

CHAPTER 1

INTRODUCTION

Internet innovation has rapidly changed the way people seek, deal and use information. It has opened avenues to every possibility, depending on how the information is being utilized or misused. With the new technologies, information can be accessed more rapidly, transactions can be done in one location, and transmitted through the information superhighway to any part of the world. As a result, it has made the world smaller and enabled to realize a borderless world concept. The Internet has given access to numerous computers that store vast amount of information. The era of Internet has affected all categories of people in our society, be they professionals, laymen and even school going children.

The World Wide Web (WWW), one component of the Internet technology has achieved global connectivity. Related to that, the role of computers has also changed from knowledge processors, word processors, or spreadsheet processors into knowledge sources [Card et al., 1996]. The content of Internet includes published sources such as books, journal articles, online newspapers, free software and patches, film, audio, video, as well as courseware. Internet is growing rapidly and links millions of users world wide. Many new technologies as well as new computer languages (such as Java, JavaScript, and VBScript) to support the Internet's interaction with its users are being developed and already in used. The Internet Model is shown in Figure 1.1 below.

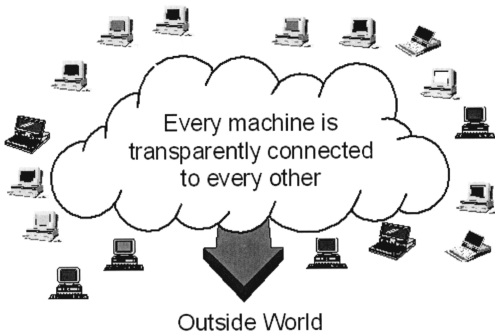


Figure 1.1 Internet Model

At the beginning of its life, Internet is very much restricted to text-based transmission of information. This is in the form of electronic mail which mostly run on UNIX, a command-based operating system. Most of today's Web software and applications are graphically represented and user friendly as well. These actually very much rely on the graphical user interface (GUI) browser technology such as the MS-Internet Explorer, Netscape Communicator and their predecessor such as Mosaic. Not only that, the current Web technology has enabled multimedia documents to be transferred so fast and easily.

1.1 PROBLEM STATEMENT

Even though the Internet has been very much different in its look since it gained popularity in 1994, still many users found it to be difficult to handle. The major problem users' encountered is getting lost in the cyberspace, a situation known as disorientation and cognitive overload [Conklin, 1987].

1.1.1 Disorientation

Disorientation is a significant cause of stress in modern life. Consider the feelings of a student who takes pride in his punctuality, already fifteen minutes late as he searches for a lecture room among a warren of lecture halls in an unfamiliar campus building. Hearts pound, sweat beads, worries form – all as a result of this student not being able to find his way.

The disorientation on the Internet might not as serious as the above student experienced. However, many Internet users experienced disorientation during their web navigation. In his survey, Conklin defines disorientation as "the tendency to lose one's sense of location and direction in a non-linear document" [Conklin, 1987].

In traditional text, it is not easy to get lost. There is the table of contents of topics with page numbers, the index with keywords and page numbers, and also bookmarks. However, in a complex hypertext network as found on the Internet with thousands of nodes and links, it is more than likely that the reader will get lost. This effect is most pronounced in badly designed systems lacking sufficient context clues and navigation aids. In addition, documents with hyperlinks should provide quick and easy methods to:

- exit the site
- return to the "home" node or title page
- retrace one's path

1.1.2 Cognitive Overhead

Cognitive overhead is a problem due to making decisions as to which links to follow and which to abandon, given a large number of choices. For example, the process of pausing for either to jot down required information or to decide which way to go can be very distracting. It can become a serious problem if there are a large number of nodes and links.

Researchers in academic institutions, industrial and commercial laboratories are working hard to produce better interfaces and tools. This is in order to help users to use the Internet more comfortably and without experiencing disorientation and cognitive overload in the maze of global computer networks. Areas being explored and researched are like information management, information retrieval and information visualization.

This thesis main interest is towards visualizing the information in the content-rich (also called content-based) web sites with emphasize on their organizational design. The examples of content-rich web sites are like The CNN Interactive (<http://www.cnn.com>), The Library of Congress (<http://www.loc.gov>), Yahoo (<http://www.yahoo.com>), ZDNet (<http://www.zdnet.com>) and MendelWeb (<http://netspace.org/MendelWeb/MWtoc.html>). In Malaysia, one of the best examples of content-rich web site is National Archives of Malaysia (<http://arkib.gov.my>). These sites hold thousands of information or may be even millions of them. A thorough analysis of content-rich web sites and examples of some such sites are discussed in Chapter 2.

1.2 SCOPE OF RESEARCH

The scope of this research is to visualize the contents of the National Archives of Malaysia web site. The goals are to reduce disorientation and cognitive overload in the part of users. The study of content-rich web sites available is conducted and how Information Visualization techniques can be applied to such web sites is reviewed. A very good example of content-rich web site is the Library of Congress (<http://www.loc.gov>). The discussion regarding this site and other content-rich web sites can be found in the following chapter.

One of the content-rich web site in Malaysia which have the similar characteristics as the Library of Congress is the National Archives of Malaysia (<http://arkib.gov.my>) web site. This web site contains many documents sorted in a few collections. They are in multiple formats such as text, image and sound. The organization of the contents of this site makes the core study in this thesis where the entire National Archives web site will be visualized.

In a proceeding paper entitled "Visualizing Internet Resources", Gershon et al. [1995] mentioned that "... another problem for Internet surfers is that often they do not know where they 'are' in information space and cannot remember how they got there. This is sometimes referred to as being 'lost in cyberspace'. One remedy is to provide users with a view of the information space available to them. The user can 'jump' from one document to another by clicking the mouse button without having to backtrack resource by resource, or in Web parlance, page by page."

Gershon et al., also suggests that the facility provided for the Internet surfers must have the following features:

1. to allow the user to view the web site which is depicted as a visual tree structure

2. the user can jump from one document to another by pointing and clicking without having to go back
3. allows the users to view the names of documents and how they are linked without actually opening and reading each document

The current National Archives of Malaysia lacks all of these features. Thus, the most obvious problem users will encounter is disorientation and cognitive overload as stated in Section 1.2. As a solution, one visualization technique will be applied to this site in order to overcome the problems mentioned above.

1.3 THESIS GOALS

Normally, users would prefer to visit and use web sites that have a comprehensible, predictable and controllable environment. This is in which they can quickly and safely explore as well as use the information. Thus the goals of this thesis are:

- i. To produce a user interface that realizes the Information Visualization technique with the features as stated in Section 1.2
- ii. To reduce disorientation and cognitive overload to users while using the site
- iii. To give a quick overview of the site to the users in terms of the surrogates and table of contents-liked metaphor
- iv. To give a sense of 'worthy to visit' to the site (or as a value-added feature* to the site)
- v. To enable user to search more effectively and dynamically as when the dynamic query is supported
- vi. To manipulate user's familiarity with Table of Contents representation adopted from printed sources such as books, journals and magazines.

Problem Statement

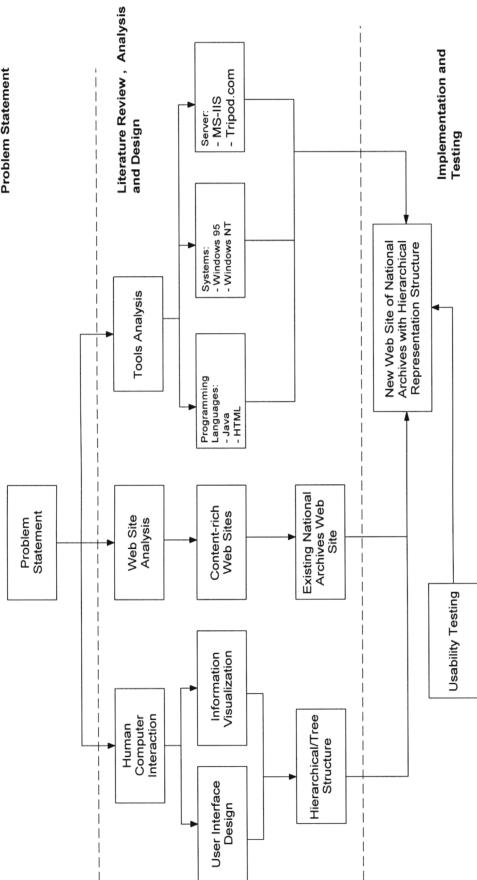


Figure 1.2 Overall Thesis Process

1.4 SIGNIFICANCE OF THE RESEARCH

This research will certainly offer great benefits to the intended users of the site. The targeted audience of this research are: the archivists, students who are doing research on the country's history, individual users and also for teachers who want to make this site as their primary resource centre for their supplementary teaching material.

1.5 METHODOLOGY

The course of action taken in carrying out this thesis is as depicted in the following Figure 1.2 Overall Thesis Process (see page 7). The figure shows how this thesis builds up starting from the problem statement as stated in Section 1.1 until the usability testing phase.

The problems are viewed from three different perspectives. First, as disorientation and cognitive overload discussed in this thesis are related to the Web's information seeking and navigation, thus several areas of interest were analyzed. They are information visualization and user interface design which actually studied under the umbrella of Human Computer Interaction discipline. Hierarchical tree structure is investigated as the technique in order to visualize the National Archives web site.

Another perspective being analyzed is the web site analysis. Different types of web sites are researched and reviewed. Examples from several content-rich web sites researched in Chapter 2 support the needs for a more structured site organization. National Archives of Malaysia web site is one of the web sites reviewed and is proposed as the case of consideration for this thesis implementation's domain.

The third perspective is in analyzing the suitable tools for the implementation purposes. Three areas of concern in this analysis are: the choice of programming languages (Java and HTML was chosen), hardware and software systems for implementation (Windows 95 and NT are selected) and server (MS-IIS and free web site hosting Tripod.com was chosen).

The final part in the Overall Thesis Process is the implementation part whereby the coding, uploading of web pages and testing being conducted.

1.6 THESIS ORGANIZATION

Chapter 2: Literature Review

In this chapter, the discussion revolves around the topic of web sites. Different types of web site genre are discussed. Five examples of content-rich web sites ranging from the Library of Congress web site to the National Archives of Malaysia web site are discussed.

Chapter 3: Information Visualization

This chapter covers mainly on the Information Visualization topic. In this chapter, the definition of information visualization is given, the framework of information visualization is discussed, and many techniques as the result of this field are presented. The good user interface principles as well as the hierarchical structure for presenting web sites content is also covered.

Chapter 4: Design

Chapter 4 focuses on the design of this thesis. A hierarchical tree structure for the National Archives of Malaysia is designed. Several user interface design

principles are studied. The knowledge gained from this study is applied to the design of the hierarchical tree structure.

Chapter 5: Implementation

Chapter 5 highlights the implementation aspect of this thesis. Concepts such as Internet, Java and web server are discussed. The algorithm and implementation for the new user interface called **CRIVE** (**C**ontent-**R**ich **W**eb**s**ite **V**isualizer) web site is also covered.

Chapter 6: Testing and Results

Chapter 6 will look into the usability testing of the CRIVE interface by comparing it with the existing National Archives web site. The analysis of the findings from the experiment held is discussed in terms of their tasks completion times and subjective satisfaction.

Chapter 7: Conclusion

Chapter 7 concludes the overall presentation of the thesis. The topics highlighted in the final chapter include: the problems faced during the course of implementation, system constraints, achievements, lessons learnt and future directions of the research.