Designing a multimedia information service for the Internet and CD-ROM

T.B. Rajashekar,
National Centre for Science Information,
Indian Institute of Science,
Bangalore – 560 012
(E-mail: raja@ncsi.iisc.ernet.in)

A. Sreenivasa Ravi,
National Centre for Science Information,
Indian Institute of Science,
Bangalore – 560 012
(E-Mail: ravi@ncsi.iisc.ernet.in)

A. Neelameghan,
Institute of Information Studies,
702 upstairs, 42nd Cross, III Block,
Rajaji Nagar,
Bangalore – 560 010
(E-Mail: srels@blr.vsnl.net.in)

Abstract:

OM Information Service (OMIS) provides Web access to selected extracts from the writings and sayings of spiritual leaders, saints, seers, mystics, and scholars, and also to biographical sketches and list of works of some of these sources. Web access has been made interesting through display of related images and rendering of background music. The user and search interfaces have been designed for the novice user. OMIS is ported from a database originally available on CDS/ISIS. The article discusses the content and structure of the CDS/ISIS database, requirements for the Web and CD-ROM versions, system design and implementation for the two media. For the CD-ROM, database access is currently provided using Windows version of CDS/ISIS operating as a helper application within the Netscape browser.

1. Introduction:

A multimedia information service called "OM Information Service" (OMIS) has recently been developed at the National Centre for Science Information (NCSI) in the Indian Institute of Science, Bangalore. The project was undertaken by the first two authors, for the Ranganathan Centre for Information Studies (RCIS), Chennai. The project goal was to port a CDS/ISIS database of selected extracts from the writings and sayings of spiritual leaders, saints, seers, mystics, and scholars to the Internet and CD-ROM. It was anticipated that Internet and CD-ROM versions would assist in providing a rich information environment for self-education, for the preparation of lectures and discourses, and also for analytical and comparative

studies. Given the popularity of the Internet and CD-ROM media, it was felt that more people would be able to access this service. The CDS/ISIS database was originally prepared by A. Neelameghan. The project, partially supported by the Defence Science Information and Documentation Centre (DESIDOC), New Delhi, started in March 1997. The Web version of OM Information Service was developed first and released for Internet access in November 1997 (URL: http://144.16.72.150/~om/). A working CD-ROM version was prepared by March 1998 making use of Windows CDS/ISIS (WINISIS) and Netscape Web browser.

In subsequent sections of the paper we discuss the content of the CDS/ISIS database, requirements for the Web and CD-ROM versions, system design, and implementation for the Web and CD-ROM.

2. CDS/ISIS Database:

CDS/ISIS database consists of three inter-related databases OM, OMBIO and OMBIB. OM is the main database containing extracts from various sources; OMBIO contains biographical details of the sources and OMBIB contains bibliographic details of the related works. A brief description of these databases is provided below.

Database OM: This contains about 16,500 records (new records added to periodically) carrying extracts and quotations from the sayings and writings of spiritual and religious leaders, saints, mystics and scholars, as well as from epics (e.g. the Bhagavd Gita, the Ramayana), vedic texts, religious texts (e.g. Upanishads, the Bible, the Koran, and Dhammapada). These extracts were taken from about 800 such sources. The fields and indexing techniques for an OM record are:

Heading (R, IT 4)
Text (R, IT 4)
Source (original) (R, IT 0)
Context (R, IT 4)
Notes (R, IT 4)
Verse number (R, IT 0)
Refrain
Secondary source
Volume
Page(s)
Record identifier (IT 0)
[R = Repeating Field; IT = Indexing Technique used. IT 0 extracts the entire line from a field; IT 4 extracts each word in the text of the field except those words included in the Stopword List OM.STW.]

The size of the textual extract in field Text may range from a few words to over 1000 words. An extract may be a saying, a verse or sloka, an anecdote, a parable, or a dialogue. For an Extract record the following fields are displayed: Text, Source, Context, and Notes, (plus Verse number for extracts from Bhagavad Gita; and Refrain for extracts from the Holy Koran).

Database OMBIO: This contains brief life sketches of about 70 sources. The fields and indexing technique for a biographical record are:

Name (of biographee) (IT 0)
Date of birth
Date of death
Period
Life sketch (R, IT 4)
Remarks (R, IT 4)

The size of a life sketch may range from 500 to 4500 words. A large record is split into two or more records.

Database OMBIB: This contains selected bibliography (115) of relevant books and tracts, as a guide for further reading. The MIBIS (IDRC) record format [1] is used in inputting these records.

Integrated access to these three databases is provided through the Pascal program 'Select.Pas' [2].

3. Requirements for Internet and CD-ROM versions:

The Internet and CD-ROM versions were expected to meet the following requirements:

1) Extracts should be searchable using keywords (e.g. 'meditation', 'salvation') or a combination of keywords (e.g. 'humility' and 'devotion'; 'liberation' or 'salvation' or 'moksha'),
2) Users should be able to select the keywords by browsing the keyword dictionary,
3) Limit the extracts retrieved from keyword search to one or more sources (e.g. 'Vivekananda', 'Jesus Christ'),
4) Facility for browsing and selection of source names and display of brief biographical sketch of these sources (e.g. 'Sankara'),
5) Facility to select source names and display list of works related to these sources,
6) Display of source pictures (images) wherever possible,
7) Selection and play of background devotional music, and
8) Support random display of extracts ('An extract to reflect on...').

A primary requirement was that the user interface and search interface should be designed for novice users. This meant that navigation within OMIS has to be based on selection of options from easily understandable and clearly specified choices and hypertext links. For searching it was essential that the user be freed from the need to understand Boolean logic and operators. This meant that the search screens have to be designed to hide the underlying Boolean operations.

It may also be noted that the system was designed to disallow users from downloading all extracts from one or more sources, for example, of Vivekananda. Use of keywords from the text of extracts is the only way a user was expected to search the database and then combine the search results, if required, with one or more sources. Thus, it would be possible for a user to get all extracts related, for example, to 'renunciation' and 'salvation', and then limit these to sayings of 'Vivekananda' and 'Ramakrishna'. The system was not to allow retrieval of all extracts from the sayings or writings of a single source.

4. System Design:

For the Internet, we chose the World Wide Web as the medium given its popularity today for most intranet/Internet applications. Web browsers like Netscape and Internet Explorer provide a very friendly end-user interface. Moreover, HTML affords quick and easy user interface design. The Web also supports easy incorporation of multimedia elements. Availability of many public-domain authoring tools for HTML and multimedia was also a favorable factor.

First design task was to fix the functionality to be supported by OMIS. On the Web a user generally expects to obtain a view of this from the home page, which typically acts as the main menu or 'table of contents'. Based on the requirements mentioned above, we decided that a user should be able to do the following from the home page of OMIS:

- Obtain information regarding the content owners, sponsors and designers
- Obtain introductory information about the service
- Obtain details regarding its usage
- Access the database of extracts, sources and related works
- Select background music
- Obtain a random extract, and
- Send feedback to the designers.

Some of the other key design decisions we made included the following:

1. The user interface will comprise a sequence of self-explanatory HTML forms.
2. Search and retrieval using CGI programs. We decided to adapt a set of programs developed earlier in the Centre for handling bibliographic databases conforming to ISO 2709 format [3] which is also the record structure used by CDS/ISIS.
3. Retain the three-database structure used in the CDS/ISIS version.
4. Support selection of keywords and source names through browsing of keyword dictionary and list of source names.

5. Support pictures as compressed JPEG images as embedded hypertext links within HTML pages.
6. Support background music as compressed, streaming Real Audio files.

Based on these criteria we designed the screen layouts for the home page and the query forms. It was important to provide a way for the user to always view the main menu and select appropriate option, irrespective of the stage of his interaction with the system. This would not only help the user to overcome the 'disorientation' problem commonly associated with hypertext systems, but also provide him/her the flexibility to jump to a different function of OMIS.

For the CD-ROM version of OMIS we decided to retain the Web design and facilitate its access using browsers like Netscape and Internet Explorer. We were aware of a few offline Web browsers that provide access to local HTML content on the hard disc or CD-ROM. However, as we realised later, this proved to be lot more difficult due to limitations of Web technology to support local database searching. This is discussed in more detail in a later section of the paper.

5. Implementation for the Web:

OMIS is implemented on a Sun Netra Ultrasparc server, named 'smart', running Netscape Web server. The server is a node on the NCSI LAN which is connected to the IISc campus intranet which is a FDDI network linking about 40 departments and centres. The intranet has a dedicated Internet link to VSNL, the primary Internet service provider in India. OMIS can be accessed from anywhere on the Internet using the URL http://smart.ncsi.iisc.ernet.in/~om/ in domain name format or http://144.16.72.175/~om/ in IP number format. We discuss below different aspects of OMIS implementation.

5.1 OMIS home page: We have made use of 'frames' feature supported by HTML to divide the OMIS home page screen into two parts. The left part is a static screen providing the 'table of contents' or menu of options available in OMIS. Content on the right side screen changes depending on the option selected in the left-hand screen or by selecting an option from choices given in the search screens appearing on the right hand side. A snap shot of OMIS home page is shown in Figure 1. Among other things, the home page provides details of organizations involved in the development of OMIS.

5.2 Database access: As mentioned above OMIS retains the three-database structure of the CDS/ISIS version, but this is made transparent to the Web user. Upon selecting this option, the user is given the choice of searching the extracts using keywords or selecting a source to obtain biographical and bibliographical details. These options are implemented as a series of self-explanatory HTML forms. Both in keyword-based searching of extracts or selection of sources, the strategy is to allow the user to browse the dictionary of terms (or of sources) by entering first few letters (e.g. 'medit' for meditation or 'vivek' for Vivekananda) and select appropriate terms or a source. Figures 2 and 3 show snap shots of term and source dictionary browsing and searching. Within keywords searching, the system supports selection of as many terms as possible (Boolean 'or' operation) and also to specify the terms, which should be present in the search results (Boolean 'and' operation). Furthermore the user can limit the search results to one or more sources (Boolean 'or' within source selection and 'and' between the search results of keywords and sources). The system provides hypertext links to source names in the extracts display screen which can be clicked to obtain biographical and bibliographical details of the source, including display of the image of the source concerned, if available in the system. A screen shot of the extracts display screen and source

display is shown in Figures 4 and 5 respectively.

OMIS also supports an interesting feature ‘An extract to reflect on...’ which can be selected from the table of contents (left screen) and clicked by the user. Every time the user clicks on this link, the system displays a randomly selected extract. This is implemented through a CGI program, which randomly selects an extract from the database, using system date and time for generating the random extract number. Screen shot of random display output is shown in Figure 6.

Following are some examples of different types of queries the system can respond to:

- Texts of what Saint John of the Cross and Soren Kierkegaard have said about 'control of the senses' or 'purity of heart'.
- What says the Bhagavad Gita about Karma? Interpretations of the relevant texts by Anne Besant, Swami Vireswarananda and S. Radhakrishnan.
- The relation between 'Salvation' and 'Renunciation' as perceived by Swami Vivekananda, Sri Ramakrishna Paramahamsa, Sri Sankara, Sathya Sai Baba, Meister Eckhart, Thomas Kempis and other spiritual leaders, and mystics.
- Biographical sketch of Prophet Muhammad.
- An inspirational verse or saying.
- List of selected works on/by the Sikh Gurus and Sufi Saints of India.

5.3 Conversion of files from CDS/ISIS to Web: We have used inverted file structures for handling searches on keywords (terms taken from the text of extracts) and source names. As it is important to retain the print layout of extracts in the CDS/ISIS version, many of which are in verse form, the OM database on CDS/ISIS was printed into text files using the ISIS print format. This was carried out for OMBIB and OMBIO databases also. From these text files we extracted the terms from fields to be indexed (source names and keywords from OM and source names from OMBIO and OMBIB databases). These files were sorted and then used for building up the index files, with term postings pointing to appropriate records in the text files. A set of CGI programs in C complete the software part. These programs are linked to the HTML search forms and interact with the index files and text files to carry out search, retrieve and display operations. Once again, to retain the original display format, we embedded the HTML output within <PRE> and </PRE> tags. This aspect is brought out in Figures 5 and 6.

Figure 1. OMIS Home Page (http://144.16.72.175/~om)

Netscape - [OM Information Service]
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Edit Reload Images Open Print Find Stop
Netsite: http://144.16.72.175/~om/
Introduction By C. Subramanian
Overview of OM System
Access OM Database
An Extract to Reflect on...
Select Background Music to Play
Comments and suggestions to:
webmaster@ncsi.iisc.ernet.in
OM Information Service
Extracts from the writings, utterances, discourses, poems, etc. of over 800 scholars, spiritual leaders, mystics, saints, literatteurs, and selected epics.
OM Information Service is prepared for the Ranganathan Centre for Information Studies, Chennai, India, by A. Neelameghan.
CD-ROM and experimental mounting on the WWW by the National Centre for Science Information (NCSI), Indian Institute of Science, Bangalore, India.
Partial Financial Support from the Defence Scientific Information and Documentation Centre (DESIDOC), DRDO, New Delhi, India.
[Note: This system is best viewed with Netscape Navigator 3.0 or above with screen resolution
Document: Done

Figure 2. Browsing term dictionary

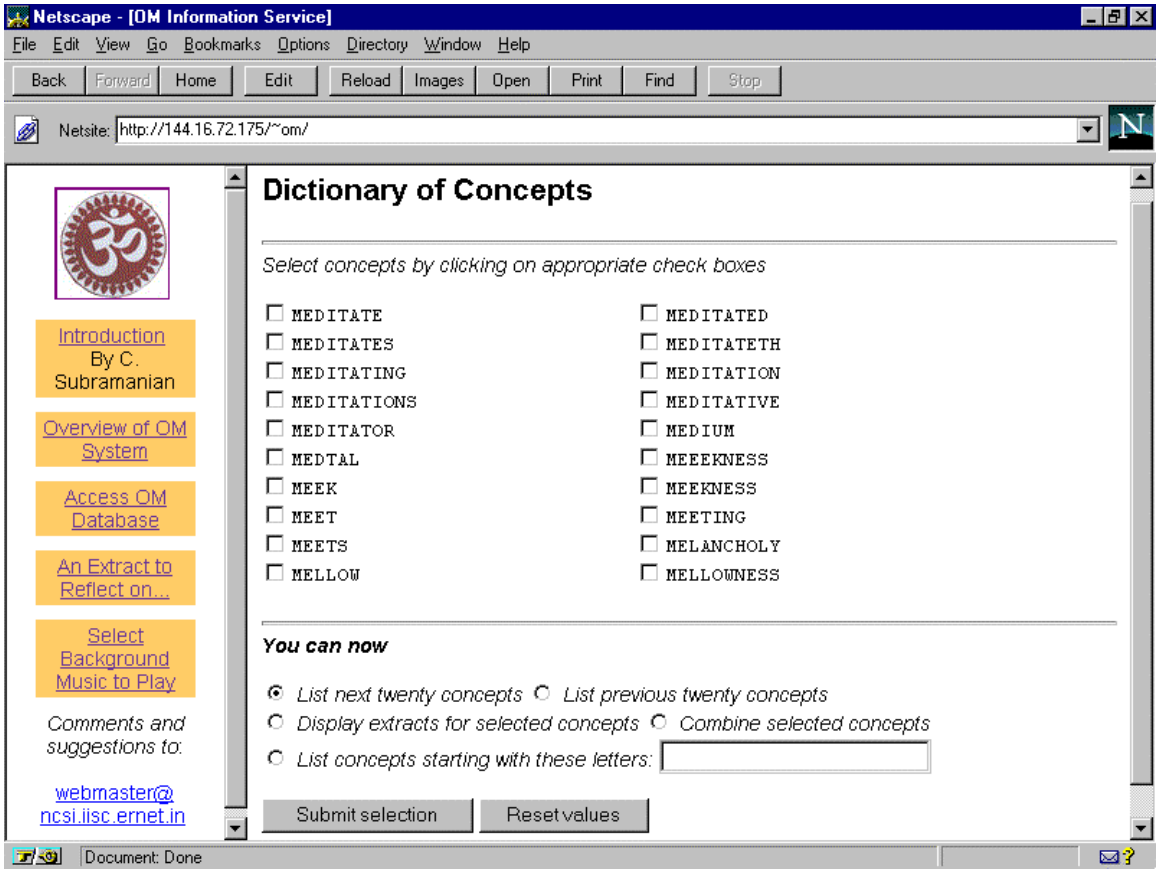


Figure 3. Browsing source dictionary

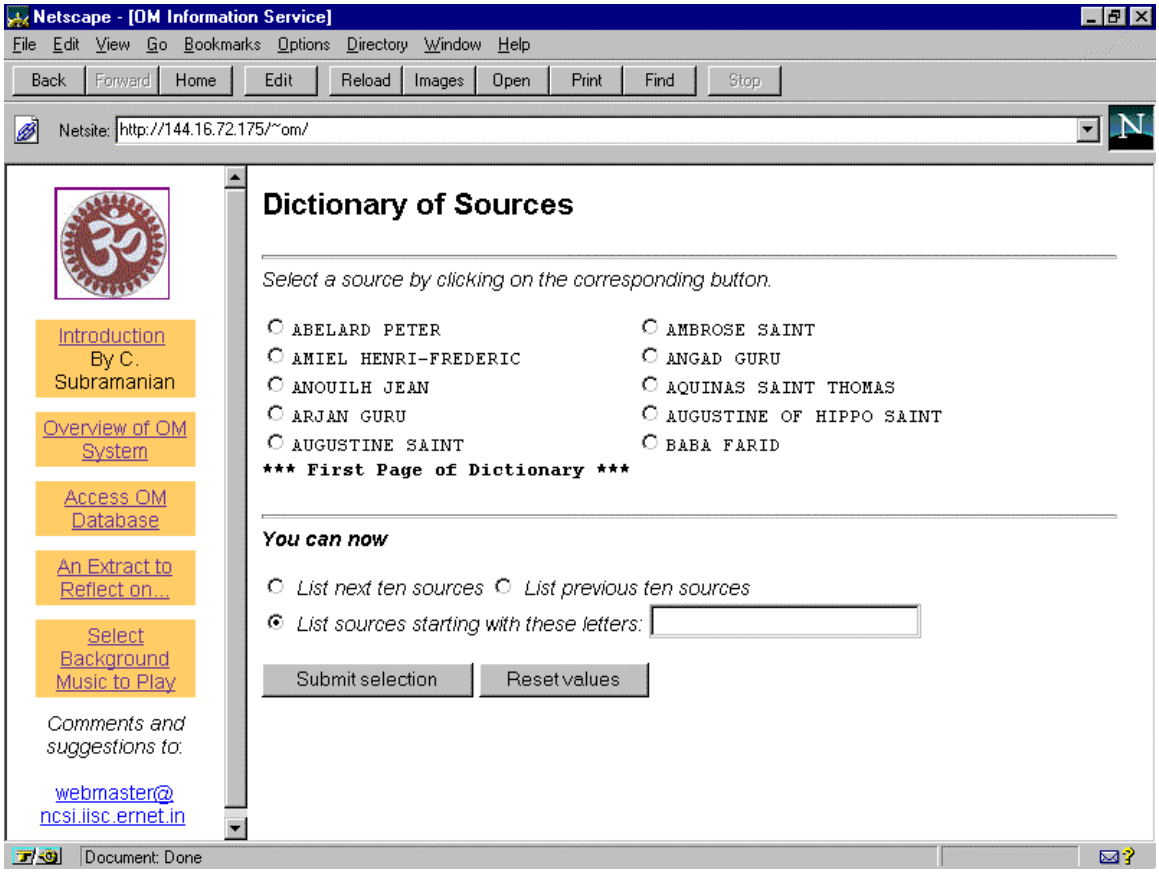


Figure 4. Display of retrieved extracts

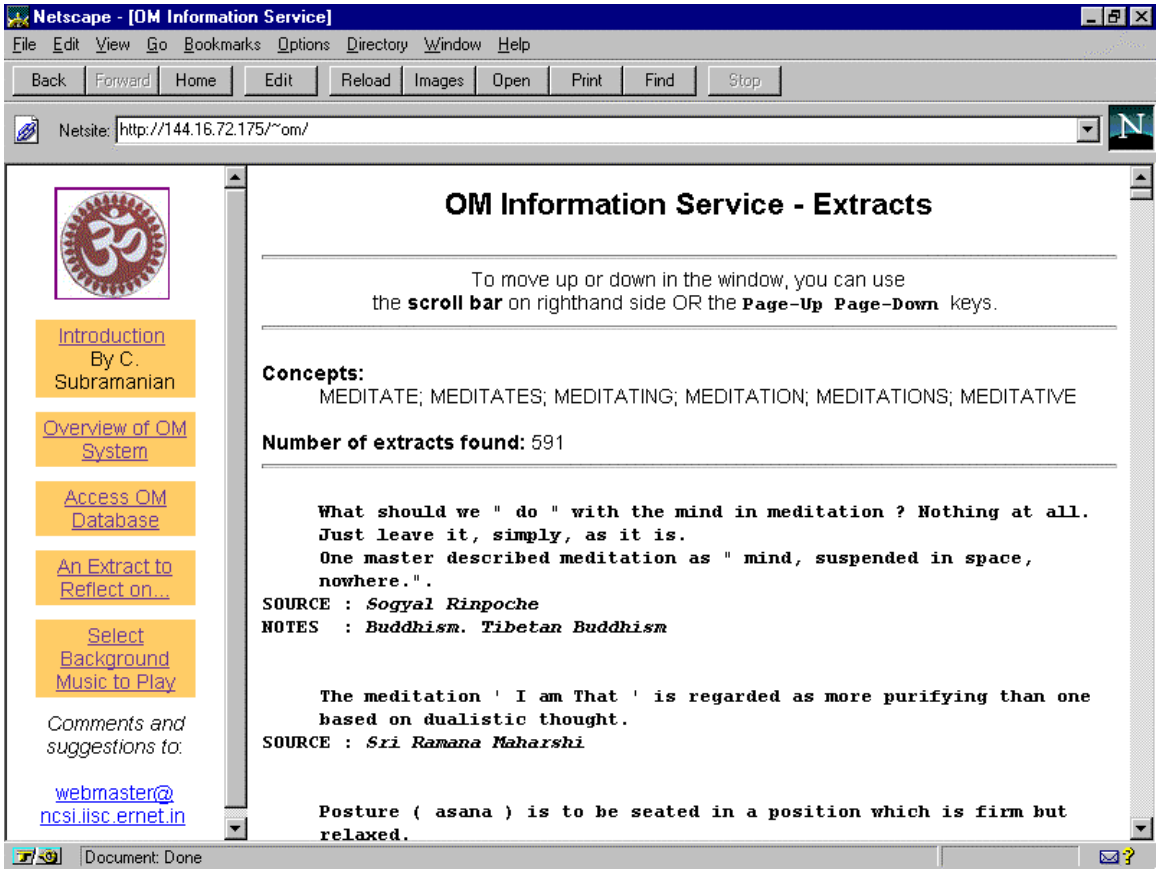


Figure 5. Source information display

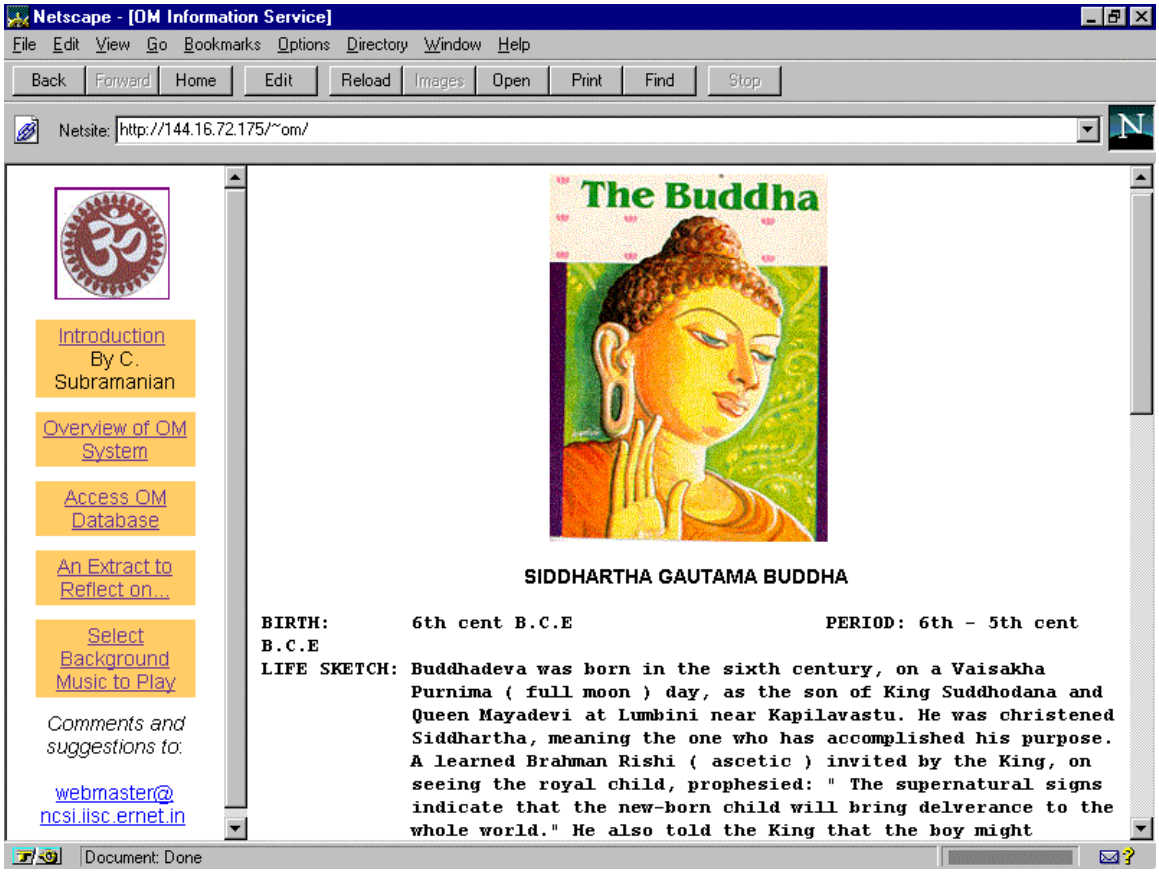


Figure 6. Display of a random extract

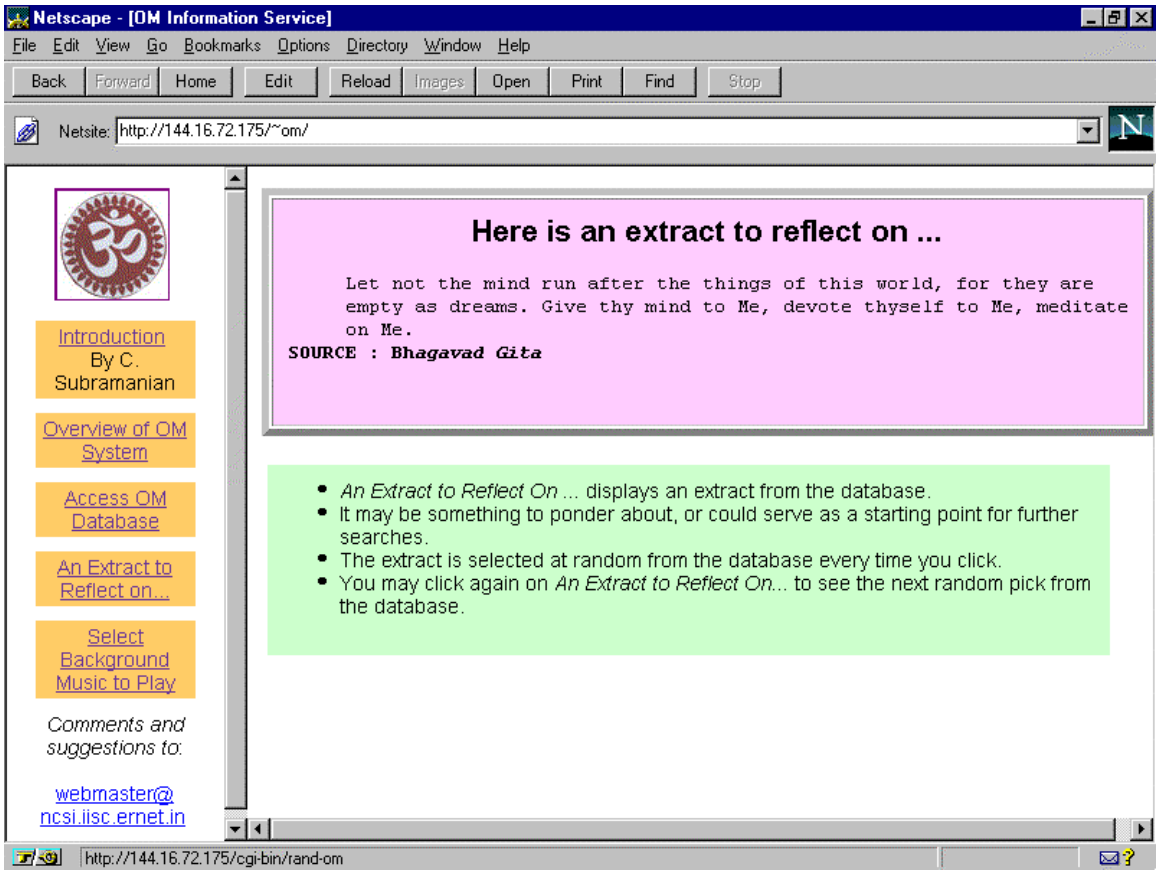
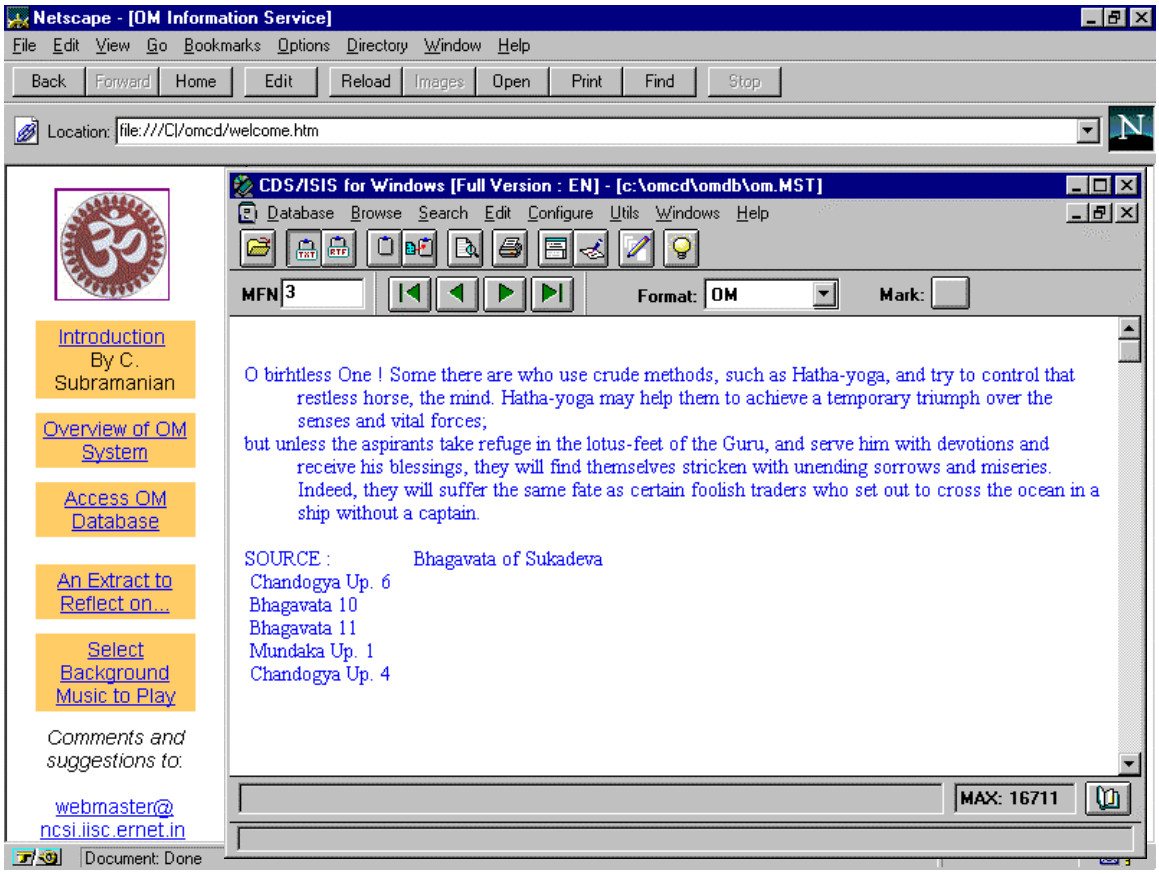


Figure 7. Windows ISIS as helper application



5.4 Disc storage: Currently OMIS occupies about 100 MB of disc storage on the Web server. Of this, about 40 MB is database files, 55 MB is audio files (8 pieces of devotional music and 1 speech by Bharata Ratna Sri C. Subramanian), 2 MB is images (about 60 JPEG files, mainly pictures of sources), and about 60 HTML files occupying about 0.5 MB.

5.5 Images: Most of the images are pictures of sources, a few are places of worship (temple, church, etc.) and religious/historical events (e.g. Geetopadesha from the epic Mahabharata). All of these were scanned on a HP 4C-colour scanner using Adobe Photoshop (version 4.0). Using PhotoShop we were able to fix the physical size of images to 3" x 2" and to store them in JPEG at medium quality level. This enabled us to keep the file size of each image file to less than 30 KB.

5.6 Audio: OMIS facilitates a user to select background music to be played while he/she is accessing the database. This selection can be changed at any point during interaction with the system. Upon selection of the option 'Select background music' from the menu on the left screen, the user is given a choice of music pieces to select from. When a piece is selected the system opens a separate window, displays a brief decryption of the music piece, before starting to play the music, in background. This is controlled through a script written in JavaScript, embedded within the music selection page.

We used the shareware version of Goldwave to capture audio files from audio CDs and cassettes. One piece ('Gregorian chant') was downloaded from the Internet. We observed that a typical 5 minutes music, captured at 22 KHz in 16-bit stereo, would occupy about 5 MB, in .WAV or .AU formats. Capturing at higher sampling rates would have resulted in much larger file sizes. Although browsers like Netscape and Internet Explorer have in built support for playing audio files in this format, files of this size are extremely slow to download and play on the Internet, even at 28.8 KBPS speeds. To handle this problem we used the free version of Real Audio encoder (available on Real Audio web site http://www.real.com/) to generate compressed .RA files from the .WAV and .AU files. Thanks to compression techniques used by Real Audio, the resulting .RA files are about 1/10th in storage size, without too much of a compromise on the audio quality. The audio quality of encoded RA files are directly dependent on the quality of the WAV and AU files. For this reason we set the Goldwave capture parameters to 16 bit stereo and sampling rate to 44 KHz. To enjoy Real Audio the user, however, is required to install Real Audio player plug-in on his/her browser, enabling the browser to interpret and play he .RA files coming over the Internet. Fortunately the player can be downloaded freely from the Real Audio Web site. Performance of Real Audio can be further enhanced using the streaming technology. Since we could not afford to purchase and implement Real Audio streaming server for this project, we used the cheaper alternative of HTTP streaming, which enables reasonable fast loading of audio files to the client machines.

5.7 Supporting continuity in search interaction: This was a key implementation issue. HTTP protocol used for WWW on Internet is a stateless protocol - neither the browser nor the server retains information about previous transactions. For example, in a search session each submission of a search form consisting of keywords or source names is a separate transaction and there is no way the user can tell the server, for instance, to combine this with the results of a previous search. This feature is common in online and CD-ROM database search systems. On the Web this feature has to be explicitly incorporated into the CGI programs. We considered two possible options for achieving continuity in a search session. One is to maintain, for each client IP number that is interacting with the server, details of the searches carried out and to use this in subsequent sessions. This would have required many disk I/Os on the server side and also cleaning up of files after the session is over. Second option was to use the 'Input Type Hidden' feature supported by HTML and to carry the details of query terms used in the HTML forms through the search session. Only disadvantage here is that the search has to be executed again for previously selected search terms, before the search results are displayed. As we did not expect users to formulate complex search expressions for OMIS we decided to use this method for maintaining continuity in search interactions.

6. Implementation of OMIS on CD-ROM:

As mentioned in the beginning, a design decision we had made was to adopt the Web solution for the CD-ROM version also. That is, all the files related to OMIS (database, html, audio, images and search

programs) will reside on the CD-ROM and a user will access the service from a Web browser residing on his/her PC. As we began implementation for the CD-ROM, we realised a major limitation that currently exists in integrating Web browsers with searchable HTML content on the CD-ROM. Web is a client-server protocol wherein browser programs like Netscape reside on client PCs and interact with a Web server (an Internet host server computer running Web server software like Apache and CERN HTTPD) on which the database to be searched and the associated CGI search programs reside. If this has to be transported to the CD-ROM platform then the user's PC, besides acting as client, also needs to run a Web server supporting CGI programs. Expecting the user to install Web server software and to configure it for accessing the content on the CD would have made the system unduly complex.

A few other solutions we considered included the following:

- Use software packages that support searching of local HTML content. Commercial packages like Verity's CD Web Publisher enable publishing of HTML content on CD-ROMs with a search engine installed on the CD for distribution. The search program can be invoked from within a Web browser. They support quite powerful indexing and search features. However, we found these packages to be quite expensive. Also, the search language is usually complex and the interface difficult to customise.
- Use offline browsers that support access to local HTML content. Offline browsers like I-View and Likse support browsing of HTML pages stored on the hard disc or CD-ROM. A few of these also support indexing and searching of HTML pages and writing CGI programs. However, offline browsers do not generally support features like Java Script, multimedia handling, frames, etc. Another key limitation is their inability to handle HTML pages at record level. They usually search on HTML pages stored in one or more directories. For our application, this meant about 16,000 HTML pages, one page for each extract!
- To code the CGI programs in Java, which can access the local storage. We could thus retain the same inverted file structure on the CD-ROM and support search through Web browsers. This would have limited the CD-ROM usage to Win'95 PCs only.

As an interim solution we decided to use the Windows CDS/ISIS (Winisis) developed by UNESCO [4], as a helper application, to be executed on top of the Web browser. Winisis is used for supporting database searching. Other aspects of OMIS, as implemented on the Web, are retained in the CD-ROM version. Winisis program is stored on the CD-ROM along with the CDS/ISIS database files, audio files, and HTML pages. The only requirement on part of the user is to define a helper application in the Web browser (e.g. under 'preferences' in Netscape) and link file types with extension '.mst' to the Winisis application residing on the CD-ROM. As the hypertext link for database access is linked to 'om.mst' file which is residing on the hard disc, when the user clicks on this link Netscape starts Winisis as a helper application. Winisis, in turn, opens the OM database for the user to carry out searches within the Winisis window. A screen shot of Winisis search window on top of Netscape browser is shown in Figure 7. The second author has recently completed testing implementations of OMIS using Java with the database loaded as tables in a SQL database [5]. We will explore the possibility of a pure Java-based solution for CD-ROM.

7. Conclusion:

Given the ease with which Web sites can be set up and Web documents published there has been phenomenal growth in the number of Web-based information sources on the Internet. It would appear that most information related activities in future will be Web-centric and Web browsers will become the common user interface for information access. Development of OMIS for the Web is in keeping with this trend. Though Web supports nice features like the design of user friendly interfaces, hypertext navigation and multimedia integration, it does have limitations when it comes to supporting powerful search and display functionalities that exist in many information storage and retrieval packages. There appear to be further limitations when we wish to package a Web-based database application to the CD-ROM. In this article we have made an attempt to discuss some of these aspects in the context of implementing OMIS on the Web and CD-ROM.

References:

1. IDRC. MIBIS Record format.
2. Sreelatha, G and Neelameghan, A. A Microisis Pascal Interface for Concurrent Multiple Databases Search and Retrieval. Information Studies Vol. 1, No. 2; April 1995; p.114-129.

3. Sreenivasa Ravi, A and Rajashekar, T.B. S.D.I software in 'C' on Unix System V Release 3: An implementation report. NCSI, Indian Institute of Science, Bangalore, February 1991.
4. Micro CDS/ISIS. UNESCO. http://www.unesco.org/ioc/infserv/cdsisis.htm
5. http://ukko.grainger.uiuc.edu/omjava/ and http://ukko.grainger.uiuc.edu/omasp/