CHAPTER 5

IMPLEMENTATION

In this chapter, the implementation of the CRIVE web site is discussed. It is in line with the intention of improving the visualization aspect of the existing National Archives of Malaysia web site. Section 5.1 will introduce the concept of Internet, World Wide Web (WWW), web browsers as well as HTML Section 5.2 presents the choice of programming language of this thesis – Java. Java will be reviewed in terms of its brief history, its features, applets, Java Virtual Machine (JVM) and the integrated software developments being chosen. Section 5.3 presents the server being chosen for developing and implementing the CRIVE web site. Section 5.4 highlights the algorithms being used together with the new interfaces designed and implemented for the CRIVE web site.

5.1 INTERNET

Internet History

The Internet began in 1969 as an experimental network connecting Department of Defense research groups [Zakon, 1998]. The development of the TCP/IP protocols in 1982, and their incorporation into the UNIX BSD 4.2 operating system in 1983 gave the system a sound infrastructure within which to grow. In 1986, the National Science Foundation (NSF) took over funding and administering the network, which had grown to several hundred hosts machines with connections to similar networks in Europe and Japan. NSFnet began with a 56Kbps (bits per second) backbone connecting universities and government research centers. In 1989 the backbone was upgraded to T1 speeds (1.5 Mbps). In

1990, NSF agreed to a proposal by IBM, MCI, and the backbone provider Merit, for the commercialization and privatization of the Internet. In 1992 the backbone was upgraded to T3 speeds (45 Mbps).

In 1995, more than 4 million hosts are connected to the Internet. NSF no longer directly subsidizes it, but instead funds a separate high speed (155 Mbps) experimental network linking a small number of national supercomputing centers. The Internet infrastructure is now provided by the major long distance phone companies as a commercial operation. Competition among carriers is expected to regulate the costs and capacities of the Internet.

The number of hosts connected to the Internet and to other, private, wide area networks (e.g. AOL, CompuServe) continues to grow exponentially. There will be a further explosion of growth when high-speed connections to homes replace the current 20-30 Kbs modems possible over existing telephone lines. How that will be delivered (phone, cable TV, or even power lines) is still an open question, but it seems likely that increased demand substantially driven by Web usage will produce cost-effective answers within 1 or 2 years.

According to PriceWaterhouse Inc. (PWC) forecast in its *Technology Forecast 1999*, it estimated that around 140 million Internet users worldwide at present and the number is expected to reach 330 million by the year 2002 [Chan, 1999].

World Wide Web (WWW)

The World Wide Web (WWW or simply "the web") grew out of a project started in 1990 by Tim Bernier-Lee at the European Particle Physics Laboratory (CERN) for allowing access to scientific papers across the Internet [Young, 1999]. He proposed a new language for the rendering of hypertext documents called

Hypertext Markup Language (HTML), and a protocol for handling these documents called Hypertext Transfer Protocol (HTTP).

Also part of the project was a "web browser", which can display HTML documents to the user, and can use HTTP to obtain the documents over the Internet. A hypertext document is one which contains references to other documents, and a browser can follow those references around the Internet as the user desires. These hypertext documents with their embedded references to other hypertext documents thus make the Internet into a "web" of information accessible to the user.

Web Browsers

In 1993 the National Center for Supercomputing Applications (NCSA) in Illinois released a robust and portable web browser for UNIX workstations called Mosaic. They also added a number of extensions to HTML, including the ability to display images on the same page as the text. Mosaic with its point-and-click graphical interface and ability to mix text and images took the Internet by storm. Many saw the commercial potential of these types of documents, and there has been a continual elaboration of multimedia effects such as video and sound. The success of web browsers has been based on their ability to present information more flexibly than printed media [Young, 1999].

Since the browser is the way that users access the web, it becomes the environment in which all web applications operate, i.e. the operating system. In that sense Netscape threatens Microsoft's dominant position in the PC market that has resulted from Microsoft's control of the PC operating system DOS/Windows.

HTML

HTML (Hypertext Markup Language) is the set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser. The markup tells the Web browser how to display a Web page's words and images for the user [Trubac, 1998].

HTML is defined in practice both by Netscape and Microsoft as they add changes to their Web browsers and more officially for the industry by the World Wide Web Consortium (W3C). A new version of HTML called HTML4 has recently been officially recommended by W3C, making this level an effective standard. However, both Netscape and Microsoft browsers currently implement some features differently and provide non-standard extensions. Web developers using the more advanced features of HTML 4 may have to design pages for both browsers and send out the appropriate version to a user. Significant features in HTML 4 are sometimes described in general as dynamic HTML.

5.2 JAVA

Java History

The origins of Java trace back to 1991, when Sun was investigating consumer electronics products. At this time, Mosaic and the World Wide Web were just interesting concepts [Cohn, 1997].

James Gosling, the father of Java, was intent on building a low-cost, hardware-independent software platform using C++. For a number of technical reasons, C++ was dropped, and a new language, called Oak, was developed, based on C++, but eliminating its shortcomings. These shortcomings include problems

associated with multiple inheritance, automatic type conversion, the use of pointers, and memory management.

Oak was used to develop a small electronics device called *7. This project resulted in the precursors of many of the components of Java: the development environment, runtime system, and API. The technology was explored in a number of consumer applications but was a little ahead of its time.

By 1994 the Web emerged, Oak was renamed Java. Today, Java has been a very popular language used for Internet applications and applets due to two reasons:

- Java has enable interactivity and dynamic effect to the WWW because previously, the WWW content is passive and static.
- Java is platform independent and "write once, run anywhere". Thus, the delivery of WWW content using Java program is now no longer dependent on the configuration of each user's Web browser.

Features of Java

The creators of Java have defined the Java language as "a simple, object-oriented, distributed, interpreted, robust, secure, architecture-neutral, portable, high-performance, multithreaded, and dynamic language." [Morgan et al., 1997].

- **Simple**

  Java started out as C++ but has had certain features removed, thus it is a simpler language than C++. The simplicity of Java is enhanced by its similarities to C and C++. Because many of today's current programmers, especially those likely to consider using Java, are experienced in at least C and probably C++, Java is instantly familiar to these programmers.

- **Object-Oriented**

  Java supports for object-orientation. Java classes are comprised of methods and variables. Class methods are the functions that an object of the class can respond to. Class variables are the data that define the state of an object. In Java, methods and variables can be declared as *private*, *protected*, or *public*. Private methods and variables are not accessible outside of the class. Protected members are accessible to subclasses of the class, but not to other classes. Finally, public methods and variables are accessible to any class.

- **Distributed**

  Java facilitates the building of distributed applications by a collection of classes for use in networked applications. By using Java's URL (Uniform Resource Locator) class, an application can easily access a remote server. Classes also are provided for establishing socket-level connections.

- **Interpreted**

  Java is an interpreted language, thus once the Java interpreter has been ported to a specific machine, that machine can instantly run the growing body of Java applications.

- **Secure**

  Java language also focus on security features especially for Internet purposes. Because Java does not use pointers to directly reference memory locations, as is prevalent in C and C++, Java has a great deal of control over the code that exists within the Java environment.

- **Portable**

  Java code is portable. It was an important design goal of Java that it be portable so that as new architectures (due to hardware, operating system, or both) are developed, the Java environment could be ported to them. In Java, all primitive types (integers, longs, floats, doubles, and so on) are of defined sizes, regardless of the machine or operating system on which the program is run.

- **Multithreaded**

  Java supports for multiple, synchronized threads which is built directly into the Java language and runtime environment. Synchronized threads are extremely useful in creating distributed, network-aware applications. Such an application may be communicating with a remote server in one thread while interacting with a user in a different thread.

- **Dynamic**

  Because it is interpreted, Java is an extremely dynamic language. At runtime, the Java environment can extend itself by linking in classes that may be located on remote servers on a network (i.e., the Internet).

Applet

In discussing Java programs, it has become standard to refer to Java programs that are embedded in another language as applets and to stand-alone programs as applications. For example, when the user uses Java to augment a World Wide Web page, the Java code is embedded within HTML code. Therefore, this is referred to as an applet. On the other hand, a Java program that is not embedded

within HTML or any other language and can stand on its own is referred to as an application.

In the context of the Web, Java applets offer the following advantages:

- Java applets are dynamic
- Because they run on the client, not on the server, thus Java applets can make better use of computing resources.
- Java is designed to be "architecture neutral" which means it can run on any platform

The new interface of the CRIVE web site is developed as an applet which embedded within HTML code of the site's web page.

Java Virtual Machine (JVM)

The Java Virtual Machine is an abstract computing machine [Morrison et al., 1997]. Like a real computing machine, it has an instruction set and manipulates various memory areas at run time. The Java virtual machine knows nothing of the Java programming language but only of a particular binary format, the class file format. A class file contains Java virtual machine instructions (or *bytecodes*) and a symbol table, as well as other information.

Java source is compiled into bytecodes that run on a virtual machine implemented by the Java interpreter and run-time system. The use of virtual machines trades efficiency for portability: It is significantly easier to implement a bytecode interpreter for a given CPU than it is to write a native code compiler. Thus the Java compiler that compiles Java source code into bytecodes needs only be written once. Each different CPU architecture and operating system then

needs to implement a Java bytecode interpreter and provide an implementation of the runtime classes.

Java Integrated Development Environment (IDE)

- **Symantec Visual Café Pro 2.5**

Symantec Café, originally released in March 1996, is the first development environment that became widely available for Java programming after the JDK. Symantec calls it an integrated development and debugging environment (IDDE).

Café is a sophisticated IDE that offers an excellent source editor with color highlighting of syntax, an editor for class and hierarchy modification, a Studio tool for interface design, and numerous example applets. To aid in the design of a class hierarchy, Café has a class editor for navigating through classes and editing class methods, and a hierarchy editor for viewing and modifying Java class relationships. Changes in the source code that affect the class hierarchy can be seen as the program is being written, instead of requiring that it be compiled before changes are reflected in the hierarchy. The programmer can also change the source code from within the class editor--clicking the function or method within a class brings up its source code in a window to edit the code.

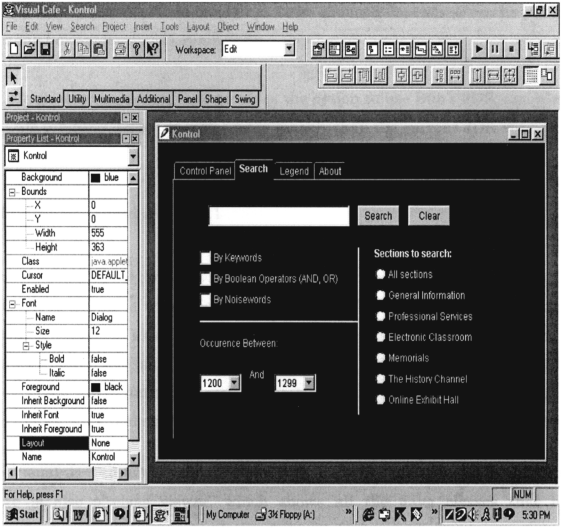

Figure 5.1 Symantec Visual Café Pro V2.5

With Café Studio, designing a graphical user interface for Java programs can be done in a visual, drag-and-drop manner (see Figure 5.1) Visual Café Pro enables programmers to develop the dialog boxes and other visual elements visually, and it creates event handlers for these components automatically. There's also a menu editor with an active window.

One interesting aspect of Café Studio is the ability it gives to design a form and dictate exactly how it looks. With the JDK and its Abstract Windowing Toolkit,

GUI designers have to allow their work to be changed depending on the platform the applet or application is running on.

For compilation purposes, Café provides the option to use Sun's JDK compiler or the Café compiler, which operates more quickly than the current JDK version.

- **Microsoft Visual J++ V. 1.1**

Microsoft Visual J++ is a Java development environment from Microsoft. It is designed to integrate with Microsoft's Visual Studio suite of development tools, Visual J++ features extensions to the Java class library that are specific to the Windows platform. Visual J++ sports the look and feel of the popular Visual C++ development environment, as well as most of its advanced features. Visual J++ provides the only real support of any IDE for fully integrating Java with ActiveX.

The Visual J++ editor is very nice and fully supports color syntax highlighting. There is also a class viewer, which shows all the Java classes that have definitions as well as the members of those classes, including properties and methods. Visual J++ has a very powerful graphical Java debugger that supports the debugging of multiple applets simultaneously from within a browser. The debugger comes complete with bytecode disassembly, bytecode-level stepping and tracing, and the ability to assign values to variables while debugging. The figure below shows MS-Visual J++ V1.1.

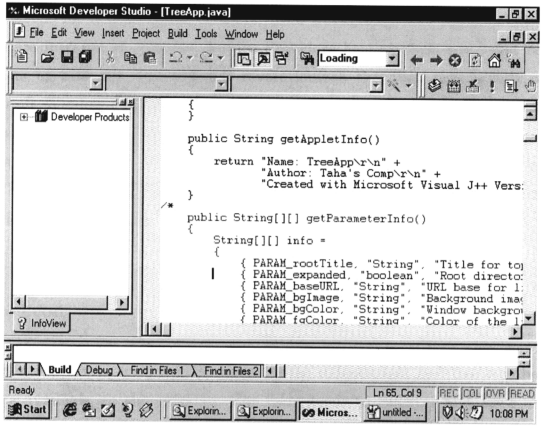

Figure 5.2 MS-Visual J++ Version 1.1

5.3 INTERNET SERVER

MS-Internet Information Server (IIS)

For this thesis, Internet Information Server (IIS) run on top of Windows NT 4.0 Server is used as one of the CRIVE site's web server. All the documents necessary for this thesis is placed on that server.

It is chosen because it is one of the web server provided by the system and is very powerful as well. It can effectively exchange information with users no matter they are located. Microsoft Internet Information Server (IIS) is designed to meet

these needs for a wide range of users, from workgroups and departments on a corporate intranet to Internet Service Providers hosting Web sites that receive millions of hits per day. IIS 4.0 revolutionizes the Web capabilities of the Microsoft Windows NT Server version 4.0 operating system by providing the easiest way to share information, build and deploy business applications, and host and manage sites. The only problem of using MS-IIS as a web server is that the CRIVE web site in discussion would not be available if the machine it runs on (Windows NT) is turn off.

Tripod.com

Another web server used for this thesis is the independent and free web hosting service available on the Internet. The user needs to register and have a legal account with the web site before a permission is given allowing the user to have his appearance on the Internet. This service started in 1995 and currently, it is a subsidiary company of Lycos, a big portal service available for the Internet users.

Figure 5.3 Free Web Hosting Server: Tripod.com

Each user is given a free space of 11MB in size. Additional web space is available for a moderate price. As the service is free and available 24 hours a day, there is no concern over downtime as in the case of using MS-IIS mentioned in the previous section.

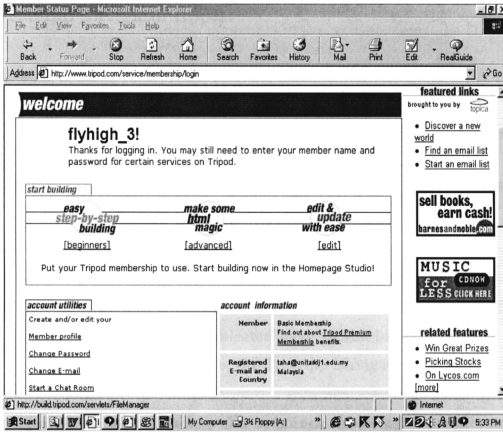

Figure 5.4 User flyhigh_3 Logged Into Tripod.com

The Figure 5.3 and Figure 5.4 show the user-friendly free web server in action.

5.4 ALGORITHM AND CRIVE INTERFACE

There are two main applets written using Java for the CRIVE web site. One of the applet is written using MS-Visual J++ 1.1 and another one is using Symantec Visual Café Pro running on a Windows 95 machine. The applet written using MS-Visual J++ is called TreeApp Navigator applet while the one written using Symantec's is called Control applet. The rest of the development especially the ones requiring HTML coding is done using a text editor.

5.4.1 TreeApp Navigator Applet

Overview

TreeApp Navigator applet provide a convenient interface for browsing the CRIVE web site. The applet operates by opening a window displaying a list of the web site contents in the left frame of the browser window. TreeApp Navigator appears very much like MS-Explorer used in Windows 95/98 operating system. It is hierarchically organized where users can expand or contract the hierarchy tree structure (see Figure 5.5). Users can easily navigate this list and select pages; the navigator displays further information about each selected page and offers the user an option of opening up the selected page in the browser window (on the right side of the frame). Design-wise, the TreeApp Navigator should be:

- Accessible – a user should be able to launch the navigator applet no matter what kind of browser they are using
- Easy to use – a novice should be able to use the navigator applet
- Efficient – code size should be kept to a minimum to reduce download time especially when the applet is accessed from Tripod.com

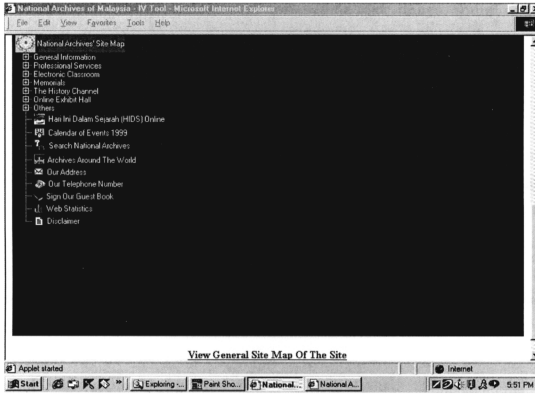

Figure 5.5 TreeApp Navigator applet

Using TreeApp Navigator

The navigator applet is fairly easy to use. It appears as a hierarchical view of the CRIVE web site. Elements of the navigator tree are either folders or documents. Clicking on a closed folder will open the folder, revealing the contents of the folder (more documents and folders). Clicking on an open folder will close the folder.

When the user put the mouse pointer over any element of the tree, a brief description of the document is displayed. This includes the directory name, the file name, the file size, the date created, and the author. When the user click on

the document, the file associated with it (whether text, image, sound or HTML file) will be displayed on the right frame.

TreeApp Navigator Parameters

There are two aspects to configuring the navigator applet: the applet parameters and the site map file (created using text editor and saved as a HTML document – named demo.htm). Below is a list of the supported applet parameters:

- baseURL: the base URL where the documents are placed
- expanded : the applet option whether to enable expand mode of the tree structure
- offset: the distance between elements
- bgColor: the applet background color
- textColor: the color of the fonts used in the applet
- fgColor: the color used for the foreground of the applet

The example of how these parameters being used in the HTML file (stored as demo.htm) is shown below:

```
<param name="baseURL" value="http://members.tripod.com/flyhigh_3/">
<param name="expanded" value="true">
<param name="offset" value="5,5">
<param name="bgColor" value="blue">
<param name="textColor" value="white">
<param name="fgColor" value="white">
```

The Site Map file (demo.htm)

Even though this file is a HTML file, it is created using a text editor and sports a very simple and convenient format. The main advantage of using this format is because of its simplicity and easy to parse by the Java program.

Each entry in the map file must be on a separate line; each line consists of two main parts. The first part is the name to refer to the document, and the second part is to refer to the document particular instances and each instance is separated by a colon (;) symbol.

The convention used is as the following:

> <param name="[*the reference number of the element*]" value="[*the title of the element to appear on the site map*] ; [*the image file to associate the element with*] ; [*the address where the element is to be read from*] ; [*optional field used to associate an image file once the tree is expanded*] ; [*the caption showing the element details on the status bar*]">

An example of how the format is used in the site map file is shown below:

> <param name="item159" value="Hari Ini Dalam Sejarah (HIDS) Online;images/hids.gif;html/hids.html;Hari Ini Dalam Sejarah (HIDS) Online - filename: hids.html - size: 4KB - date created: 14041999 - created by: Taha's Comp - parent:">
>
> <param name="item160" value="Calendar of Events 1999;images/cal99.gif;html/calendar_cal.html;Calendar of Events 1999 - filename: calendar_cal.html - size: 36KB - date created: 14041999 - created by: Taha's Comp - parent:">

<param name="item161" value="Search National Archives;images/find.gif;html/search.htm;Search National Archives - filename: search.htm - size: 2KB - date created: 13041999 - created by: Taha's Comp - parent:">

<param name="item162" value="Archives Around The World;images/web.gif;html/internet_oa.html;Archives Around The World - filename: internet_oa.html - size: 3KB - date created: 17041999 - created by: Taha's Comp - parent:">

TreeApp Navigator Classes

The TreeApp Navigator that have been developed consists of six separate classes. The first, TreeApp Navigator is the actual navigator applet. This applet is embedded in a Web page and displays as the table of contents which also sports a tree-like structure, hence the thesis is titled hierarchical representational structure.

The second class is the FolderItem class which monitors and controls the operations done in a folder element.

The third class is the Item class which is the individual element, be it a folder or a document. It controls the behavior of the item such as to where to open the item and so on.

The fourth class is the ItemAction class which controls the behavior of the item such as to where to open the item and so on..

The fifth class is the ScrPane class which provides and displays a hierarchical view of the site contents using a helper class called TreeControl class (the sixth class) which represents the entries in the site list.

TreeApp Navigator Elements

Once the TreeApp Navigator applet is activated, it will show the elements (documents) which reflect the contents of the web site it is representing. For the CRIVE web site, the elements especially the HTML documents, image files and the text files are all downloaded directly from the existing National Archives web site. Some of the elements were slightly modified to improve their 'look and feel' as well as to suit their new appearance on the TreeApp Navigator applet. Some of the elements were added and some are removed altogether.

5.4.2 Control Applet

Overview

Control applet provides the navigational options to help users navigate and use the TreeApp Navigator applet more efficiently. The applet is displayed on the top-most part of the right-hand side of the frame of the browser window. Below the Control applet is where the users will see the TreeApp Navigator applet. The following figure shows the Control applet is in its active mode.

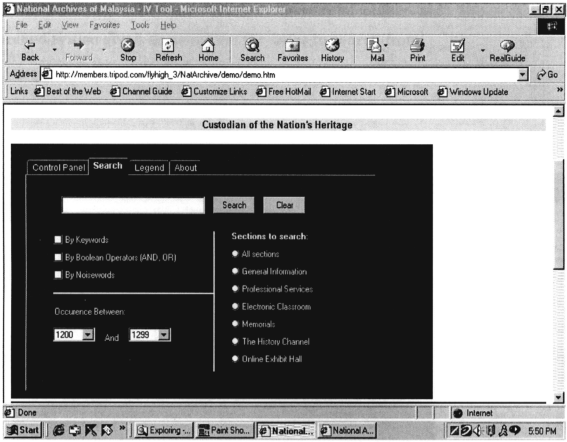

Figure 5.6 Control applet

The Control applet is using a tab panel approach in order to save the screen real estate. It has four tabs and each tab serves different purpose. When the web page of the site is loaded, the Search tab panel is shown first (set as default).

Search Tab

Figure 5.7 below shows the Search tab in its active mode. Four main options were offered in this tab. The user can search anything they wish to see from the site. The entry is typed into the text field and once finished, the user will hit the Search button to execute searching. Users are given choice to search either by keywords, by boolean operators or by noisewords. They are also given a chance to

find file being searched according to its year of occurrence. The section to be searched is also given as an option for users to search from.

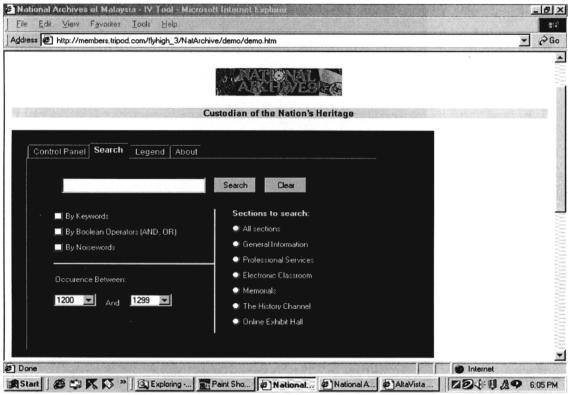

Figure 5.7 Control applet: Search tab

Control Panel

The Control Panel tab offers two options. The first one allow users either to keep the table of contents structure in expand or contract mode. They can also set the table of contents structure to show titles and icons, or just titles or just icons as they wished. The Figure 5.8 below illustrates the idea.

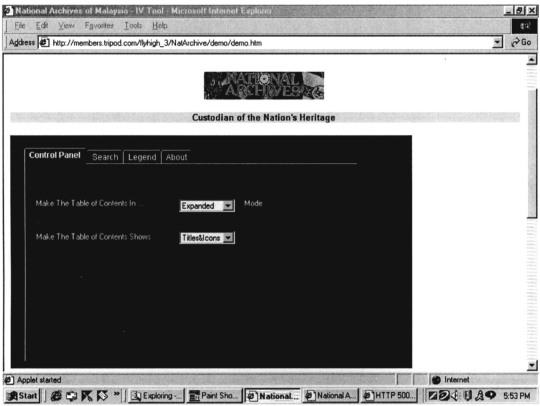

Figure 5.8 Control applet: Control Panel tab

Legend Tab

The Legend tab shows all the icons being used in the table of contents structure seen on the TreeApp Navigator applet. There are seventeen icons used and associated with the documents as found in the navigator applet. Figure 5.9 shows the Legend tab in its active mode.

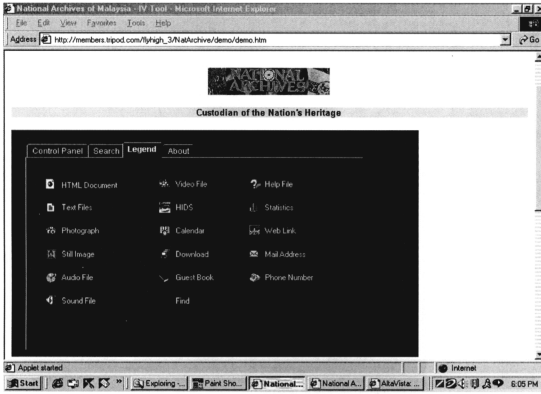

Figure 5.9 Control applet: Legend tab

About Tab

The About tab shows a brief information about this program and new user interface. Figure 5.10 shows an active About tab panel.

Figure 5.10 Control applet: About tab

5.4.3 The CRIVE Interface

The implementation of CRIVE interface is discussed in this section. The site can be accessed either from the MS-IIS web server running on the Windows NT machine or from Tripod.com. The discussion of the web server was covered in Section 5.3.

The CRIVE Main Page

Figure 5.11 below shows the main web page of the CRIVE web site once the user accessed it. The loading process can take a few seconds as the applets were loaded to the web browser.

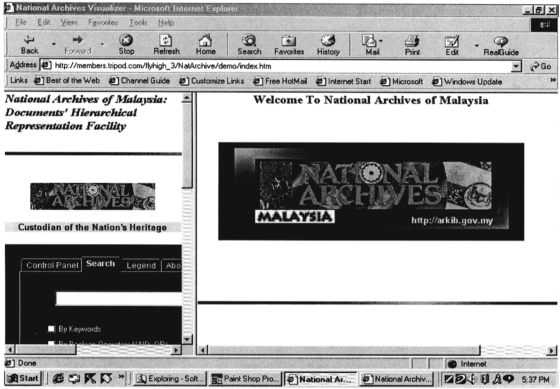

Figure 5.11 The Main Page of CRIVE Web Site

Description

The users can see two frames dividing their web browser. On the right frame is the entry page for the CRIVE web page while the one on the left is showing an applet called Control. It consists of four tab panels and users can click on each of them to view more what is offered by each of the tab panel. The discussion of the Control panels was covered in Section 5.4.

The Welcome Page

Figure 5.12 below shows the Welcome page as it is suggested in the table of contents structure (TreeApp Navigator applet). From the figure, the page is displaying the Welcoming information together with the image files of the minister in charge of the Archives.

Description

The General Information folder is expanded and several elements of HTML type are shown. The currently active document is the one referring to the Welcome document. It can be seen that Welcome in itself is a folder and it can be expanded to view more documents under it.

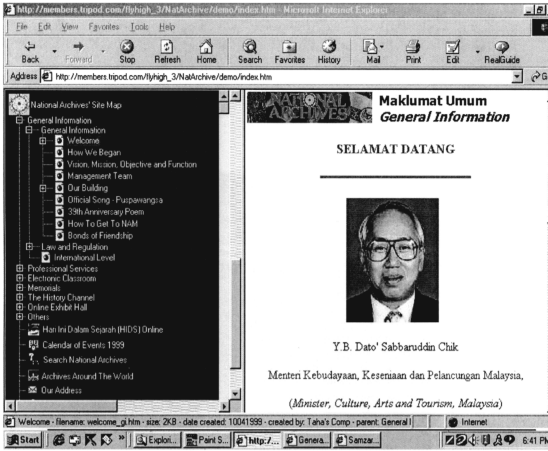

Figure 5.12 The Welcome Page of the CRIVE Web Site

The General Site Map Page

The General Site Map page (see Figure 5.13 and Figure 5.14 below) is launched when the user click on the "View General Site Map of the Site" hyperlink below the TreeApp Navigator applet.

Figure 5.13 The interface shows a hyperlink titled "View General Site Map of the Site"

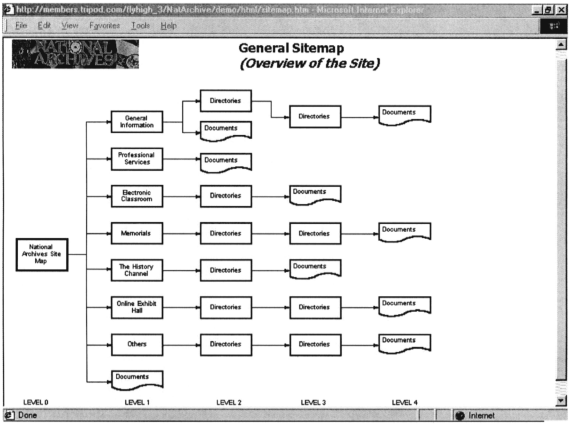

Figure 5.14 The General Site Map shown as an image file

Description

The purpose of this link is to show the user beforehand the go deeper into the hierarchy. It is also intended to give the users the overview of the depth of hierarchy offered by the site. Some users may get disoriented and cognitively overwhelmed if they knew that the web site is organized into a very deep hierarchy. As this web site only have 5 levels of hierarchy, thus the user may not feel being too overwhelmed.

The Document Detail's Page

Figure 5.15 below shows the size of document as it is shown to the user when the mouse pointer is placed over the element.

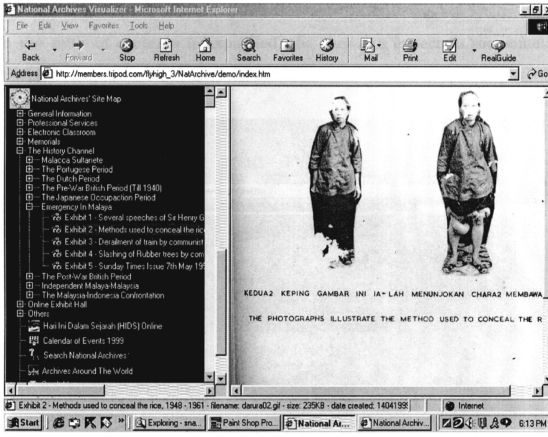

Figure 5.15 The CRIVE Web Site

Description

From the figure above, the user can see the size of the document (an image file in this case) when they look at the browser's status bar. The active document is also highlighted in red as can be seen at the table of contents (TreeApp Navigator applet). The users can also get other details of the picture: its title: Exhibit 2 –

Methods used to conceal rice, 1948 – 1961; its file name: darura02.gif; size: 235 KB and date created: 14 April 1999; and the other details are partly hidden.

The Table of Contents in Expand Mode

When the table of contents is expanded, the user can expect to see something similar to the one shown in Figure 5.16 below.

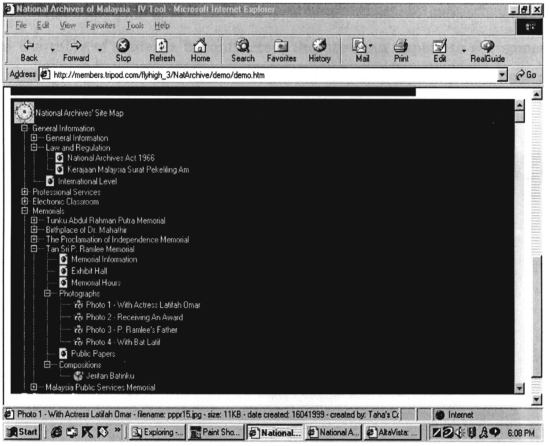

Figure 5.16 The TreeApp Navigator applet in Expand Mode

Description

The tree is expanded and the user can see how many levels the hierarchy is in any folder they wished to see. From the figure above, the user can expand the Memorials folder and see the active one currently is the Compositions folder under the Tan Sri P. Ramlee Memorial folder. An under the Compositions, users can listen to a song titled "Jeritan Batinku" as suggested by the title and also the icon.

The Table of Contents in Contract Mode

When the table of contents is contracted, the user can expect to see something similar to the one shown in Figure 5.17 with a very less information compared to the one shown in the previous figure.

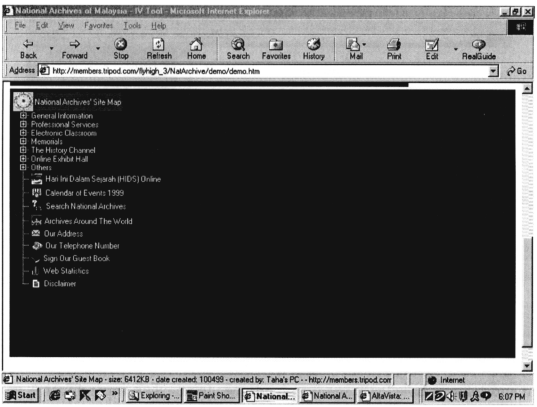

Figure 5.17 The TreeApp Navigator applet in Expand Mode

Description

The tree is contracted and less information available to the users. However, once the user wants to get more, they can click on the folder twice or simply clicking on the plus (+) symbol next to the folder name to expand the table of contents back. The figure shows that the size of the whole site is 6412 KB and other details are available as well.

The Page Showing Relation of Icon and File Type

When the user navigate the table of contents, they can see that all documents are associated with an icon. The examples are shown in Figure 5.18 and Figure 5.19 below.

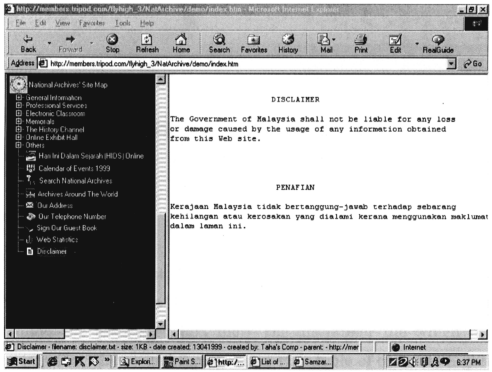

Figure 5.18 The new web site shows the document displayed in relations to its file type (text file).

Figure 5.19 The new web site shows the document displayed in relations to its file type (image file)

Description

As both figures above suggested, the icons represent the file type associated with it. The first figure highlights "Disclaimer" on the table of contents and the right frame displays the associated text file – the disclaimer file which in turns is actually a text file. The second figure shows that table of contents highlights the title "Datuk Seri Anwar Ibrahim" and the icon used is the type of still image. In order to know the type of file an icon is representing, users can refer to it from the Legend tab panel of the Control applet. This is one of the features that the existing National Archives web site does not have.

The next chapter will discuss on the usability testing aspect and then review the findings of the experiment conducted.