PLACING SRI LANKA ON THE IT MAP: 
A CASE STUDY

By

Dileeka Dias and Gihan V. Dias

Abstract

This paper presents experiences in the on-line publication of The Sunday Times on the Internet. The initial design methodology used according to the accepted norms of Internet publishing is discussed in detail. The formatting of the electronic edition, and the software tools developed for its automation are also presented. A summary and an interpretation of the access statistics for The Sunday Times Internet edition is also provided.

The Internet has become an ideal medium for electronic publishing of news today. Not only is it accessed by a very large number of people, but it is also fast, economical and widespread. The World Wide Web, one service offered on the Internet, is particularly suitable for on-line publication, due to its powerful capabilities in integrating text and graphics and in archiving and cross-referencing of the published material.

1.0 Introduction

1.1 Background

The worldwide dissemination of news on Sri Lanka's political, economic and social climate is of utmost importance today. Access to independent sources of such information is a significant factor in the establishment of credibility, good international relations and potential investor confidence in Sri Lanka. The availability of Sri Lankan news in a timely manner is also a valuable facility to the large expatriate Sri Lankan community.

To this end, Information Laboratories (Pvt.) Ltd. has been providing a voluntary news-clipping service on the Internet since 1992. Selected items from major local newspapers were distributed regularly over the Internet. The distribution was carried over the sinet mailing list, to which a large community of people interested in Sri Lanka subscribe to. These news items were also archived in the sinet archives for later access by interested parties.

The above news-clipping service was particularly appreciated during the periods leading up to the general and presidential elections of 1993 and 1994. Recently, this news clipping service, Spotlight on Sri Lanka, was expanded to include articles of relevance to Sri Lanka from foreign newspapers and magazines as well.

With the advent of the World Wide Web, the news-clipping service moved from electronic mail to this more versatile service. As a result, there was a large number of requests for a more extensive publication of independent news on the web. In response, the process of on-line publication of The Sunday Times was initiated, and was gradually expanded to the full-blown on-line publication it is today.

This on-line publication is a collaborative project between The Sunday Times, Lanka Academic Network (LAcNet) and Information Laboratories (Pvt.) Ltd.

1.2 Objectives

The objectives of this venture are briefly summarised below:

- Dissemination of independent news, views and features on Sri Lanka to the worldwide community.
- Provide the expatriate Sri Lankan community with a quick and efficient way to respond to or to contribute their views to the local media.
- Initiate the building up of a Sri Lankan news archive for later reference by the academic, journalistic and business communities.

1.3 Overview of Paper

The technical contents of this paper begins with a short introduction to the Internet and the World Wide Web. How the Internet is being used at present for news distribution is illustrated through specific examples. The design of The Sunday Times web site is presented in

Dr. (Mrs) Dileeka Dias,  
BSc Eng, MSc, PhD, MIEEE Director, Information Laboratories (pvt) Ltd.

Dr. Gihan V Dias,  
BSc Eng MSc, PhD, MIEEE, AMIE (SL)  
Senior Lecturer, Department of Computer Science & Engineering, University of Moratuwa
detail. A description of the implementation follows, together with the tools developed for the automation of publication. Advanced features included into the web site such as archiving and searching facilities are also presented. The paper concludes with an analysis and an interpretation of access statistics pertaining to the web site.

2.0 The Internet and the World Wide Web

2.1 What is the Internet?

There's no agreed-upon answer that neatly sums up the Internet. However, some common definitions often found are:

- A network of networks based on the suite of Transmission Control Protocol/Internet Protocol (TCP/IP) protocols
- A community of people who use and develop these networks
- A collection of services and resources that can be reached from those networks.

It is the view of the authors that the Internet would not be what it is without any one of the above three components.

The Internet started in the late 1960s as a project funded by the US Department of Defense, and gained quick and widespread popularity among the academic community. A short history of the Internet may be found in [Dias, G., 1996]. However, it was not until a few years ago that the potential of the Internet for publicity and marketing was discovered. Since then, the Internet has grown dramatically, the number of host connected to this worldwide network increasing by more than a million per year. The most popular services on the Internet are Electronic Mail, Conferencing or USENET, File Transfer or FTP, Telnet, Gopher and the World Wide Web. Detailed descriptions of these and other Internet services is found in [Liu 1994].

2.2 What is the World Wide Web?

The World Wide Web, or WWW as it is commonly known, is one of the newest and most dazzling of the services available on the Internet. It combines information in the form of text, graphics and sound in a powerful interactive format. Further, it has powerful linking capabilities which allow users to navigate the Internet in search of information stored on computers all over the world. These features have led to an explosive growth of this service which was launched in 1992.

![Figure 1](image.png)

This section provides a brief overview of the Web. For more details, the reader is referred to [Dias, D. 1995].

Like all other Internet services, the World Wide Web is based on a client-server architecture. It is implemented by a collection of servers connected in a web-like manner as shown in Figure 1.

These servers are accessed by users of the web using client or browser software. The requests for information by the client from the server and the server's responses are exchanged using the HyperText Transfer Protocol (HTTP). The browser formats the data received from the server for displaying to the user.

There are a number of different browsers, one of the most popular of which is Netscape. These browsers are either text-based or graphical. Graphical browsers can display both text and graphics, while text-based browsers such as Lynx can only display text. Most browsers can also use other Internet services such as email, usenet, FTP, Telnet and Gopher.

2.3 Documents on the Web

Documents on the Web reside on web servers, and are organized as pages, which are text documents interspersed with formatting commands written in a language called the HyperText Markup Language (HTML). HTML describes the structure of the document as well as provide pointers or links to various information sources from which the document has built up its content. Therefore, the contents of a particular document may include documents, pictures and other resources which may physically reside in servers elsewhere in the world. The physical locations of the resources accessed are transparent to the user.

The point of entry to a set of web pages on a particular
topic is called the *Home Page*. It contains a summary of the most important information on the topic, and a guide to the rest of the pages.

Each document on the web is identified by a Universal Resource Locator (URL), a unique address. The URL contains the address of the server on which the resource resides, the path to the file, the filename, and the protocol by which the document is to be retrieved.

Figure 2 shows some sample URLs. The URL in Figure 2(a) points to a file `homepage.html` residing in a server called `www.myservice.com` in the directory `mywebdir` and is to be retrieved using the HyperText Transfer Protocol. The URL in Figure 2(b) points to a file `chapter1.zip` which can be retrieved using the File Transfer Protocol (FTP) which resides in `myftpdir/mythesis` on the same server.

### 2.4 Hyperlinking

The most attractive feature of the Web is *hyperlinking*. Once a user retrieves a web page, they can find more related information simply by selecting these hyperlinks which appear as text highlighted in some manner. Hyperlinks consist of two parts, the *anchor text* which is displayed in a highlighted manner by the browser, and the invisible URL which points to the resource described by the anchor text. When a hyperlink is activated, the file to which the hyperlink points is identified and retrieved over the network from wherever it resides. This may be a text document, an image, a graphic, an audio clip, or a video clip. In this manner, the user can navigate the web of servers along any path that he/she prefers.

### 2.5 The Web as a Tool for News Distribution

The number of newspapers on-line has grown from 20 in 1993, to 100 at the end of 1994, to over 800 at the start of 1996. The range of delivery styles and news sources on the Internet is wide-ranging newspapers, news wires, news magazines, television networks and radio [Quintana 1996]. Many of these newspapers are available on the Web in their entirety at no cost. Some are only partially available and require a fee for full access.

Some examples of newspapers on the Internet include USA Today [USA], The New York Times [NYT], Wall Street Journal [WSJ], The Washington Post [WP], San Jose Mercury News [SJM], The Irish Times [IT], Le Monde [FRANCE] [LM] and The Daily News (Sri Lanka) [DN].

News wire releases such as those from Associated Press [AP], and Reuters on Yahoo [YAHOO] are available on the Internet. These news sources are updated continually throughout the day.

News magazines available on-line include Time [TIME], Business Week [BW], and The Economist [ECON]. The stories in these magazines sometimes have links to the home pages of the individuals or organizations mentioned in the article. This is an example of where the technology is fully used for the benefit of the reader.

Cable news networks such as CNN provide up-to-the-minute coverage of news including video clips. Cable sports networks such as ESPN provide up-to-the-minute results on sports scores and include photos and in-depth stories.

Major television networks also have descriptions of their television shows, transcripts of news broadcasts, and timetables for their programs. Some major television networks on the Internet with news content include NBC (NBC) and CBS (CBS). CBS has a daily on-demand video of their *Up to the Minute* news broadcast.

Radio stations such as CBS Radio [CBSR] are also making available their broadcasts on the Internet. Examples include CBS Radio.
Indexes to news sources on the Internet can be found at the Internet Free Press (http://www.newmedia.slis.uwo.ca/ftp/index.html).

3.0 The design of The Sunday Times web site

3.1 Objectives

The web enables electronic publishing of information very easy and attractive. Text, images and audio/visual information can be incorporated into the same document. However, in using these features, the authors must pay special consideration as to how the potential audience will access the documents. Such considerations will result in attracting a wide audience to the web site and maintaining their interest.

- The major design considerations of the web site are summarised below:
- To enable users with low-bandwidth Internet connections to access the material conveniently.
- To allow users to access the pages at the site without intolerable delays, and the accompanying excessive access charges.
- To allow users the maximum flexibility to navigate the web site according to their own preferences.
- To facilitate browser-independent viewing of the published material.
- To provide facilities for users to communicate with the publishers.
- To regularly enhance the web site by adding new features to retain the interest of the readership.

Most of the above considerations concern the Internet access cost and access bandwidth. Internet access providers generally charge their subscribers based on access time and the volume of data transferred. The access bandwidth varies widely from user to user. Most users may have dial-up access using 14.4 kbps modems. Even institutions which may have higher-speed dedicated access may have a large number of simultaneous users, resulting in congestion of the link. In both cases, having large graphics, audio/video files etc. increases the access time to intolerable levels. It has been found in surveys that if the wait-time for a page exceeds 5 seconds, the user will most probably lose interest and go elsewhere. It has also been found that 40% of users have no interest in accessing graphics [Emery 1995]. Even in cases where high-bandwidth access is available, it is important to consider the size of web documents, as the user may have to pay for the amount of data downloaded and may not be willing to access pages with high graphical content.

The next most important aspect of web publishing is to make the published material browser-independent. The publisher has no control over which browsers users will use to view their pages. Some users will use high-end graphical browsers, some will have their browser’s graphical features turned off to save time and some will use text-based browsers. Therefore, the publisher has to ensure that all the information contained in the pages will be conveyed to the user irrespective of the browser being used.

The following are some of the steps taken in order to achieve these objectives:

- Each graphic is designed to be small in size, and is reused whenever possible.
- A proper balance between text and graphics is maintained.
- A compressed version of the weekly issue is made available for users to download and read at their leisure.
- Hyperlinks are used to provide maximum navigational flexibility.
- The material is broken down into pages of approximately 15 kbytes wherever possible, in order to reduce access times.

Whenever large graphics or images are present, a “thumbnail” version is inserted into the documents, and a link provided to the full-sized version.

Testing of the web site is carried out by several browsers of different kinds.

3.2 Structure of the web site

The Sunday Times web site has four major inter-linked components: the Home Page, the current issue of the newspaper, a compressed version of the current issue and the archive. This is made available for downloading by FTP by readers who wish to read it at their leisure, without incurring the per-minute Internet access charges. The structural organisation of the web site structure is shown in Figure 3.

3.3 The weekly issue

Each week’s issue is organised into a set of sections. The issue’s home page displays a summary of the
week's issue, and provides links to all the other sections. Efforts are made to organise the sections in a similar manner to the organisation of the printed issue. Therefore, the main sections are:

- the front page
- the business section
- the general news/comment section
- the sports section
- the editorial/opinion section
- the "Plus" section

Figure 4 illustrates the organisation of each weekly issue. Each section begins with a list of titles of the articles contained in that section. Each title links to the actual article, which may be in the same file or in a different file. In addition, each section has a prominent link to the next section in the logical order shown in Figure 4. Further, every section has a set of menu-like links to every other section, so that if a reader does not wish to follow the logical order, he can browse the issue according to his preferences. These links are not shown above to maintain clarity.

Attempts are made to limit the size of all files to under approximately 15 kbytes in order to avoid excessively long downloading times. Hence, the number of files in each section varies according to the number and the size of articles therein. However, the number of files is transparent to the reader, as links are provided from each file back to the list of contents of that section.

Each page in each section contains a link to the section's archive. Therefore, a reader, having read a section or an article in the current issue can browse previous articles in that particular section. This is described in more detail in the following section.

3.4 The Archive

The Sunday Times archive is used to collect past issues in an organised and an easy-to-access manner. It is envisaged to preserve the archives for a period of at least one year.

The archive is accessed via a set of index pages. Each index page lists the articles in that particular section from all on-line issues up to and including the current one. These index pages are available for all the sections as well as some of the popular columns in each section. The organisation of the archive is shown in Figure 5. The set of index pages serves as a guide to each section in the previous issues, providing a summary of the contents of the particular section of each issue. Hyperlinks from the dates in these indexes carry users to the corresponding section in the issue of that date. Hence, the archive does not have reproductions of the previously published material, but simply consists of a set of hyperlinks categorised by section, to the original material.

3.5 Comparison with Other On-line Publications

A number of other on-line publications, some of which were mentioned in Section 2.5 were studied during the initial design stage of "The Sunday Times on-line edition. Following are some of the observations made, and their influence on our final decisions.

Some newspapers attempt to make the online version appear similar to the print version. For example, "The Times Fax/New York Times [NYT] is formatted in Adobe Portable Document Format (PDF). Documents encoded in PDF. Although preserving the appearance of
the newspaper to a certain extent has its advantages in attracting readers who are familiar with the print version, it does not fully utilise the interactive features available on the Web. Therefore, it was decided to structure the Sunday Times on-line edition similarly to the printed paper, but not preserve the actual appearance beyond reproducing the title graphics.

One approach to help readers with their information load is to offer a headline page, and a short summary of each article, with links to the full story. News wire releases from Associated Press [AP], and Reuters on Yahoo! [YAHOO] as well as several newspapers mentioned in Section 2.5 take this approach. This was seen as an excellent feature. However, due to our lack of journalistic background, it was not possible to implement this feature immediately. However, a listing of headlines of each section with links to the corresponding article was implemented at the initial step, to be later expanded to include a summary of each article.

Another important consideration is the availability of back issues, which is largely missing in most online news systems on the Internet. Most on-line issues are kept only for a very limited time period. An exception is The Hindu [HINDU].

Searchable archives was another useful feature observed in on-line publications such as the Nando Times [NANDO] and The Hindu [HINDU]. It was decided to implement an extensive archive and index the material according to issue as well as section. A search engine will be implemented at a later stage to search for information using these indexes.

4.0 Implementation

4.1 Initial implementations and problems encountered

The first phase of the implementation of the web site was done by scanning articles from the printed newspaper. This presented severe limitations in speed, and instead of the full implementation described above, only a limited number of articles were published.

In the second phase, a set of diskettes containing the articles were obtained from The Sunday Times was obtained, and these electronic versions were formatted and published on-line. The initial problem encountered during this phase was the conversion of text and images from Macintosh formatted diskettes provided by The Sunday Times to the formats used by the DOS and Unix environments in which the electronic publication was prepared. This problem was solved after several utilities for format conversion were found in Internet software archives. At this point, the full implementation as described above was found to be feasible.

However, two other important aspects to be considered were the quick formatting and uploading of the weekly issue and maintaining the accuracy and the uniformity of the issues, with a minimum of manual work. To this end, a set of templates for each issue were developed and tested, and a set of automation tools were developed. These are described in the following sections.

4.2 Templates and their use in implementation

A template for a web page consists of a file containing the required title etc., banner graphics and one or more menus of hyperlinks linking each template to others in the set. It also includes clearly marked insertion points for the date and page contents which are updated each week. Once a template is available, inserting the content article into it completes the web page.

A set of such templates for each page in the structure shown in Figure 4 were designed, and the interconnecting hyperlinks were exhaustively verified. In addition, a set of archive index pages were also prepared.

For each weekly publication, a copy of the set of template files is made, and the proper date and material are inserted. As all the necessary links and graphics are in place, this enables the completion of the weekly material without errors and a minimum amount of time, and ensures uniformity of the issues.

After the weekly issue is completed, uploading of the archive index is carried out by inserting the summary of each section into the appropriate index page, accompanied by the appropriate hyperlink.

4.3 Automation tools

The first tool developed was for HTML formatting of articles, particularly to insert paragraph and heading tags at the appropriate positions. A flow-chart depicting the functions of this tool is shown in Figure 6(a).

The second tool was for updating of the archives. With this tool, the titles of the articles in each section are extracted and are inserted into the appropriate index file, together with the date and a link to the day’s section. This is depicted in Figure 6(b).

5.0 Access statistics

Access statistics are extracted from a set of log files which every web server retains. Each time an object stored within it, either a file or a graphic is accessed by a browser, the server makes a number of different entries in a number of log files. These files can be analysed in different ways to extract useful and interesting statistical facts regarding the visitors to a web site.
For the first twelve weeks following the launch of The Sunday Times web site, these log files were analysed, and we present some of the results obtained in this section. The particular statistics that are presented are the total readership as a percentage of the readership of the first week, the number of accesses vs. day of the week, and the relative number of accesses for each section. We also present the web sites from which the highest number of referrals were made for The Sunday Times web site. These statistics are then analysed to see to what extent the original objectives of the project are satisfied.

5.1 Relative change in the weekly readership

Figure 7 shows the weekly readership (number of total accesses or the hit rate per week) relative to the first week of publication.

These statistics show a sharp decline after the first week, and a gradual increase thereafter. The decline can be attributed to the fact that out of the visitors who were attracted to the site due to the initial announcement, some did not find the material of sufficient interest to return. However, the gradual increase following this indicates that serious readers have started developing an interest in the site. By the 12th week, the readership was 140% compared to the initial reader-
Figure 7. The relative readership vs. week

Figure 8. Percentage readership vs. day of week
ship, and this number has been continually growing after the 12th week. Therefore, these numbers show that the web site is being increasingly appreciated by those interested in Sri Lanka.

5.2 Readership vs. day of the week

These statistics show the reading pattern for several weeks. The total number of hits per day is normalized with respect to the total number of hits for the week, and is shown in Figure 8 for weeks 1, 6 and 12.

The curves corresponding to weeks 6 and 12 show that the readership of the on-line edition has fallen into a pattern where the most number of accesses are on Sunday itself. Also, it can be seen that by the 12th week, 44% of readers have visited the site by the next day. After that, the readership declines, and another small increase is seen on the following Saturday, which may be attributed to the readers who have not had time to visit the site during the working days.

![Figure 9. Readership by section](image)

These statistics show that our objective of making news available expeditiously to people outside Sri Lanka has been achieved, and in fact, the on-line edition is read by more than half the interested people within two days of publication. This is in contrast to the two-week lag period that would normally be encountered in delivering the newspapers by post to subscribers abroad.

5.3 Readership by section

Figure 9 shows the accessed statistics sorted by section and averaged over the 12 weeks. It is seen that the most-accessed is the general news/comment. The next most popular sections are the editorial/opinion section and the front page. These three pages account for 72% of the accesses.

These statistics reveal that the Sunday Times on-line edition is mostly accessed for its news/opinion content.

5.4 Top referers

The web sites from which visitors were referred to a particular site is another interesting statistical fact. Investigation of the referer statistics shows the obvious fact that the most number of referrals have come from the Information Laboratories web site. Among the next highest referers are the Associated Newspapers of Ceylon Ltd. (ANCL) web site, and one called www.tamil.org. The former is not surprising as it is a popular site for Sri Lankan information. However, the latter reveals an interesting trend. This web site devoted to collecting information of interest to the Tamil community, and pointers to other sources of such information. It is interesting to note that a large number of visitors, presumably belonging to the Tamil community find the Sunday Times web site useful. This augurs well for the international and inter-communal credibility of Sri Lanka.

The next highest number of referrals are from the various search engines available on the Internet such as Infosseek, Webcrawler, www.altavista.digital.com, Yahoo and Excite. This indicates that there is high interest among the Internet community for news regarding Sri Lanka.

6.0 Summary

This paper details the objectives, the design and the implementation of a World Wide Web site for The Sunday Times. The major objective in launching this project was to disseminate news, opinions and features on Sri Lanka to the worldwide community in a timely manner. A secondary objective was to gradually build up a unique and an extensive news archive on Sri Lanka.

The design of the web site was discussed in detail, including the overall structure and the structure of individual elements. Tools developed for the automation of the updating of the web site were also discussed.

Finally, some access statistics for the web site were presented and interpreted in terms of how well the original objectives appear to be satisfied. It was revealed that most people access the web site and read the weekly issue within two days of publication. The statistics also reveal how the readership has changed over the first 12 weeks since publication, and what the most popular sections of the newspaper are. Statistics also show that the newspaper holds a high degree of credibility among the Tamil community worldwide.
7.0 Acknowledgments

The authors gratefully acknowledge the support given to this project by the Editor and the editorial staff of The Sunday Times. They also thank the directors of LACNet for providing the funding and for the many valuable comments regarding the design and improvement of the site. The untiring efforts of the staff of Information Laboratories (Pvt.) Ltd. for their continuing support of this project is also appreciated.

8.0 References


[USA] USA Today, http://www.usatoday.com/


