues who took 6.033 five years ago. That is, they understand the underlying system, network, and naming concepts and have a fair amount of experience applying them in various situations, but they have not thought carefully about the particular problem you are dealing with. Assume that your paper will also be used to convince your friend's computer guru that you have the right idea. Finally, give enough detail that he can turn the project over to that guru for implementation with some confidence that you won't surprised by the result.

When evaluating your report, your instructor will be looking at both content and writing.

Content considerations:

- Does your solution actually address the stated problem?
- Do you explain your decisions and the trade-offs?
- How complex is your solution? Simple is better, yet sometimes simple won't do the job. But unneces-

sary complexity is bad.

- Does your solution fit well with the rest of the system? If your solution requires modifying every piece of hardware, software, and data in sight, it won't be credible, unless you can come up with a very good story why everything needs to be changed.
- How extensible is your design? Are there opportunities for later addition of desirable features that you decided to omit?

Writing considerations:

- Is the report easy to comprehend?
- Is it well organized and coherent? Does it start by summarizing the problem and your solution? (Remember, you aren't writing a mystery novel. Perhaps you should start with an abstract (See HREF=http://tute.mit.edu/course/21/21.guide/www/abstract.htm).
- Does it explain the approach or architecture conceptually before delving into details, components, and mechanics?
- Does it use diagrams where appropriate? (A frequent problem when people use word processors is that they try to express everything in words, either because the word processor doesn't make it easy to include diagrams, or they haven't ever learned how to use the drawing features. Pictures can communicate some ideas far better.)
- Does it use the concepts, models, and terminology introduced in 6.033? If not, does it have a good reason for using a different universe of discourse?
- Does it address the intended audience?

You can find other helpful suggestions on writing this kind of report in the M.I.T. Writing Program's on-line guide to writing Design and Feasibility Reports (HREF=http://tute.mit.edu/course/21/21.guide/www/feasibil.htm).

Phase Two writing considerations: If you are enrolled in the 6.033 writing practicum, you don't need to do anything special; your practicum instructor will explain how the report will get you credit for the Phase II writing requirement. If you are not enrolled in the practicum, AND you want us to forward your design project report to the writing program as your phase II writing project, please say so on the cover page, and make sure that your report is at least 8 pages long. Note also that the writing program has a rule that they will accept only reports that earn a B or better from the class in which they originate. Finally, be aware that the second design project will probably be a team project, and thus much more difficult to tailor to the needs of the writing program than this one.

Collaboration: This project is an individual effort. You are welcome to discuss the problem and ideas for solution with your friends, but if you include any of their ideas in your solution you should explicitly give them credit, and you should be the sole author of your report.

Schedule: Your report is due in recitation *Thursday, March 20, 1997.*