Final Report

Development of a Biodiversity Information System

Research Project No. KFRI 363/2000

P.Vijayakumaran Nair

Kerala Forest Research Institute
Peechi 680 653

July, 2005

Project Proposal

Code	KFRI 363/2000
Title	Development of a Biodiversity Information System.
Investigator	Dr. P. Vijayakumaran Nair

Objectives

1) To design database and user-friendly search mechanism for storing information on plants and animals.
2) To compile satellite images and maps of forest areas in the region and develop software for their storage and retrieval.
3) To integrate various information such as bibliographic, flora and fauna distribution and maps.
4) To supply tabulated results, images and maps to collaborating agencies.
5) To prepare documentation for the techniques and software developed.

Outline of Research

Information technology is a field undergoing rapid changes. A few years ago design of a biological information system would have centered around a data base package with graphical interface, where as now it would be more of a web/html based mechanism. During the course of the project, technology may change again.

Design and implementation of the information system would involve the following aspects.

1) Conceptualizing of an indexing scheme involving keywords from flora fauna, maps and bibliography. As per present plan, the keywords will have indication of the data base and record number. This would be an index, sorted alphabetically so that binary searches are possible. Integration of different subjects would be achieved through this index. There will be separate index for each subject, the combined index will contain all subject indices.
2) The search will be by selecting key words from the index. Selection will be carried out in an associative way so that subject links can be made. The search will yield hit records which can be displayed in ways specific to database.
3) Data would be kept in format appropriate for the subject. There will be option for exporting the database to popular formats. A structure with variable length fields will be used for storing bibliographic information. Fixed field length structure will be sued for flora and fauna information. Database structure capable of storing BLOBS (Binary Large Objects) will be used for storing images and maps.
4) The information system will be brought out in CDROM format. There will be a graphical interface from which specific subjects or total index can be searched. Satellite images and maps will be produced in self running CD ROMS also.
5) For programming, Delphi, Visual Basic and HTML are intended to be used. Documentation will be prepared in HTML and MS Word formats.

Date of commencement	August 2000
Date of completion	December 2003
Funding agency	MoEF through IISc Bangalore

Contents

List of figures

Acknowledgements

We are thankful to Dr. JK Sharma, Director, Kerala Forest Research Institute for his encouragement and guidance. We are thankful to the Ministry of Environment and forests, Government of India for sponsoring the project and Prof. Madhav Gadgil for allotting the component Development of 'Biodiversity Information System' to us. Dr. R. Gnanaharan, Research coordinator monitored the progress of the project closely and examined the final report.

We are indebted to Ms Sarojam and Mr KH Hussain of KFRI Library for many of the ideas and suggestions in developing the bibliographic data base. Ms Kavitha Varma perfected the plant identification part.

Drs. C. Renuka, ARR. Menon and M. Sivaram offered many useful suggestions as editors for the final report.

Preface

'Development of a Biodiversity Information System' was initially envisaged as a subproject to develop teaching material for an integrated biodiversity project. Due to funding constraints the scope of the program was reduced and attention was shifted to examining the design and implementation aspects of biodiversity information system and solving specific problems. Initial attempt was to develop an information system which could store all aspects of a region such as distribution of flora and fauna, their identification, ecology, bibliographic data, maps, etc. This would in a way be an encyclopedia approach. This approach will have severe constraints for updating. The technical skill required for such a task was another constraint. Therefore focus was shifted to developing specific databases for different tasks. It was possible to devise a mechanism through which these databases could be searched together. It was soon apparent that trying to combine different databases is too ambitious a task and independent databases focusing on specific topics would be more practical. Even in the case of independent database approach there were challenges in design and implementation. Topics of immediate relevance were given priority. Lack of user friendly bibliographic software was one of the issues addressed first. A usable package could be developed for this purpose. Mechanism was devised for plotting distribution of plants and animals. A common interactive program was developed for identification of plant or animal taxa. Tools were developed for solving problems associated with satellite images. Tools were also developed for customizing three dimensional data. These packages are described individually along with their design and implementation details.

Some of the components were developed into stand alone CDROMs while others were incorporated into tool sets. The final report of the project has already been submitted to the sponsors. The present report is submitted as final report of KFRI Research project KFRI 363/2003.

Abstract

The main objectives of the project were to develop a user-friendly information system dealing with various aspects of biodiversity of plants and animals. Program for managing bibliographic references, plotting plant/animal location, program for identification of plants/animals, processing satellite images and terrain visualization are some of the topics covered. Issues relating to integration of databases are also discussed.

A program for storage and retrieval of bibliographic references is developed. The program has many features such as form based editing, single file storage, automatic generation of index, multiple keyword searching and formatted print generation. An important feature of the program is its ability to store original articles if necessary.

Identification of plants or animals is a problem that has received much attention from the point of view of computerization. Present attempt is to develop a program to identify members of taxa such as genus or family. A data matrix of character presence is used as input. The user can select identifying characters in any order. The program would at each stage show a list of members having the characters selected. Illustration and explanation is provided for selecting characters. Data of the required nature is available for certain groups of organisms. The program also gives an indication of number of character required for a singular match. There is need for plotting distribution of particular plant or animal. Several approaches are possible, including the ones based on GIS databases. In the present situation, geographic coordinates or Survey of India (SOI) topo-sheet numbers can be specified. The background map can also be specified. Data on plant distribution is available from floras. There is also provision for displaying abundance data. This program employs simple text format for maps and input data.

Graphic representation of biodiversity parameters would help in better interpretation of the subject. Methods were devised for this purpose which can generate high quality pictures. These programs can also carry out many of the data summarization and computation tasks. Traditionally satellite images are analyzed using remote sensing programs. Some of the entry level commercial programs are incomplete and many steps of data preparation and conversion have to be done by the users. Many of these programs do not support opening of Indian Remote Sensing satellite (IRS) images. A program is developed for viewing IRS images. Possible utilization of high quality satellite images

available free of cost is also described. The program has capability of extraction of parts of IRS images.

Terrain visualization is becoming a popular tool. Various GIS programs provide this facility to varying degree. Some times freeware programs designed for a specific task are able to yield better results. In this part, methods are described for preparing Digital Elevation Models (DEM) from contours and displaying them using freeware programs. Sources of high quality data available free of cost are also provided. Methods for integrating different types of databases from common interfaces, use of common search engines are considered. Such an approach is more suited for encyclopedias. For databases which are constantly updated, stand alone database with search capability is recommended.

A program has been developed for generating HTML index and link system automatically. Using this program, a collection of articles, theses or booklets can be converted into a self contained CDROM.

1. Introduction

The main objectives of the project are to develop a user-friendly database on biodiversity of plants and animals, to compile and process satellite images and to develop mechanism for storage and retrieval of maps. One of the objectives of the project was to develop a storage and retrieval system for data in different formats such as bibliographic data, maps, satellite images, taxonomic information and detailed descriptions. The programs and sample databases mentioned are included in a CD ROM along this report.

1.1 Bibliographic data

Bibliographic data is an important component of biodiversity data. Several methods were tried for storage and retrieval of this type of data. Commercial programs deal with the subject in different ways. Most of them are meant for reference collection at individual level. Bibliographic software such as CDS/ISIS, Reference Manager, Scimate, etc have severe limitations. Several library automation packages attempt to process bibliographic search and retrieval along with other tasks such as book lending and purchase in a library. Available software has limitations in terms of field length, search ability, and facility for keeping long fields such as abstracts and facility for linking documents. A program for storage and retrieval of bibliographic references is developed. The program has many features such as form based editing, single file storage, automatic generation of index, multiple keyword searching and formatted print generation. An important feature of the program is its ability to store original articles if necessary.

1.2 Taxa identification

Identification of specimens is an important item in any biodiversity study. Conventional method involving dichotomous keys is very laborious and time consuming. A software is developed for interactive identification of plants or animals. Identifying characters can be specified in any sequence. The program was further generalized to be suitable for any taxa such as genera or families. This approach has many benefits and some defects. Benefits are that the user can start identification from any character. Computer could eliminate items that do not posses the selected character. The computer could also be made to show illustrations and details of traits selected. On the negative side, a presence / absence matrix of all the characters has to be made in advance. A dichotomous key would possess only values along the diagonal cells. Traditionally such full

data matrix is used in identification of microbes and in these cases the mechanism would be easy to implement. Sample cases have been built up using a variety of data.

1.3 Distribution of plants and animals

There is often need for plotting distribution of particular plant or animal. Several approaches are possible, including the ones based on Geographic Information System (GIS) databases. In the present situation, geographic coordinates or Survey of India (SOI) topo sheet numbers can be specified as location. The background map can also be specified. The data file and map has simple structure description which is included. Data on plant distribution is available from floras. There is also provision for displaying abundance data along with distribution. There is provision for generating good quality maps. The program employs simple text format for maps and input data. The same task can be carried out using GIS programs. The present module came out handy in preparing stand alone data entry/display package for specific species.

1.4 Graphical representation of biodiversity

Graphic representation of biodiversity parameters would help in better interpretation of the subject. Methods are devised for this purpose which can generate high quality pictures. These programs can also carryout many of the data summarization and computation tasks. Even though multiple values can be represented by bar diagrams or line graphs, it is possible to devise special display techniques which bring out the importance of different parameters readily. Proportion of sand, clay and silt in soil can be shown in this way. It is also possible to compute a new parameter from existing parameters. For example Importance Value Index (IVI) is computed in this fashion. In these displays, two major parameters are made to vary along x and y axis. The property values are properly scaled to draw polygons. Depending upon the magnitude of the values, the polygons take different shape and positions in the graph. Association between different samples also become clear in this way.

1.5 Satellite images

Satellite images are a field with a great deal of developments. Our country has been able to make rapid strides in the field of acquisition of satellite images. We are still dependant on proprietary software for processing. Many of these software has not sufficient provision for reading data from Indian

satellites. Attempt is made to develop software for processing of these images and to carry out necessary data conversion. NASA has put images from Landsat series in public domain. Available images for Kerala have been repackaged in ready to use form. NASA has placed Digital Elevation Data (DEM) also in the public domain. These images are also compiled and bundled with non proprietary software for processing.

1.6 Terrain visualization

Terrain visualization is becoming a popular tool. Various GIS programs provide this facility to varying degrees. Sometimes freeware programs designed for a specific task is more easy to use and yield better results. In this part, methods are described for preparing DEMs from contours, displaying them using freeware programs and about sources of high quality free data available for the task.

1.7 Map compilations

Maps are prepared using standard GIS package such as Mapinfo. These maps are exported to bitmap or .PDF format for making searchable compilations. Two attempts were made to make a collection of maps. In the first case section level maps of Periyar Tiger Reserve were combined into a CD. In the second case, forest atlas of Kerala was compiled.

1.8 Integration mechanisms

Methods for integrating different types of databases from common interfaces and use of common search engines are considered. It would be often desirable to click on parts of a displayed picture to obtain further details about the topic. Depending upon the platform involved, web based, Graphical User Interface (GUI) based, etc, the mechanism to be used would vary. A program has been developed for generating HTML index and link system automatically. Using this program, a collection of articles, theses or booklets can be converted into a self contained CDROM.

1.9 Review of literature

Several methods are available for bringing together information. Depending upon whether the information system is implemented in a CD ROM or web site, appropriate methods will have to be used. The system can be based on Hyper Text Markup Language (HTML) or other suitable format. MS Office formats and .PDF can also be used in some parts.

2. Materials and Methods

Two sets of programs were used in the present project. The first set consists of programs the author had written. These were written using the Rapid Application Development (RAD) tool Delphi (Delphi 2003). The second set consists of public domain programs. In the former case design and operation details are described, in the case of the later, only operational details.

2.1 Bibliographic data

Since none of the commercially available software are fully suitable for maintenance of personal reference collection, a software was created for storing and display of bibliographic information. The following were the design specifications 1) Software is designed using MS Windows user interface and can be run under Windows 98, XP or above. 2) Database consists of a single file. Index will be generated internally when required. 3) Necessary database fields are already predefined. Fields are provided for author (multiple), title, year, key words, reference, address, link and abstract. These fields have practically no length restriction. Original articles can be stored if necessary. 4) Search capabilities are built-in. Technical details are described below. File structure and search engine needed considerable effort to develop.

File structure

Variable field and record size impose many constraints on limiting database to a single file. The program uses several files internally, but the details are hidden from the user who operates on a single file. Index files are generated at the time of opening the database.. Ten predefined fields are used in the database. They are

1) Author: Multiple authors to be separated by **%** symbol. Each author will be included in the index.
2) Title: Title of the article. Key words have to be marked using angle brackets <KW>.
3) Year.
4) Reference: Publisher in case of books, name, volume, issue and pages in case of journals.
5) Key words: To be separated by % symbols. All key words will be used for generating index.
6) Address: For address of authors.
7) Link: This can include reference to a linked file. The linked files have to be kept in the folder \links.

8) General purpose field 1
9) General purpose field 2
10) Abstract: Key words have to be marked using angle brackets <KW>.

Data entry/Edit: A work sheet appears for inputting references. Facilities are provided for storage of frequently used text. Sample work sheet is shown in the results section. Input collected using worksheet is formatted and stored as one record and added to the database. Records for editing can be invoked by record number. Required record is opened and loaded into work sheet. During the save operation, constituent files are combined into single database file with extension .BIB

Indexing: Index consists of key word and record number, sorted separately alphabetically and in the order of records. Index includes each and every author, key words in title, additional key words and key words marked in abstract.

Export /import: The database can be exported to a text format. Facility is provided for importing text in delimited format and few other popular formats.

Search engine: Search engine developed for the purpose works by the criterion of examining keyword (s) in the record. Two approaches were tried, list based selection of key words and hierarchical selection. In the package both methods can be used. In the first method, the user selects key words from a list. If multiple terms are selected, they are combined into a search expression. In the second method, once the first key word is selected, subsequent entries will be from records containing the first key word. Binary search method is used for searching the index files.

Hard copy generation: Formatted output with highlighted author and reference is provided.

Technical details

The following highly technical programming and file management aspects are meant for developers attempting to design bibliographic databases. Bibliographic data brings up several challenges to the programmer. One could easily write a program using fixed size fields as was done during the days of Dbase/Foxpro. This would have difficulty when there are a large number of authors or when there is long abstracts. In some programming languages, it may be difficult to use fields longer than 256 characters. One way of over coming this is using large field size. Even when this is feasible, the database would contain large blank areas there by increasing the file size. It would also require special effort to access authors other than the first one. The second approach is

to create the fields using variable size fields with delimiters and store them as binary objects. This is feasible, but the burden of storage management and record generation would fall on the programmer. Programs such as CDS/ISIS use a variable size field approach. In this case two files are employed. One file stores the beginning of individual records and the second file contains actual data. At the time of editing, this would create enormous complexities as the modified text has to be accommodated in the initial area. One way would be to append the modified record at the end of the data base. This would also increase the file size unnecessarily. Such an approach was tried in the present case also. Though initially successful, the complexity made maintenance of the code difficult. The latest version uses an altogether different mechanism. The database is divided into chunks of 1000 records and stored in subfolders. Each record is stored as separate file. The record number can be derived from the name of the subfolder and individual record file names. The user sees the whole database as single file because, during saving all the records are copied to a single file. The process is reversed while opening the database. This mechanism was unthinkable in the DOS or early Windows period as PCs lacked speed and capacity to handle very large number of files. The present mechanism is usable if the data base contains few tens of thousands of records. With more records, system constraints may appear. In short, the present mechanism is suitable for personal level databases of medium size. For more ambitious tasks, professional programs will have to be used

2.2 Taxa identification

Identifying characters of taxa are coded as a matrix. The user interface is divided into four boxes, one for character list, another for display of illustration and explanation of character and yet another for displaying the list. The character list is loaded as a list of check boxes. As the user checks, the characters, matching items are listed. Common data input mechanism can be worked out. Sample data and working is described under the results. Sample user screens are shown in the results section.

Technical aspects

The user interface employs standard components only. Search and matching algorithms are also straight forward. Some amount of character weighing could be achieved by simulating selection of remaining characters, one by one, behind the scene by the computer. This would identify candidate features selecting which the steps in pinpoint identification can be reduced.

2.3 Distribution of plants and animals

Distribution data is supplied as a text file with name and location details. The display box is designed in two sizes, low and high resolution which differ only by the size of the display box. The program displays the background map to fit the display window. The input file is read line by line and location marked by symbol. For the map, text format of Mapinfo can also be used.

Technical aspects:

The background map is a feature encoded vector text file. Maps in the Mapinfo Interchange Format (.MIF) can also be used for the purpose (Mapinfo, 2003).

2.4 Graphical representation of biodiversity

A program is developed for displaying four parameters as a multi-dimensionally scaled diagram. The program can summarise enumeration data from sample plots. This produces visibly different appearance for even slightly different data and helps to interpret differences and characteristics. The values are scaled to occupy a fixed maximum size on the screen and the remaining values scaled in proportion to this. Data file can be supplied as a delimited text file. Working of the program is illustrated in the results section.

2.5 Satellite images

A program was developed for keeping set of satellite images together and for selectively displaying required areas. Two modules were written for displaying IRS IB and C images (NRSA, 1995). The program reads the header portion of IRS images and extracts bands. Red-Green-Blue (RGB) display of extracted bands is made using histogram equalization option. The program reads the header data of images and displays on screen. User is allowed to select required regions interactively.

Technical aspects

The program deciphers row, column and band combinations from the header and copies required regions in usable format. Required region of each row is written to individual bands. The images written to Idrisi format in byte binary format with ASCI header file (Eastman, 2004). User interface permits selecting data file and output location.

2.6 Terrain visualization

The main source of data is Shuttle Topo Radar Mission (SRTM) data of NASA. The files downloaded from the internet can be displayed using freeware such as 3DEM (ref.). The author had digitized 1:1000,000 scale topo sheets to produce DEM of Kerala of about 500m resolution. Later NASA data for whole India at 1000m and 90 m data as degree tiles for the Kerala region is included in a set of CD containing DEM data. The CD cannot be made self contained as the freeware programs used for displaying the images do not have an external integration mechanism. The freeware programs are also contained in the CD, which the users may run. Overlay files for the degree tiles are provided. These over entire extend of degree tile there by overcoming need for geo registration. Tools are developed for splitting and joining of DEM files.

2.7 Map compilations

Most of the maps are prepared using standard GIS package such as Mapinfo. These maps are exported to bitmap or .PDF format for making searchable compilations. Two attempts were made to make a collection of maps. In the first case, section level maps of Periyar Tiger Reserve were combined into a CD. In the second case, Forest Atlas of Kerala was compiled. The maps in bitmap or .PDF format are copied to folder. An HTML file with links to individual files in a hierarchical manner is created using package such as MS Word. This file is referenced in an auto run file so that the CD containing the files in the folder would auto start when inserted into a PC. Listing of html and auto run files are provided in appendix. Links are given with out reference to folder name so that the mechanism will work even when copied to a folder in the hard disk.

2.8 Integration mechanisms

There is no standard method for integrating the databases. The mechanism employed in encyclopedias is not suitable for the task under consideration. Working of a search and display mechanism developed is described in the results section.

3. Results and Discussion

3.1 Bibliographic data

A software was designed and successfully implemented for storage and retrieval of bibliographic data. In the biodiversity context, literature would consist of floras, handbooks, field books, monographs, keys, forest management plans, etc. The main features of this program are 1) Software package consist of a single .EXE file. 2) Database consists of a single file 3) Fields such as author(s), year, title, reference, key word(s), abstract and link are already provided, users do not have to define data base structure. 4) The database is designed as variable length records, enabling entry of records with several authors or long fields such as abstract. 5) Key words can be marked in title or abstract. 6) Links can be given to full document in HTML, MS Word or PDF format. 7) User interface consists of browsing records, work sheet based data entry/edit and search engine based data retrieval.

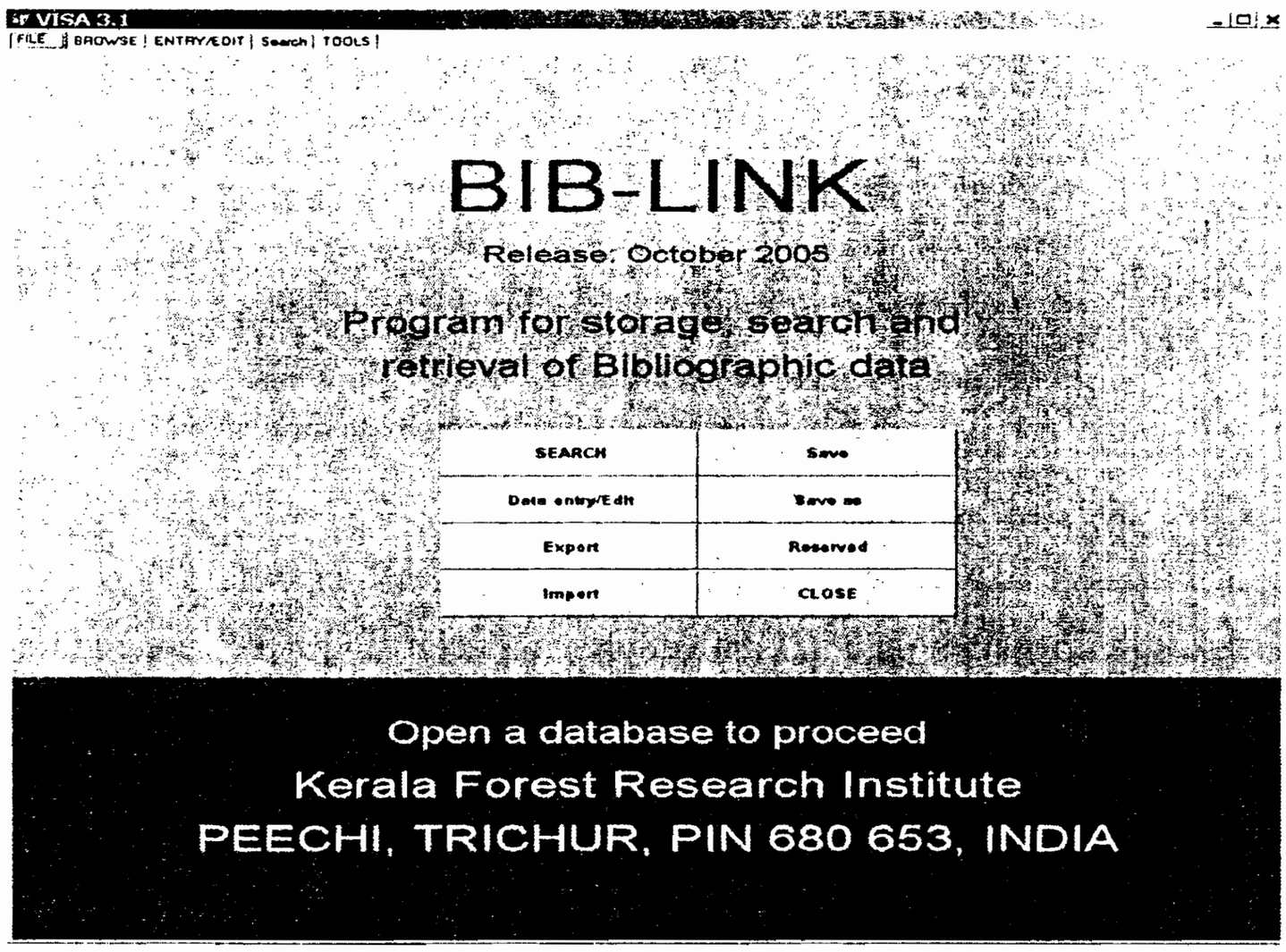

Fig. 3.1.1 Bibliography program. File open dialogue when 'Open' button is pressed. The tabbed note book containing edit and search pages become visible.

As in any typical MS Windows program, the main options are creating a new database, editing it and saving changes. Being a bibliographic database, there is provision for searching and exporting to other formats. The user interface is designed as a tabbed page. Separate pages are provided for data input/editing, searching and displaying results. It is also possible to edit records containing search results.

Working of the program

When the program is invoked, a title screen followed by the main menu would appear (Fig. 3.1.1). The first option is to create a new database. Next option is to open an existing database. In case of new database, the user has to supply a file name. After this, data can be entered. The data entry work sheet is shown in Fig. 3.1.2. The same work sheet is used for editing opened files. Desired record can be edited by changing the record number. There is provision for editing search results also. Data can be viewed in the browse window also. Eight records are shown at a time. PGUP and PGDN keys can be used for navigation (Fig. 3.1.3).

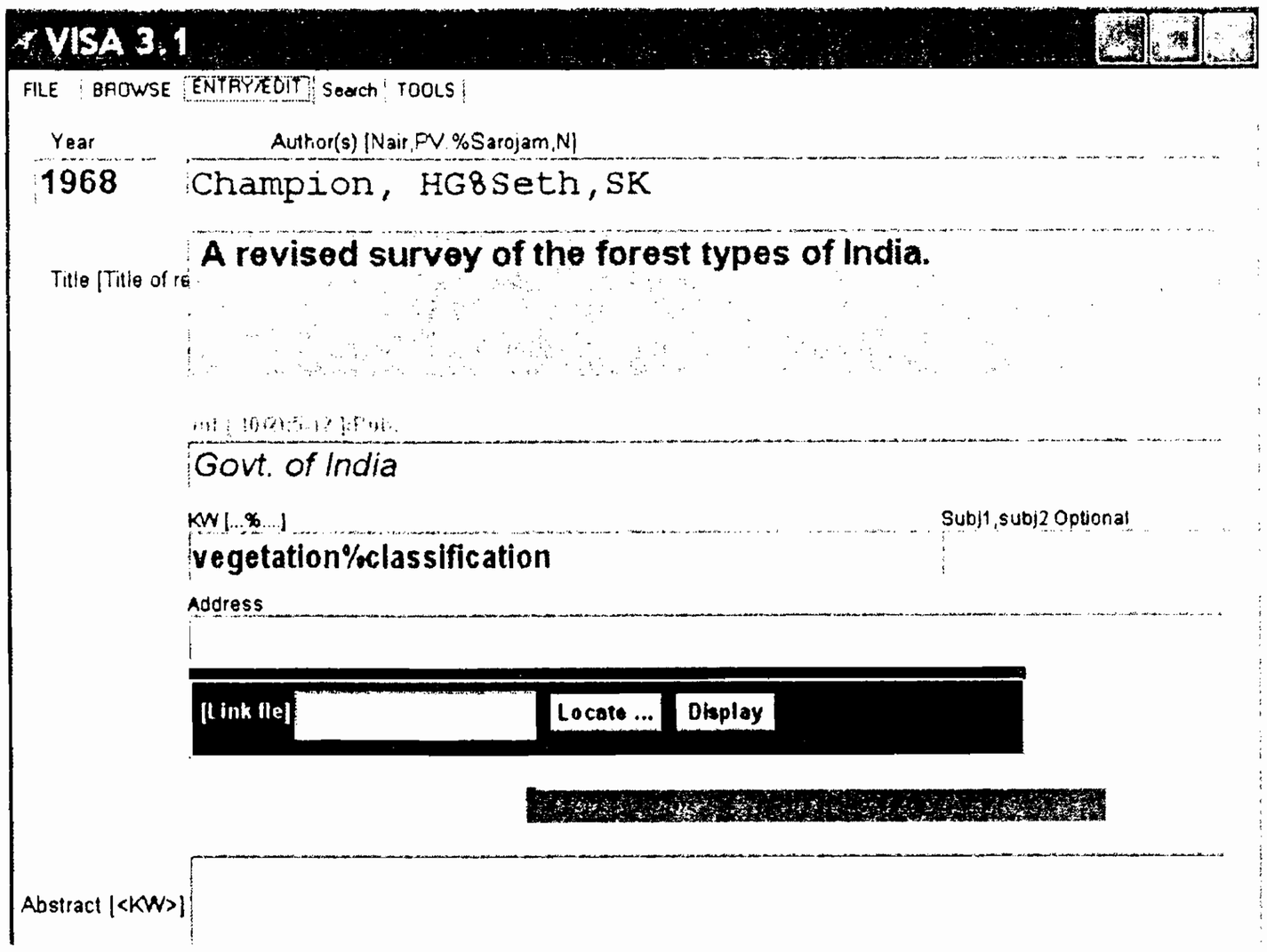

Fig. 3.1.2 Bibliography program. Edit work sheet. Each field can be entered or altered.

The program has very user-friendly search capabilities. Key words can be selected from the list. The database can be searched by author, subject or year. This package provides a combined index of authors, year and subjects. Names of broader topics and region are also included in the keywords so that search is possible by country or topic. The search results can be saved to a file. This file can be opened in any word processing program. The user can also enter key words in the search box directly. Key words can be truncated using the '*' symbol (Fig. 3.1.4). The user has to press the SEARCH button for searching the database. A maximum of three keywords can be specified at a time. Symbol + is used for combined search. The words can be truncated. Some valid search expressions are shown below. Instead of typing the entire keyword, search is possible by typing the first few letters of the search term like '**Bibli**' or '**Bibli***' for '**Bibliography**'. Searching by a term such as '**insect**'

would bring up a large number of references. One can narrow down the search by searching for '**pests**', '**borers**' or '**defoliators**'. One can obtain references on pests and their control by searching on '**pest+control**'. Both pest and pests would be automatically included. One can further refine the above search by '**pest + control + bio**' to get references on biological control of pests.

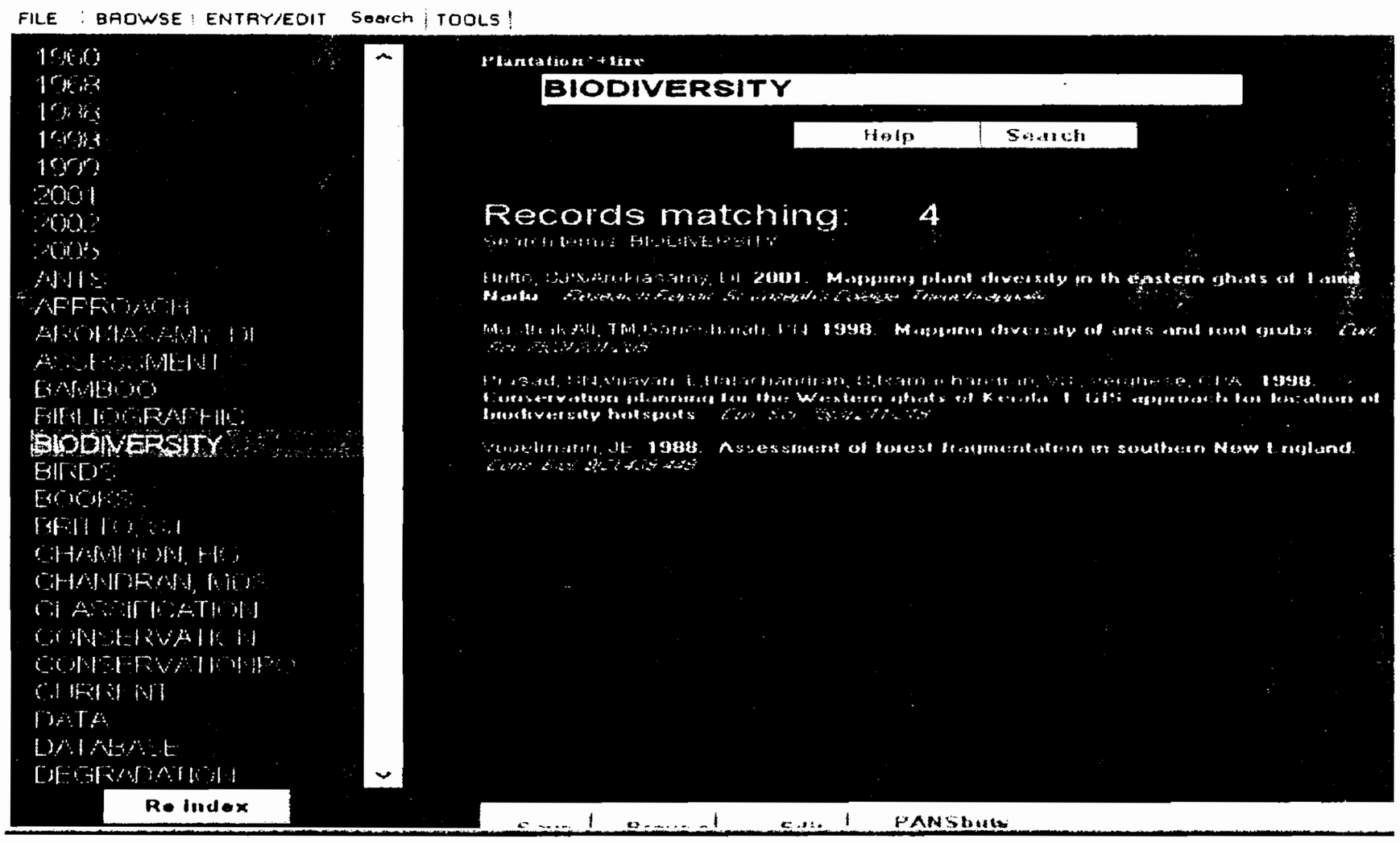

Fig. 3.1.3 Bibliography program. The search screen. Key word for search can be selected from the list. Matching records will be displayed.

While searching for '**Plantations**', the keyword can be truncated to '**plantation***' in which case both '**plantation**' and '**plantations**' would be covered. The * symbol is optional and is always added at the end of each search term. The term '**planta**' would also produce the same result, but '**Plant**' alone would cover '**plants**', '**planting**', etc. in addition. This would generate a large number of references. To narrow down the search further, additional keyword such as '**diseases**' or '**fungal**' can be added.

Plantation + disease
Plantation + disease + fungal

In the same way, '**seed + germi**' would show references containing articles about seed germination. '**Wood + treatment**' would show references in these topics. Search can also be made by author and year fields. A search of '**Nair + 1990**' would show references containing these two parameters. This expression can be further modified as '**Nair + defoli* + 199***'. This expression would bring up articles on defoliation by author Nair in the 1990s. A search by '**Africa + plantation**' would bring up references indicated by the keywords. The search results can be saved

in Rich Text Format (.RTF) which can be opened in most word processors such as MS Word and printed if necessary.

The program was further enhanced to link full text and graphic images. This would enable users to store the documents along with the database, Even items such as scanned visiting cards, news paper cuttings, reprints, research reports, etc could be stored and retrieved using the program. A routine was written for preparing hierarchical index in hard copy from the index of the database. Key words from each record would be permuted and the permuted list sorted alphabetically. Repeating key words are removed and record numbers added.

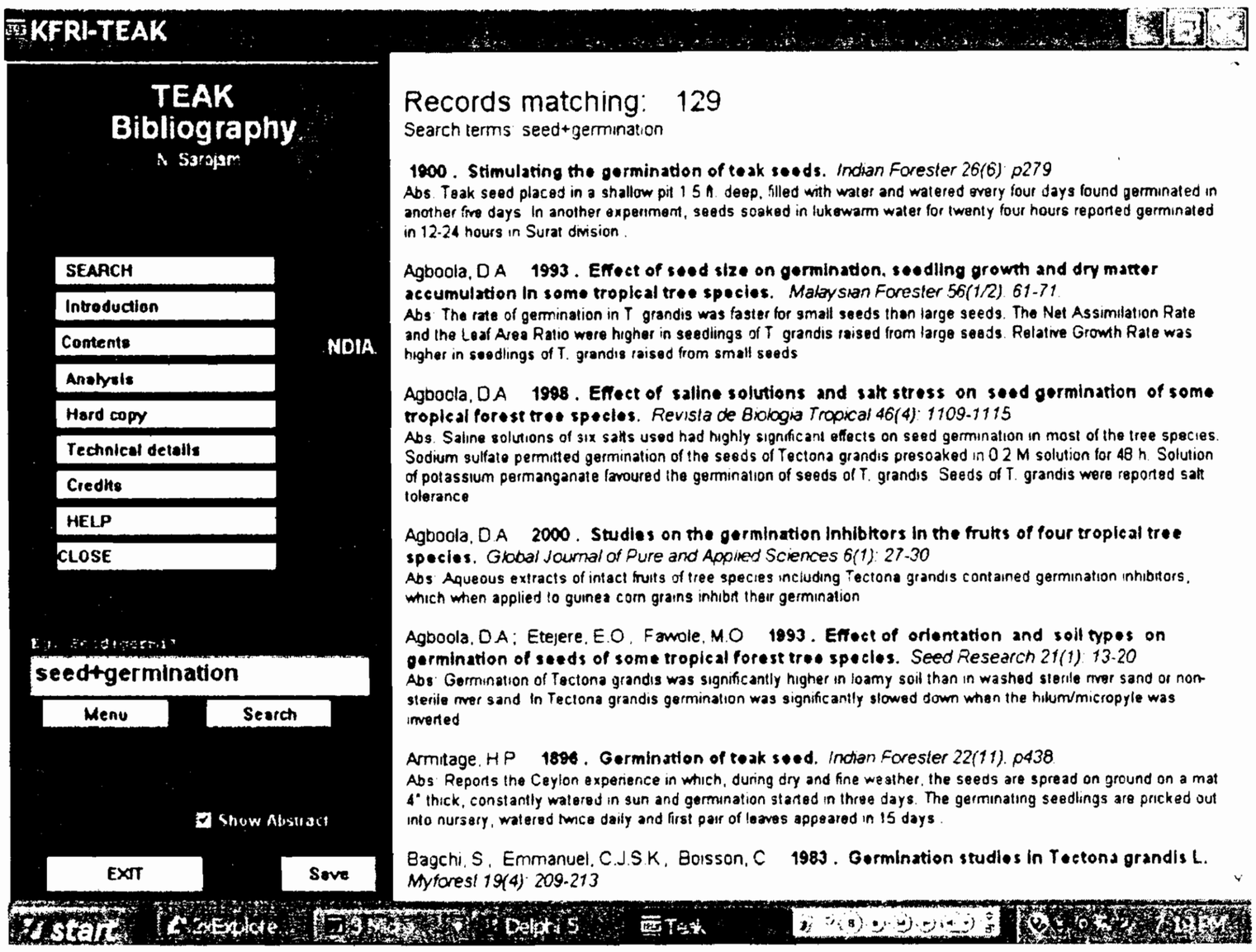

Fig. 3.1.4 Screen shot from a full fledged implementation of Bibliography on teak. The tabbed note book has been replaced with a windowed screen.

As a stand alone program

The bibliographic software can be used for making collection of bibliographic references. This SW alone is sufficient to bring out a collection of research reports or reprints in a self contained CDROM /hardcopy document. Search engine and formatted outputs are built in. Facility for storing full text make it suitable for storing visiting cards, news paper cuttings, photos, audio/video files, etc. Sample screen shot from a stand alone 'Teak Bibliography' prepared in this way is shown in Fig. 3.1.5.

3.2. Taxonamic identification

As mentioned in the methods section, a program was written for identification of taxa (TIP). The program's hardware requirements are minimal. Any PC running Windows 98 or higher should be able to run this program. The user interface is simple and intuitive. To make best use of the program, the user should be familiar with the taxa under consideration. In spite of the visual and descriptive aids provided, a person totally unfamiliar with the taxa under consideration may not be able to use this program efficiently. Presence / absence data of all characters for all members of the taxa are stored in a data file and compared with user input.

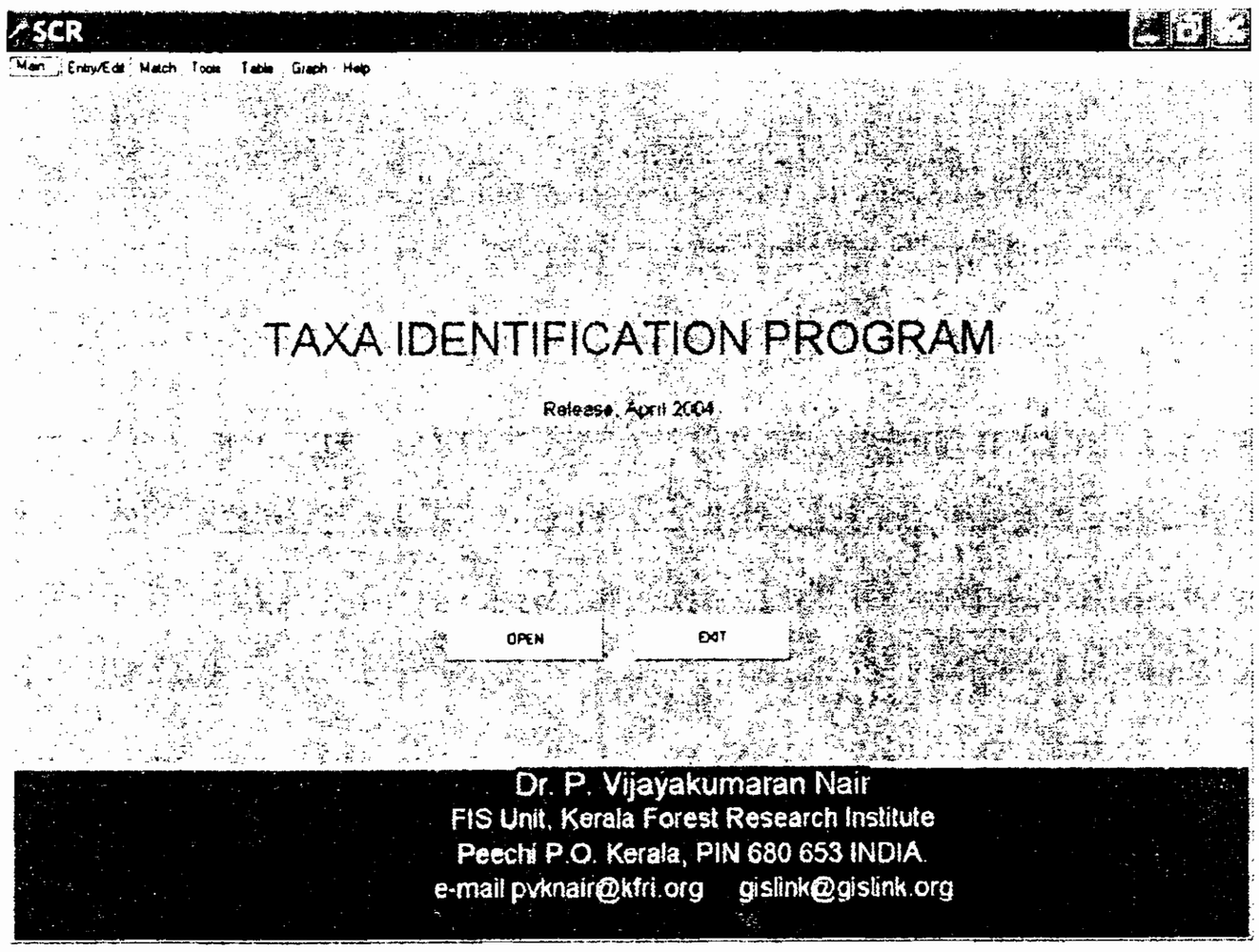

Fig. 3.2.1 Title screen of the Taxa Identification Program. The title is included in the first page of the tabbed note book.

The package was compiled as an installable software. The user has to run the program **`Setup.exe`** and the program will guide the user step by step on further action. The user may accept default directory names or specify his choice. The title screen will be displayed first (Fig. 3.2.1). The user interface is arranged as a tabbed notebook. There are pages for data preparation, identification and tabular output. Buttons are provided for getting additional explanation on each character, selecting character and matching taxa. The identification page consists of windows for selecting character, showing illustrations, explanations and taxa list (Fig. 3.2.2). When the user clicks on character name (not ticking the box), the character gets highlighted, illustration and explanation on that character will appear on the explanation box.

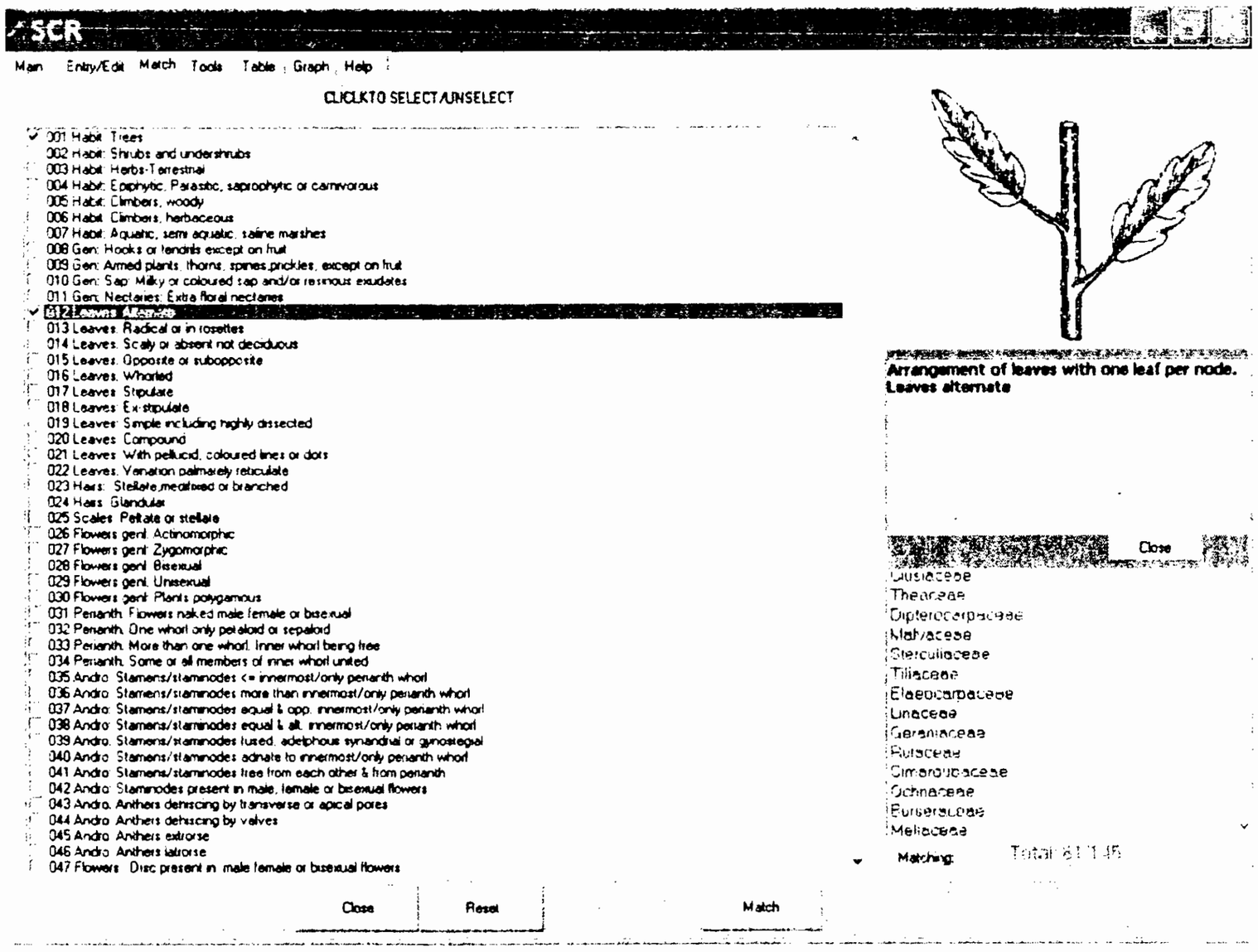

Fig. 3.2.2 The identification page. There are windows for displaying character, illustrations, explanation and list of taxa.

Further action depends upon user input on the screen. Clicking on character (not ticking the box), the character gets highlighted; illustration and explanation on that character will appear on the explanation box. The user has to tick the characters he is certain. After this, the match button can be pressed. The yellow window at right show the taxa matching the set of characters selected. The blue window above this shows the weight of remaining characters. A value of 1 indicates that by selecting this character, conclusive identification will be achieved. A value of 2 indicates that two possibilities will result. This may enable to choose required character and reach identification faster.

The program is designed as a general purpose one. One can use it for different taxa such as genera, family, etc by changing the data files. Sample implementations are shown below.

Case 1: Punched card key for identification of plant families

Data from the punched card key of Saldhana and Rao (1975) was implemented as a first case. This was a suitable system because it works on a character matrix for all characters in plant families. The punched cards themselves did not become very popular because (1) Difficulty in the alignment of the cards – a slight fault in the alignment of even a single card would hamper the identification process. (2). Difficulty in eliminating unlikely families - since in a single family, a wide range of characters is found, elimination process would be sometimes difficult. The advantages of computer

implementation over the use of individual cards could be listed out as follows: 1) The data present in the cards could be converted into a tabulated form, from which the data could be retrieved at any time very easily. 2) In the computer, the changes could be incorporated very easily wherever required unlike that of punched cards where changing each character requires either punching new slots or blocking already punched slots. 3) The computer could make the user understand the technical terms with the help of illustrations and additional explanations. 4) Selection of characters would be easier since all the characters were visible at a time on the screen and there was no need of ransacking the cards. 5) At each step of the selection of the characters, the computer could exhibit the names of the families, which possess that particular combination of characters. 6) The computer could give suggestions on the character to be selected further in order to attain the elimination of maximum number of families.

Computer implementation brought out additional information. Few of the families when examined were found to contain characters, which were wrongly placed in them and in few other cases some important characters were found to be missing. So the modification was done in such a way that the necessary characters were added and the improper ones were deleted.

Case 2: Synoptic key for Microbes

The synoptic key employed for identification of microbes is another suitable candidate for implementation of this type of identification mechanism. Here also a matrix of all the characters are available. In addition to these two cases, implementation from a dichotomous key was also attempted, but with out additional data many of the constraints of the dichotomous key would continue.

Technical aspects

This technical information is intended for programmers attempting similar tasks and key developers only. The program consists of one file, **`Tip.exe`**. It can be copied to a folder. Data files can be in a subfolder. Two Asci files **`Matrix.tax`** and **`charlist.tax`** contain the data. Text files and picture files are also kept in the subfolder. Data files are stored under a subfolder. Card.txt: Asci file that contains the family names and presence absence (0/1) of each of the 76 characters in each family (in the case of the punched card implementation). P001.jpg to P076.jpg: illustrations for the plant characters, stored in subdirectory Pathe\data, where path is the location from

where the program is invoked. Text files T001.txt to T076.txt provide additional description for each plant character.

Matching: The selected character is identified in the loop, its number extracted and used for finding the illustration and explanation file. When the match button is pressed, following actions take place. Initially all families are selected. After this, the program scans character by character. If the character is checked, the program scans all families in a loop and finds if that family has that character or not. If a family does not have that character, the family is dropped. The remaining families after running through all checked characters are the possible families. These names are displayed in the yellow box.

To get the next best character, the following mechanism is used.

Clicking on the MATCH buttons initiates a series of events.

1) First it does a scanning of all character in the matrix of char. The logic is quite simple, First all families are marked as '1' in a vector. Then each family is examined and if a character not present in the family is ticked, that family is marked as '0' in the vector of family list.

2) Displaying list of families having the selected character is now quite easy as the list is stored in the vector in the process described above.

3) Predictive weights of remaining characters. Because the computer operates very fast, operations not possible with index cards can be tried. Each of the character is taken and checked what happens if that character is ticked. This is done by calling the Match family function itself, which was described earlier. The function Count family returns the number if families matching. Since the predicted list is kept sorted on the number matching, most promising character (1) for narrowing down to a family would appear at the top. Further characters with diminishing potential appear subsequently.

The TOOLS buttons brings up a popup list of several types of lists that are of interest to users (Figs 3.2.3. and 3.2.4). GUI elements are positioned using values proportionate to screen size so that the boxes will appear the same irrespective of screen resolution.

Fig. 3.2.3 The TIP program has a provision for generating tables such as list of characters

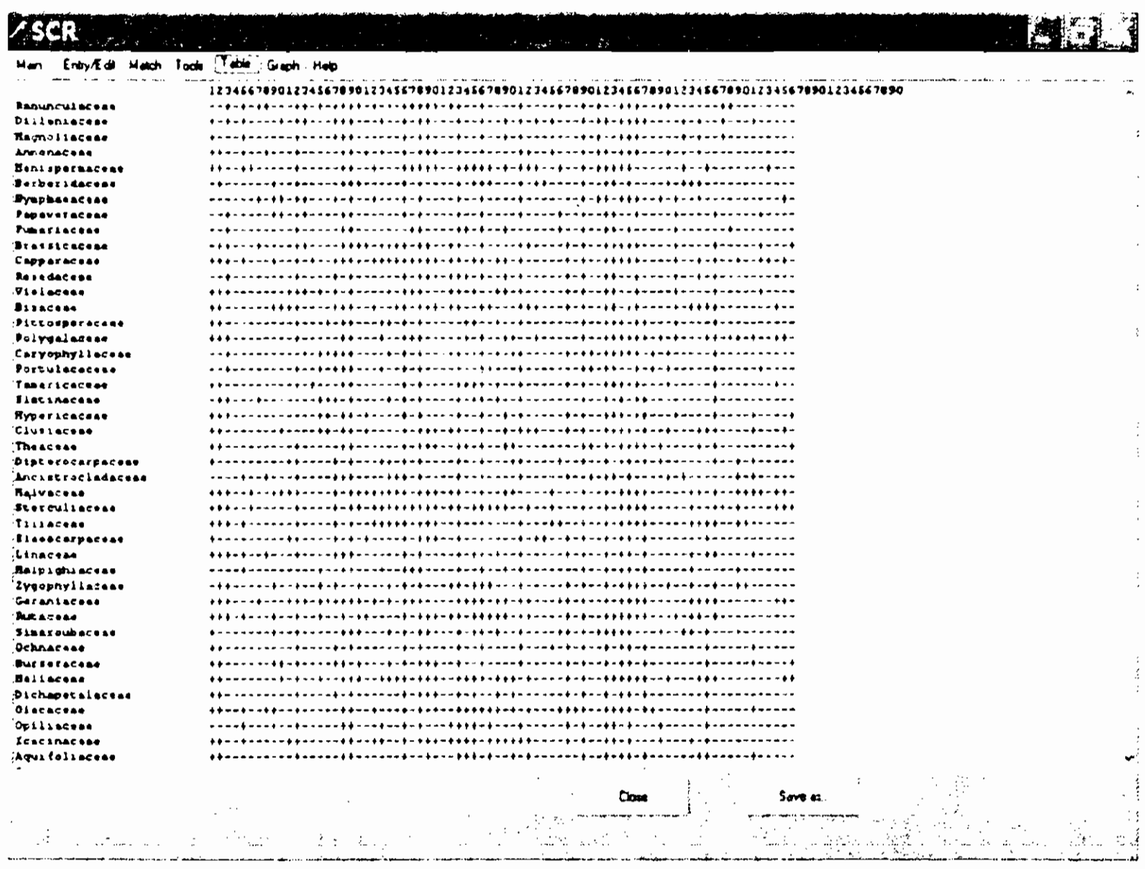

Fig. 3.2.4 The TIP program has a provision for generating occurrence table from character matrix.

3.3 Distribution of animals and plants

A program for displaying location of selected plant / animal on a map was developed as part of this project. The maps can be in WMF or ASCII format. Presence / absence is recorded using Survey of India map naming conventions.

3.3.1 Species information

Software is developed for storing detailed information on species. The software displays the index of the subject selected, and displays article stored in MSWord or PDF format. Information is compiled for common trees, birds, mammals and medicinal plants. Sample result is shown in Fig 3.1. A file containing name of species and name of its data file name is made first. The controlling program reads this file and displays the names as a selectable list. Details could be subdivided into groups such as plants, trees, birds, mammals, etc.

3.3.1 Mapping location

The background map is a feature encoded vector ASCII file. Distribution data is also supplied as an Ascii file with name and location details. The display box is designed in two sizes, low and high resolution which differ only by the size of the display box. The program displays the back ground map to fit the display window. The input file is read line by line and location marked by symbol.

The program has two parts. The first part plots data in a text file. The data has to contain location information in terms of x, y coordinates or topo sheet names. The second part permits displaying a background map in text format or the interchange format of Mapinfo. (There is no universally popular text file based interchange format for vector data yet). Two sets of data were plotted using this program. 1) Distribution of *Mikania* in the state. 2) Distribution of birds and plants from Western Ghats data base. When the program is run the user is given an opportunity to specify the background map and input file.

The program initially depended upon other utilities written by the author for preparation of map. This dependence was avoided by adopting a format such as the .MIF/MID files of Mapinfo. This way map preparation can be carried out in Mapinfo. Mapinfo exchange format is versatile, it is ASCII text, can contain points, poly lines, polygons and text. Interchange formats of

other GIS programs can handle only one type of object at any given time. The program was not developed into a stand alone package as there was no immediate demand for the same. Wide spread availability of GIS programs also obviate the need for such a package.

3.4. Graphical Representation of Biodiversity

A software is written for computation of various diversity indices and computation of biodiversity parameters such as Importance Value Index (IVI) from enumeration data. The software examines richness of higher taxa in addition and produces graphic and tabular results. This program was initially written to read name and girth data of trees and to do various computations on IVI. Data is supplied as a text file in a certain format. The data has to be arranged into sub locations, vegetation, species and sub plots. The program has data entry forms which facilitate entering data in this format. The program was made into two parts to process plot data and summarized data separately. Plot data consist of species name and girth recorded plot and location wise. The program allows data to be divided into regions, vegetation types and plots. In the case of a study on sholas, sholas in five different parts of the sate were treated as regions. In a study at Periyar, the five constituent ranges were considered as regions and deciduous, evergreen and shola as vegetation types. In a riverine study, low, middle and high elevations were treated as regions and proximity to river class treated as vegetation type.

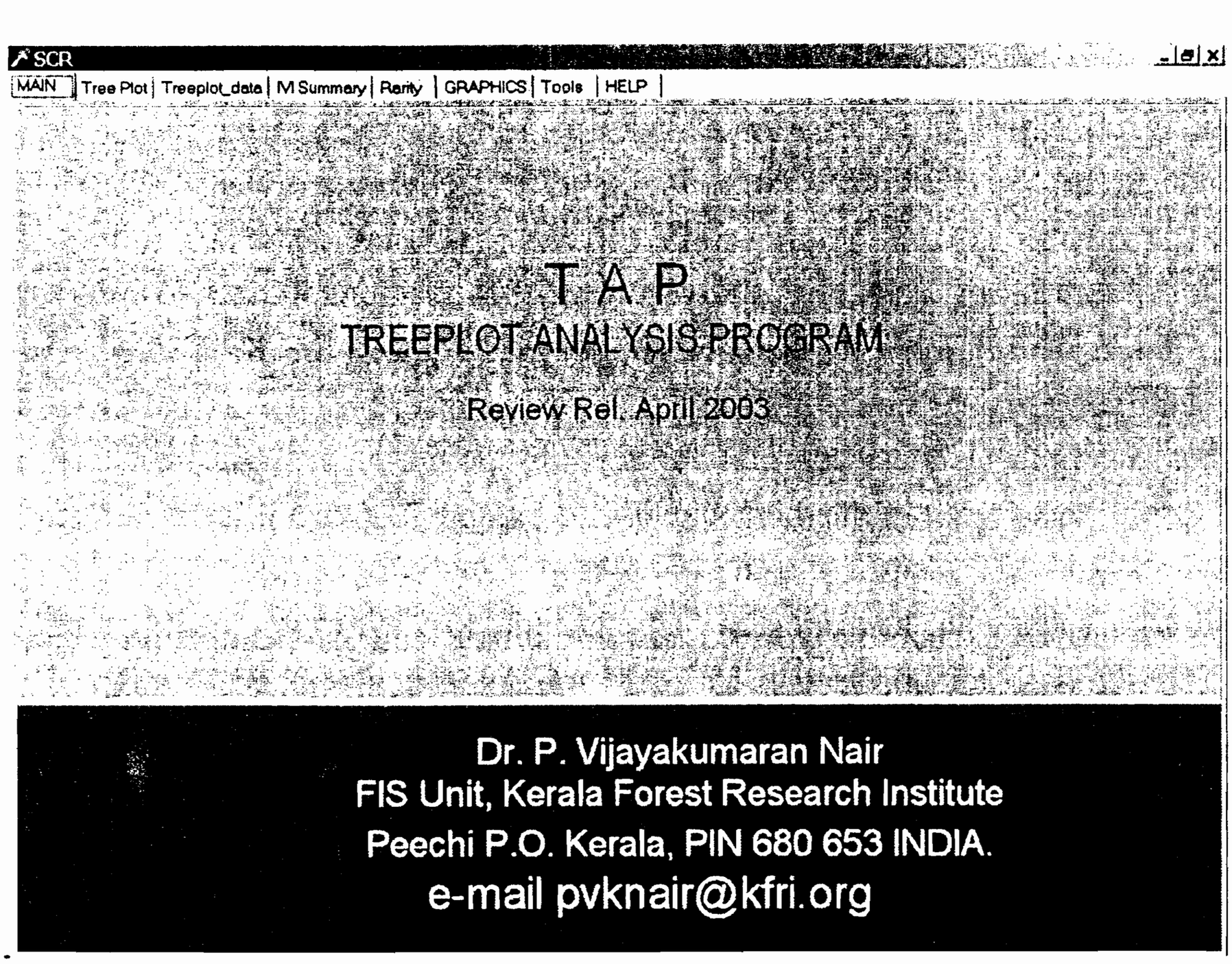

Fig. 3.4.1 Title screen of 'Treeplot' Analysis Program. The title is on the first page of the tabbed book.

Part I Analysing enumeration data

The main GUI in this program too is a tabbed book. The first page is the title page which contains a button for quitting the program as well (Fig. 3.4.1). The first page of the book contains interface for opening file, choosing categories of data, the tabular display area and the graphical display area. Tabular display and graphical display will also be displayed in full size in separate pages with option for saving to file. Plot data can be analyzed in a variety of ways. The user can select the location, vegetation, species or plot. In each of these all data or selected set can be analyzed. This makes it possible to pool and analyze the data in a variety of ways. Modules necessary for computing all biodiversity parameters are present in these modules.

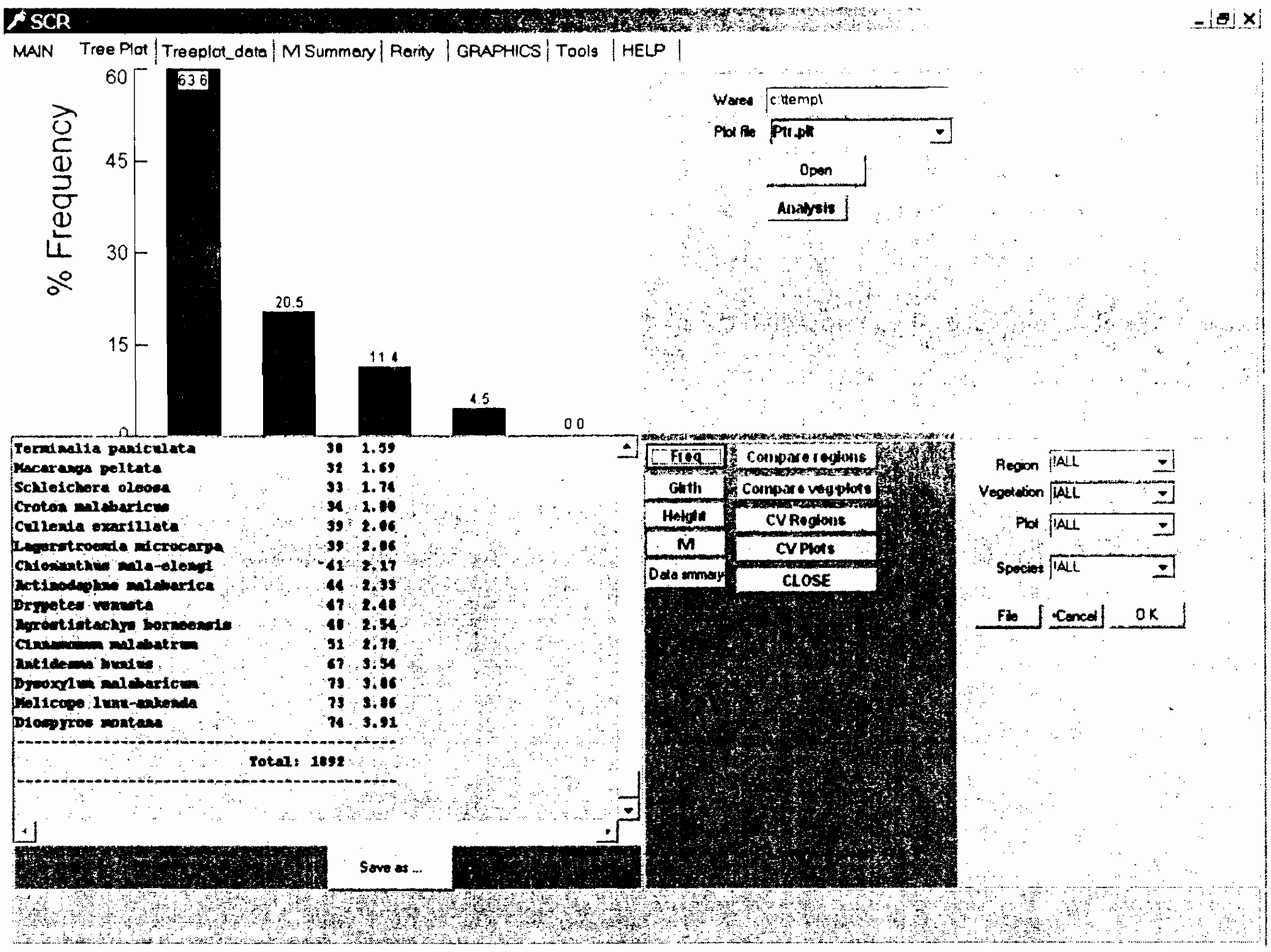

Fig. 3.4.2 The analysis page of the tabbed note book contains windows, buttons and options for the analysis.

Analysis can be carried out for entire data [ALL] or for selected region, vegetation, plot or species (Fig. 3.4.2). These offer a very wide variety of choice for summarising and understanding biodiversity parameters. The following types of analysis and diagrams can be generated.

Frequency: Frequency is computed species wise in tabular form. This, in combination with facility for choosing region, vegetation type and plots become a useful tool for initial summarising of data.

Girth: Girth of trees can be plotted on screen. The program carries out the scaling automatically and produces bar graph with appropriate legend.

IVI: The program produces IVI tables at the click of a button. The program computes necessary parameters such as relative density, girth and frequency for raw data. This also in combination with facility for choosing subcategories becomes a powerful tool. Outputs consist of a conventional IVI table and diagram which places the constituent species according to their girth and IVI values. Trees having similarity in values get grouped and appear as clusters (Fig. 3.4.4).

Compare regions: The program has option for generating a table comparing regions. This facilitates side by side view of parameters (Fig. 3.4.3).

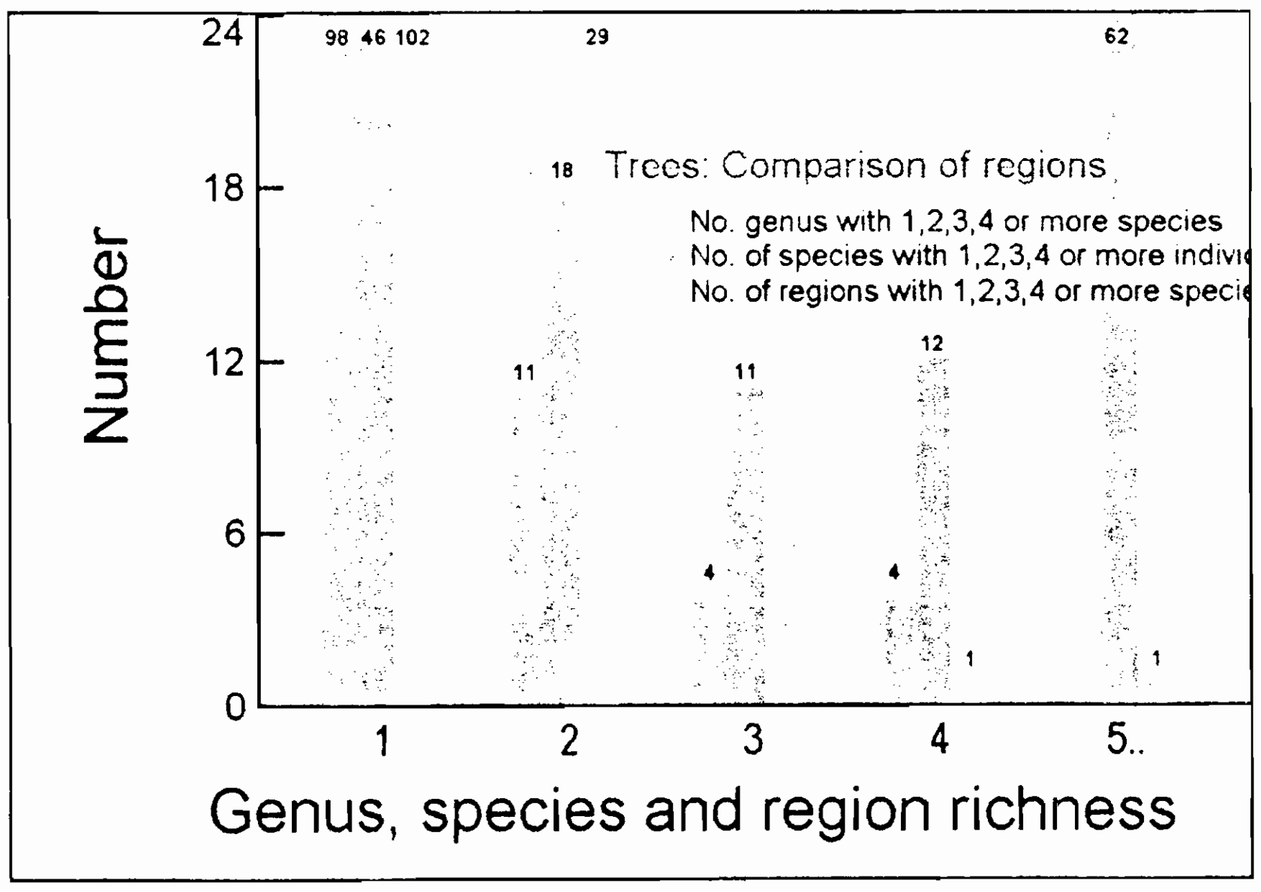

Fig. 3.4.3 Individual, species and genus richness of different regions.

Compare vegetation: The program has option for generating a table comparing vegetation types. This facilitates side by side view of parameters.

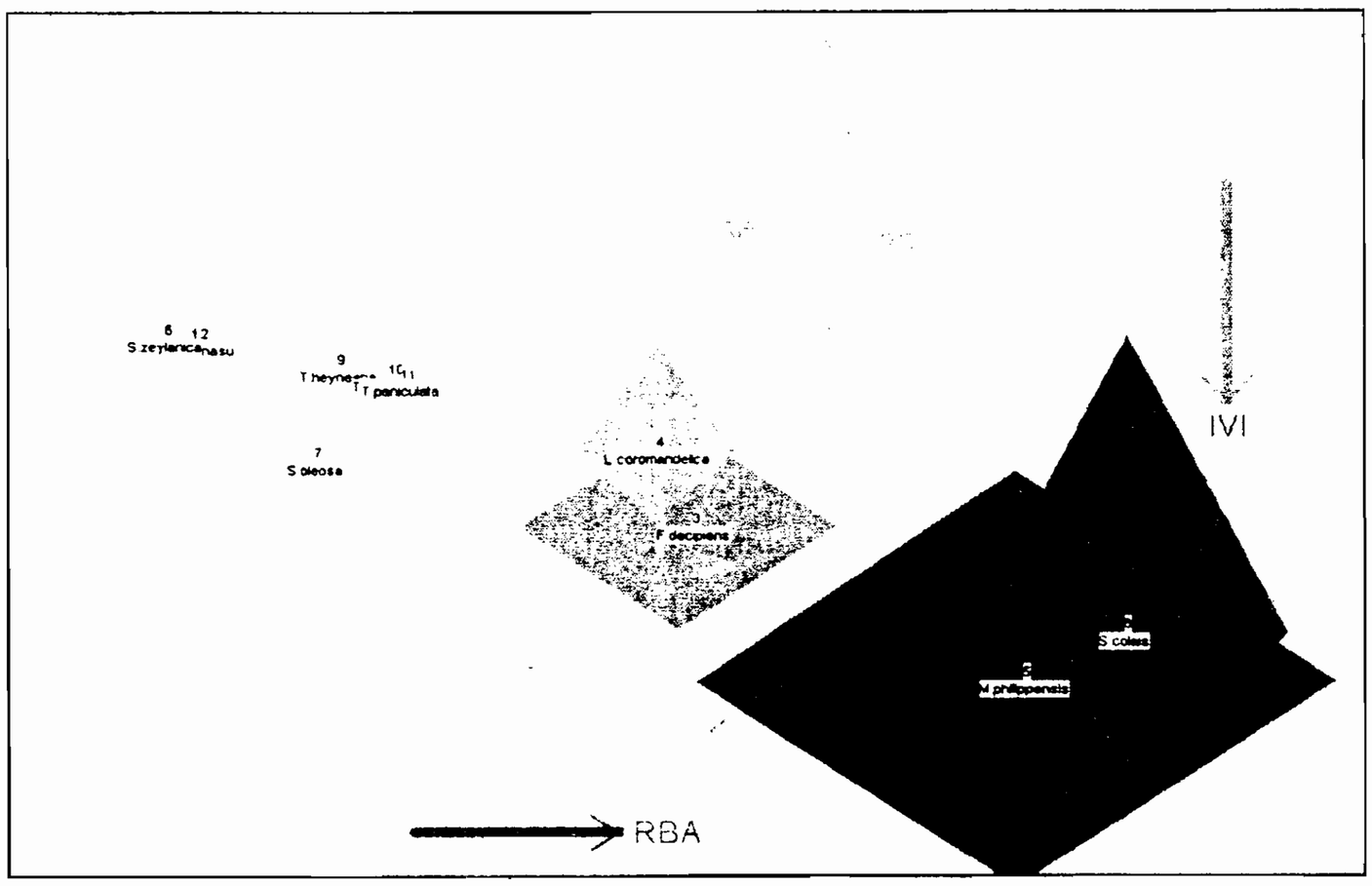

Fig. 3.4.4 Visualisation of AVI parameters.

Part II Plotting summary data

The program was further extended to plot as many as four parameters. Using this data available in many published sources can be plotted. The data can be supplied as a text file with name followed by values of parameters separated by commas. Options are provided for specifying headings and parameter names. The diagram can be saved in many formats.

3.5 Satellite Images

Satellite images are very useful for getting a summary status of land use in an area. After the launch of Indian Remote Sensing Satellites (IRS), the cost of satellite images have come down. But only a person well versed with remote sensing techniques can make use of it. National Remote Sensing Agency (NRSA) who supplies these images provides only a DOS based program for viewing individual bands of the image. Modern PCs offer tremendous amount of processing power. Earlier image processing was done by agencies such as NRSA or Regional Remote Sensing Service Centres (RRSSC). Even these agencies used to process images in small pieces because of the limitations of hardware and SW at that time. Entire images could now be processed at home by users. Commercial image interpretation software is complex and costly. These tools fill the gap to some extend.

The tools presented here fall into three categories. 1) Tools for packaging, display and extraction of images from IRS1B, C and D satellites (Lis II, III and IV). 2) Packaging public domain data belonging to Kerala region from Landsat and TM 3) Packaging public domain SRTM data along with freeware for processing.

1. Display and packaging of IRS Data.

A set of 54 images belonging to different regions of the Western Ghats from 1962 to 1965 was available in CCT format at CES, IISc as part of the Western Ghats Biodviersity Progam. This collection, valuable for biodiversity studies was fast becoming unusable as CCT readers were becoming obsolete and CD readers and writers were replacing them. Working tape drives were located and the data transferred to hard disk. The data was converted to CD ROM format and incorporated into a CD with a controlling program which enabled selection of images based on map and location query. The CD was compiled as a self contained CD ROM which when inserted in a PC would display a map of Western Ghats and allow user to select the image with a mouse or from a list of states and districts. The whole or selected parts of selected image in uncompressed and written to user specified location, band wise in easy to read format for further processing. Details are shown in Figures 3.5.1 to 3.5.4.

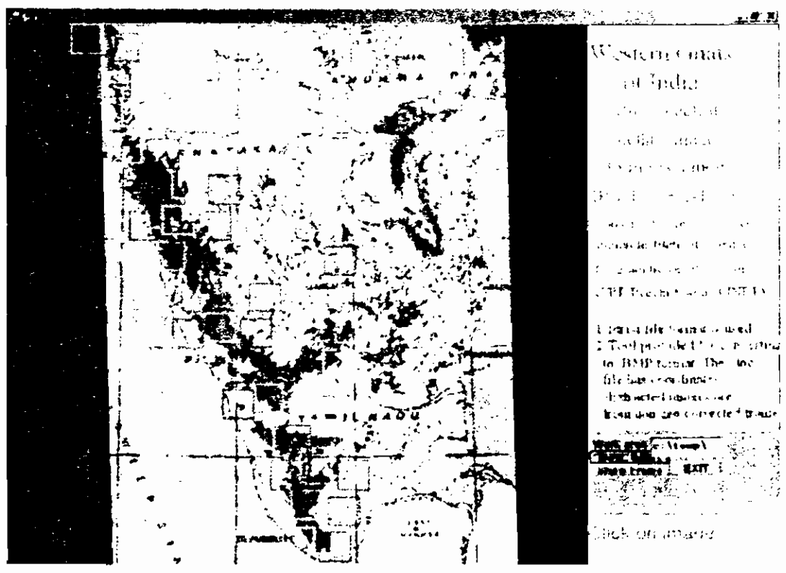

Fig. 3.5.1 Title screen of compilation of IRS 1B images of Western Ghats

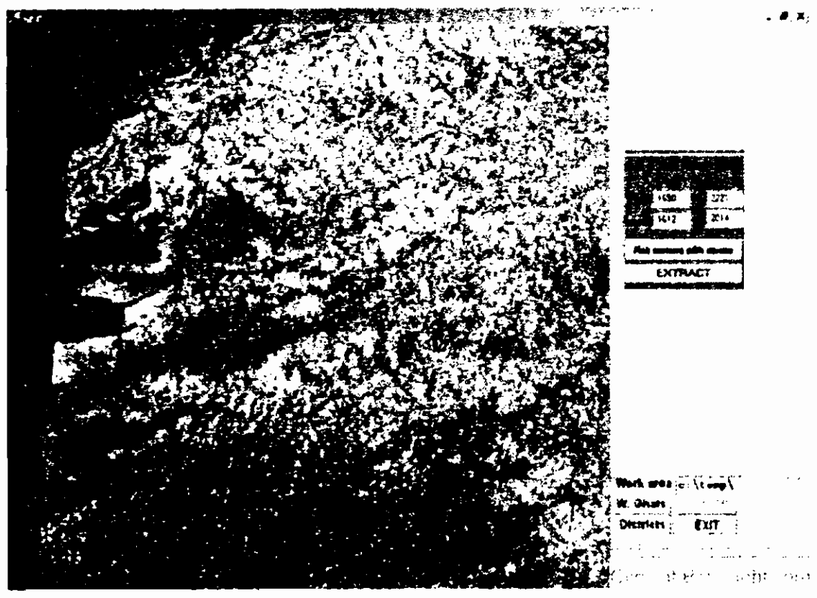

Fig. 3.5.2 Selected frame of IRS 1B is displayed on the screen.

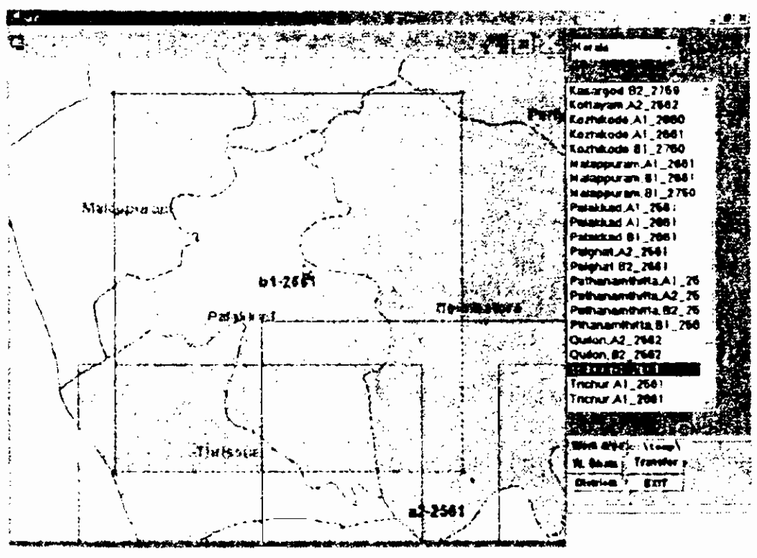

Fig. 3.5.3 The district selected is highlighted and the image coverage indicated.

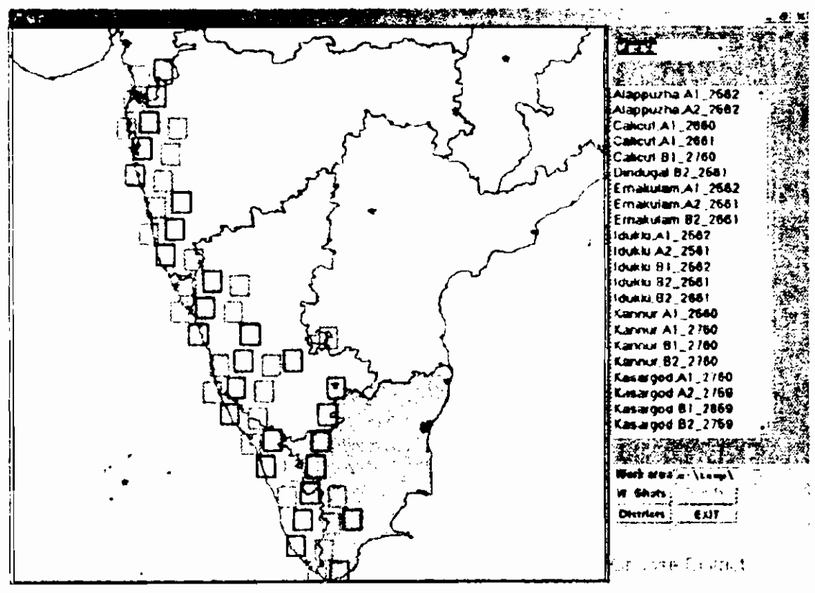

Fig. 3.5.4 Modified interface with facility for selecting image by state and district.

A similar compilation was done for IRS IC and D images of Kerala. Six images covering whole of Kerala could be placed in one CD ROM. Separate sets were compiled for periods such as 1997-98, 2000-01, 2004-05, etc. As in the case of IRS B images whole, or selected part of the image could be extracted interactively. Figures 3.5.5 and 3.5.6 show sample screen shots.

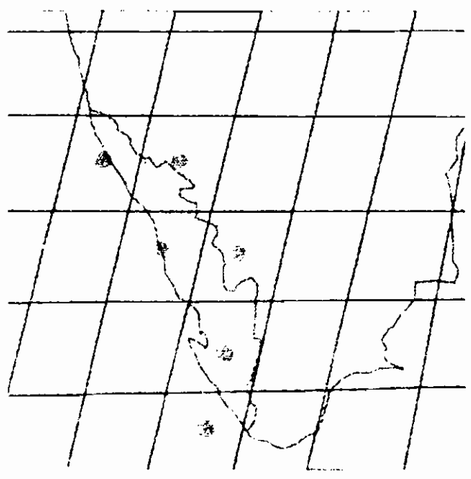

Fig. 3.5.5 IRS IC Frames available

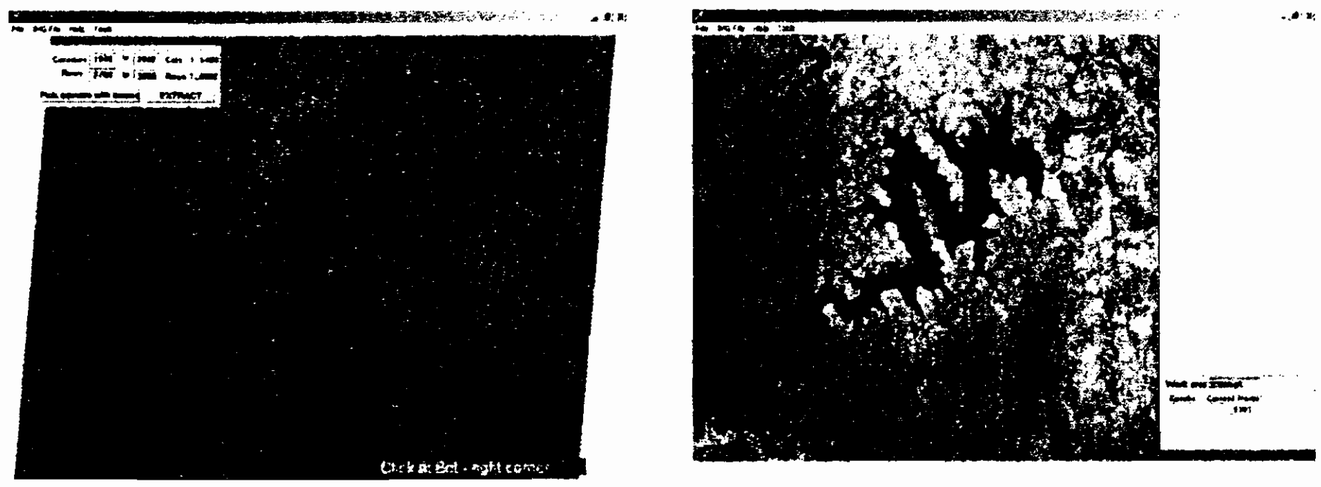

Fig. 3.5.6 Corresponding frame of IRS1C extracted.

2. Repackaging of NASA Data.

Compiling a set of satellite images of public domain NASA, MSS data for 1973,1990 and 2000 for Kerala region. The images are included in a CD the contents of which can be viewed through an htm linked file. The purpose is to make the data available to users in its original form. Therefore no viewing program is provided, even through screen shots of views are provided. A detailed mosaiced FCC for Kerala region is provided in bit map form. Figures 3.8.7 and 3.8.8.

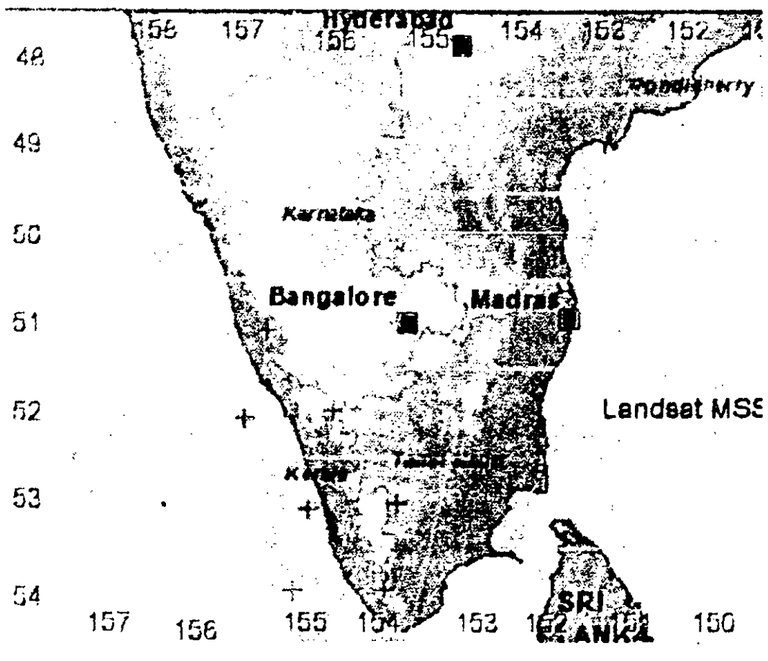

Fig. 3.5.7. Nasa Landsat frames of Kerala and adjoining areas

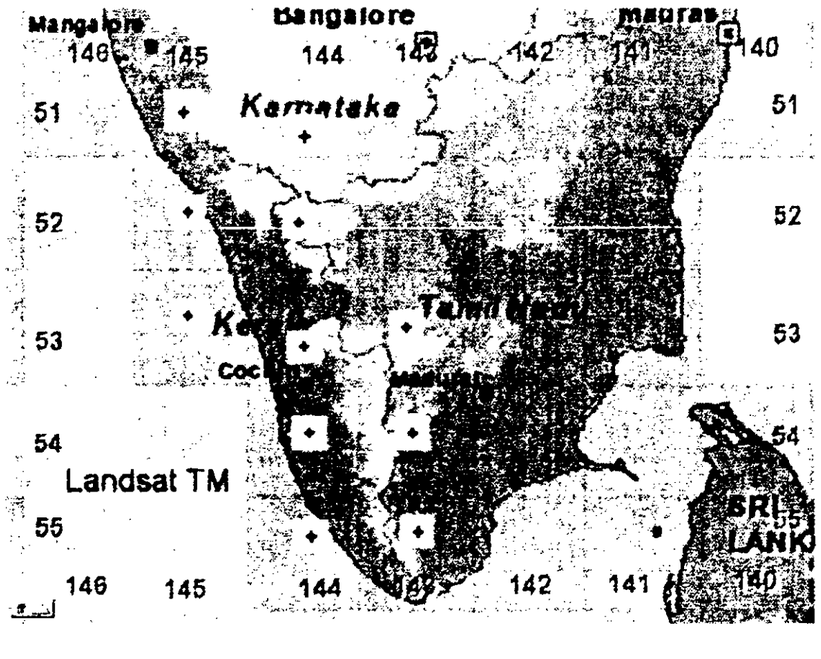

Fig. 3.5.8 Nasa TM frames of Kerala and adjoining areas

2. Repackaging of SRTM Data.

SRTM data belonging to different locations of Kerala is brought together in a CD. Details are given under section 3.7.

Technical details:

IRS images contain header files which provide band details and row/column size. From this images could be extracted by reading serially and writing to appropriate files. The corner coordinates are stored in the controlling program for generating images for selection. Correspondence is established between screen coordinates and images displayed on screen so that required regions could be extracted. Histogram equalization technique is used for generating rich RGB views. In the case of NASA data, browse images available, along with data is used for preparing material for interactive selection. Tools for subsetting DEM data was based on geographic coordinates.

Baseline maps of Periyar Tiger Reserve

The maps and documentation are prepared sectio
PDF format. The design was to create a title page in .HTM
maps could be viewed in Internet Explorer. Documents and
hierarchically in tabular form. Hyper links were made to
Screen shot of the title screen is shown in the discu
documents in PDF format had the advantage of portabili
kept in MS Word and Mapinfo format for the purpose of fu

Forest Atlas of Kerala

In the case of Forest Atlas also a similar method w
main HTML page had links to topics and maps grouped int
forest divisions. The opening screen is illustrated in the dis

Keeping the documents in .PDF format had the ad
portability. Original files are kept in MS Word and Mapinfo
purpose of future editing.

3.7 Terrain visualization

Shuttle Topo Radar Mission (SRTM) data for Indian region comes in two resolutions so far, 1000m and 90 m. The former cover whole of India and is suitable for general views of area. Tools are developed for format conversion and sub setting of regions. Ninety meter data come in degree tiles. Tools are developed for splitting and merging of these files. Sample operations are illustrated in Figures 3.7.1 and 3.7.2.

SRTM data downloaded can be displayed using freeware such as 3DEM. The author had digitized 1:1000,000 scale topo sheets to produce DEM of Kerala of about 500m resolution. Later NASA data for whole India at 1000m and 90 m data as degree tiles for the Kerala region became available and is repackaged as a CDROM. The CD cannot be made self contained as the freeware programs used for displaying the images do not have an external integration mechanism. The freeware programs are also contained in the CD, which the users may run. Overlay files for the degree tiles are provided. These cover entire extend of degree tile there by overcoming need for geo registration. A self contained CD is prepared for DEM data.

Terrain visualization is becoming a popular tool. Various GIS programs provide this facility to varying degrees. Some times freeware programs designed for a specific task is able to yield better results. In this part methods are described for preparing DEMs from contours, displaying them using freeware programs and about sources of high quality free data available for the task.

Three types of data can be used .

1) Point data in programs such as surfer. After preparing contour lines and then grid map.

2) From contour lines of topo sheets. Contour lines can be scanned and geo-referenced topo-sheets. A field in the table would contain the elevation values. GIS programs such as Mapinfo has modules for interpolating the contours and preparing three-d surface.

3) DEM data is available from sources such as SRTM. Data in 100m and 90 m resolution are available. These can be directly displayed using Raster GIS or terrain visualising programs.

As part of the exercise, routines were developed for cropping required regions from the DEM data.

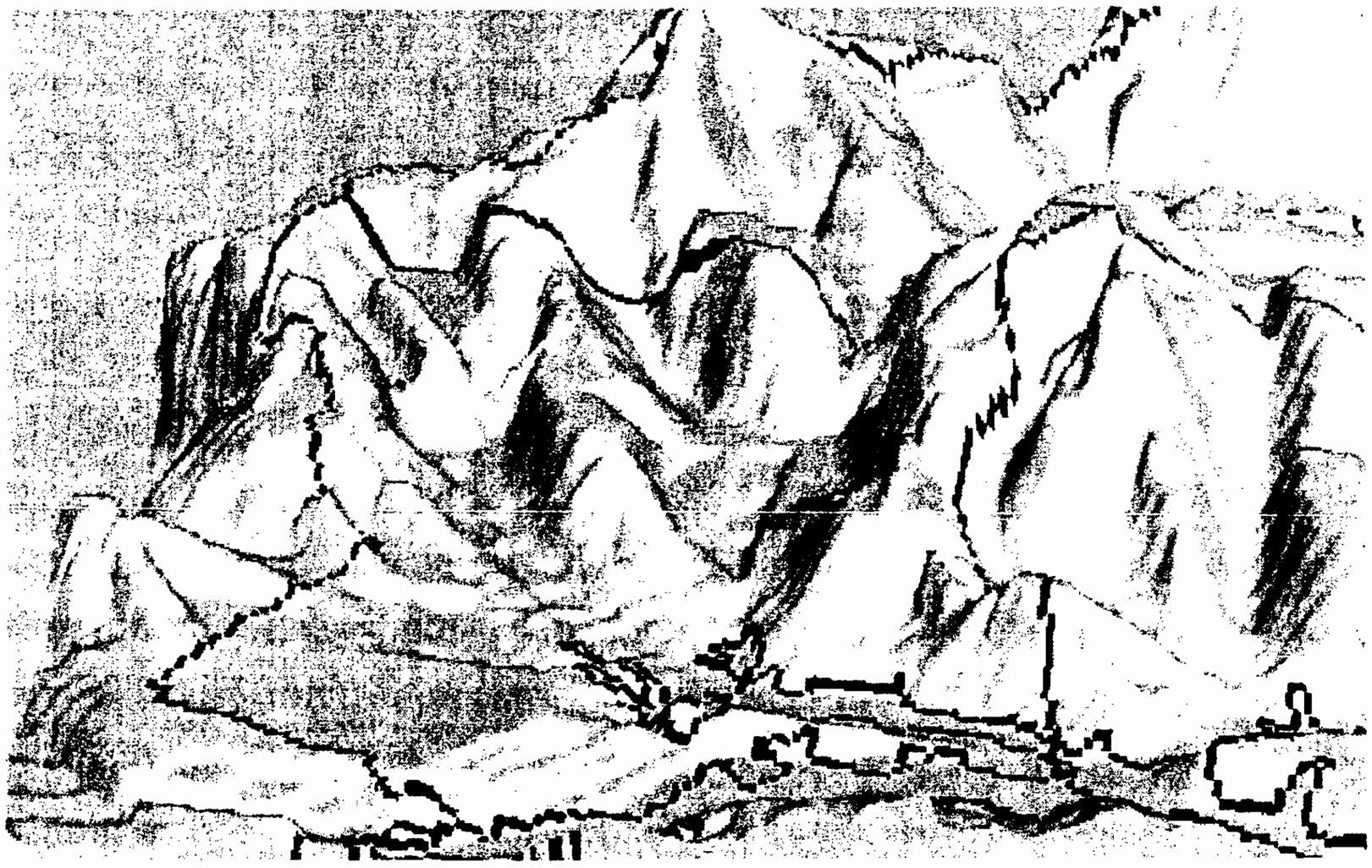

Fig.. 3.7.1 Three dimensional view generated from contours using free ware program

Fig. 3.7.2 Vegetation layer is overlaid on three dimensional image.

4. Discussion

Integration Mechanisms

Integration mechanisms employed in several information packages were analysed. The initial objective was, to provide information on bibliography, taxonomy, maps from a single source. This was in a way too ambitious. Even encyclopedias do not support indexed storage of all data, often maps are stored as pictures only. In the present scheme, the pattern of an encyclopedia was not suitable as the users were expected to add or modify the information as they use it. An attempt was made to implement a prototype which could search several types of information but the idea was given up in favor of listing the databases and the user searching inside each data base. Implementation details of the two schemes are presented.

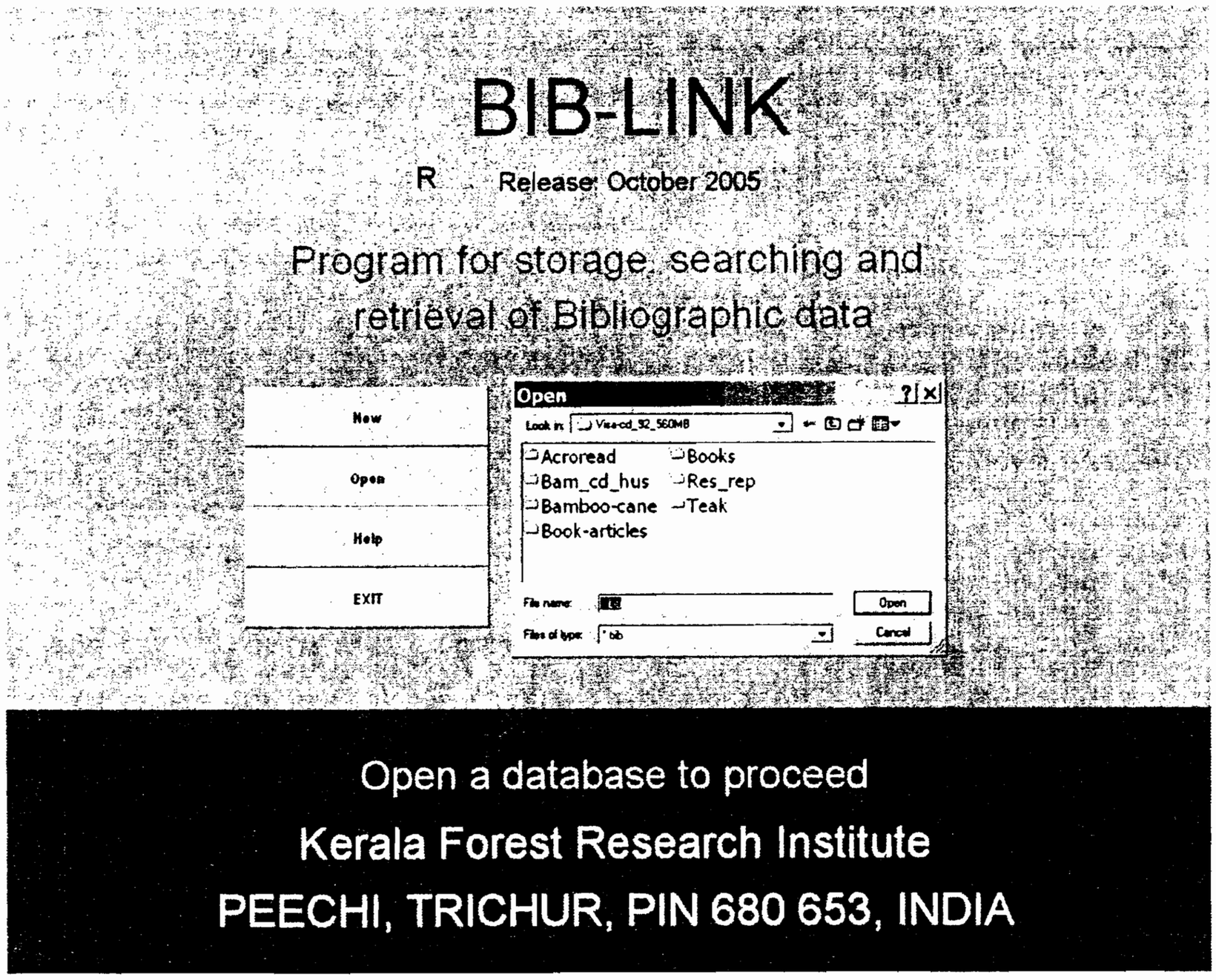

Fig. 4.1 Opening screen of a compilation of different databases.

Search over Multiple data bases

In the standard use, index file contains only the key word and record number. In order to operate on multiple data bases, indication of database was also included in the index. The first 20 letters were used for the key word, next three letters for the database code and the next nine letters for record number. In this way a prototype with smaller data sets could be built. Display mechanism

had to be separately devised depending upon the database. But as the data base size increased, unexpected problems arose. One of difficulties was the limit of items that could be displayed at a time in a graphical user element called list box. Most software design tools limit it to 32,000. Packages such as encyclopedias overcome this limit by maintaining separate lists of multimedia, location, etc. The index can also be broken as per the first letter of the word. Even though it was possible to go ahead with the multiple database design, during trials it was found that users prefer to look under separate categories such as bibliography, taxonomy, location, details, etc. There fore further developments were carried out on this line and several CDs could be compiled using individual database mechanism. Sample screen shots for the package is shown in Figs 4.1 to 4.4

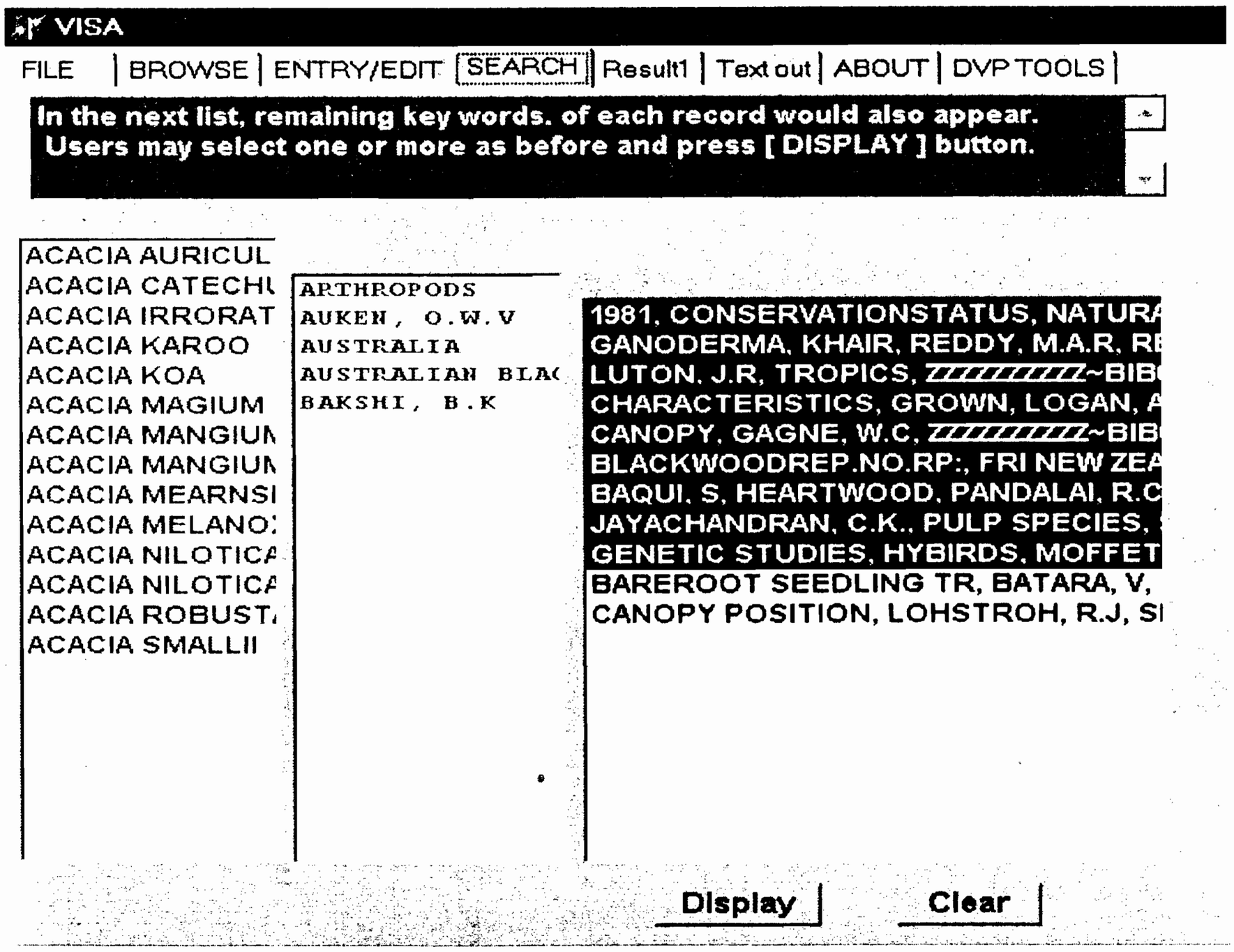

Fig. 4.2. Screen shot of the search procedure employed

Search in individual databases

An information package can be designed on single topic or multiple related topics. Packages on Bamboo or Teak contain single data base. Bibliography on reprints, books, research topics form a set of related data base brought out in single CD. Slightly different approaches are used in these two cases. A controlling program developed under MS Windows GUI form the first component. In the case of single data base, the screen can be divided into areas

for displaying main menu, search window, search results, documentation, etc. Sample case studies are given later. In the case of CD containing multiple data bases, the main menu can be selection of database and then search and display of results carried out on selected database.

It is possible to produce automatic links and contents for documents arranged into hierarchical folders. The fllwing code can be used for the purpose. It traversed the directory tree in a recursive manner, preparing index files for each.

```
Procedure Folder_to_htm(path)    //Recursion capable procedure
  Write the header of path.htm
  Write the table heading for a two column table
  Scan the conttents of the folder sequentially
  begin
    If the item is a file
        Prepare a row of htm table using the file name.
        //file name is used in the link and name fields.
        //User may change the second manually
    If the item is a folder
        Prepare a row of htm table using the folder name.
        //Name is folder name, link is \foldername\foldrname.htm
  end
  Write line to close htm table
  Write the tail of path.htm and close it.
End of Procedure
```

Integrated Packages

These control mechanism were effectively used in preparing a self contained CD ROM for base maps of Periyar Tiger Reserve. Index.htm was used as the starting point which runs from the auto run option in the CD ROM. The contents are displayed in the form of a table. Links are provided to locations in the table. Actual documents are in MS Word, Adobe Acrobat or MS Power point presentation. The browser program will do the work of opening appropriate packages.

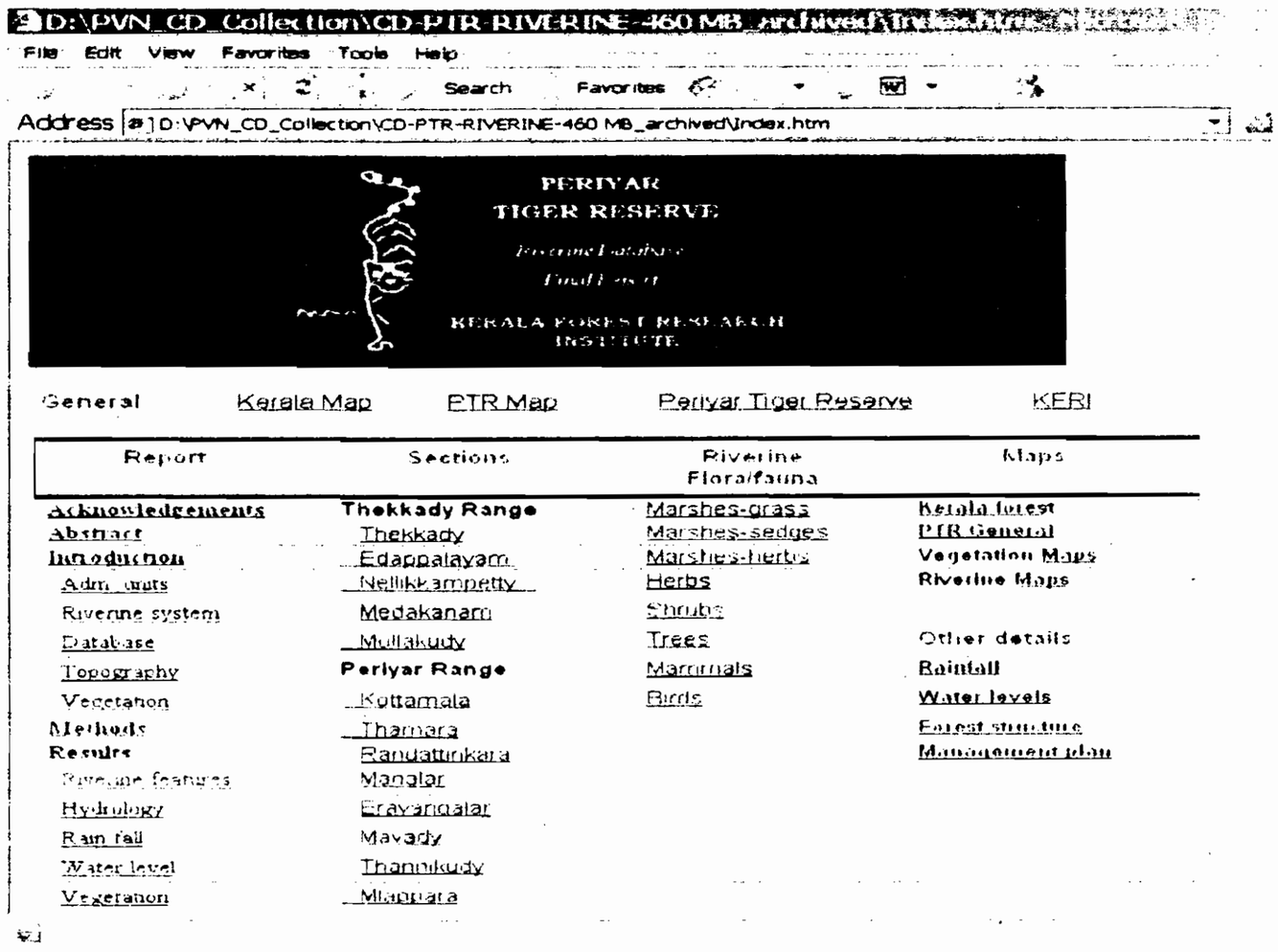

Fig. 4.3 Title screen of map compilation of Periyar Tiger Reserve

More or less same mechanism was used in preparation of a CD ROM of Forest Atlas of Kerala.

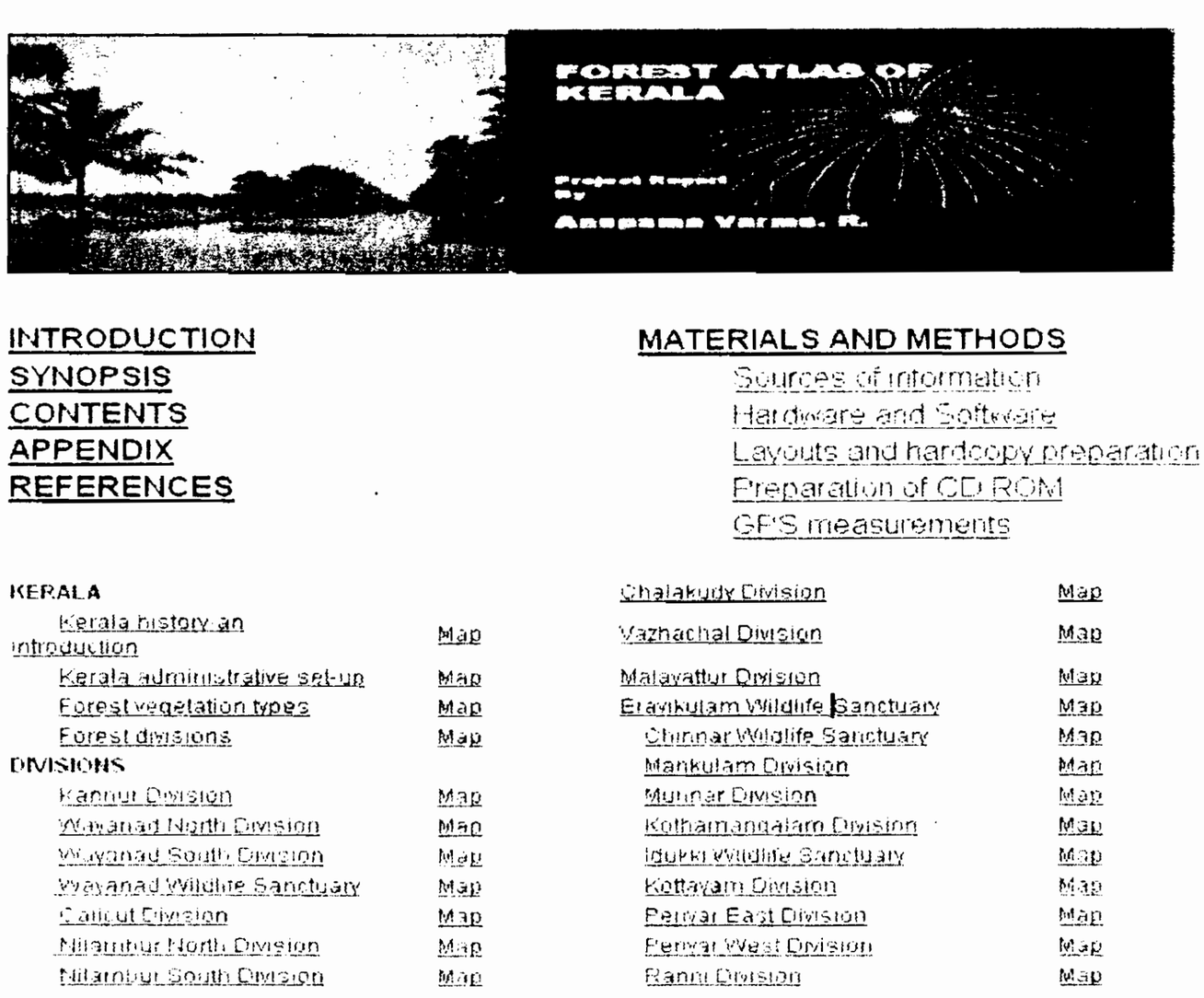

Fig. 4.4 Title screen of map compilation of forests of Kerala

A controlling program becomes necessary for a combination of data bases. In the case of an information system for Periyar Tiger Reserve, buttons were provided for selecting subjects, clicking of which will load a list of topics. Which also contain the name of the associated document. The program uses a feature available in MS Windows programming, to open the package associated with a particular file.

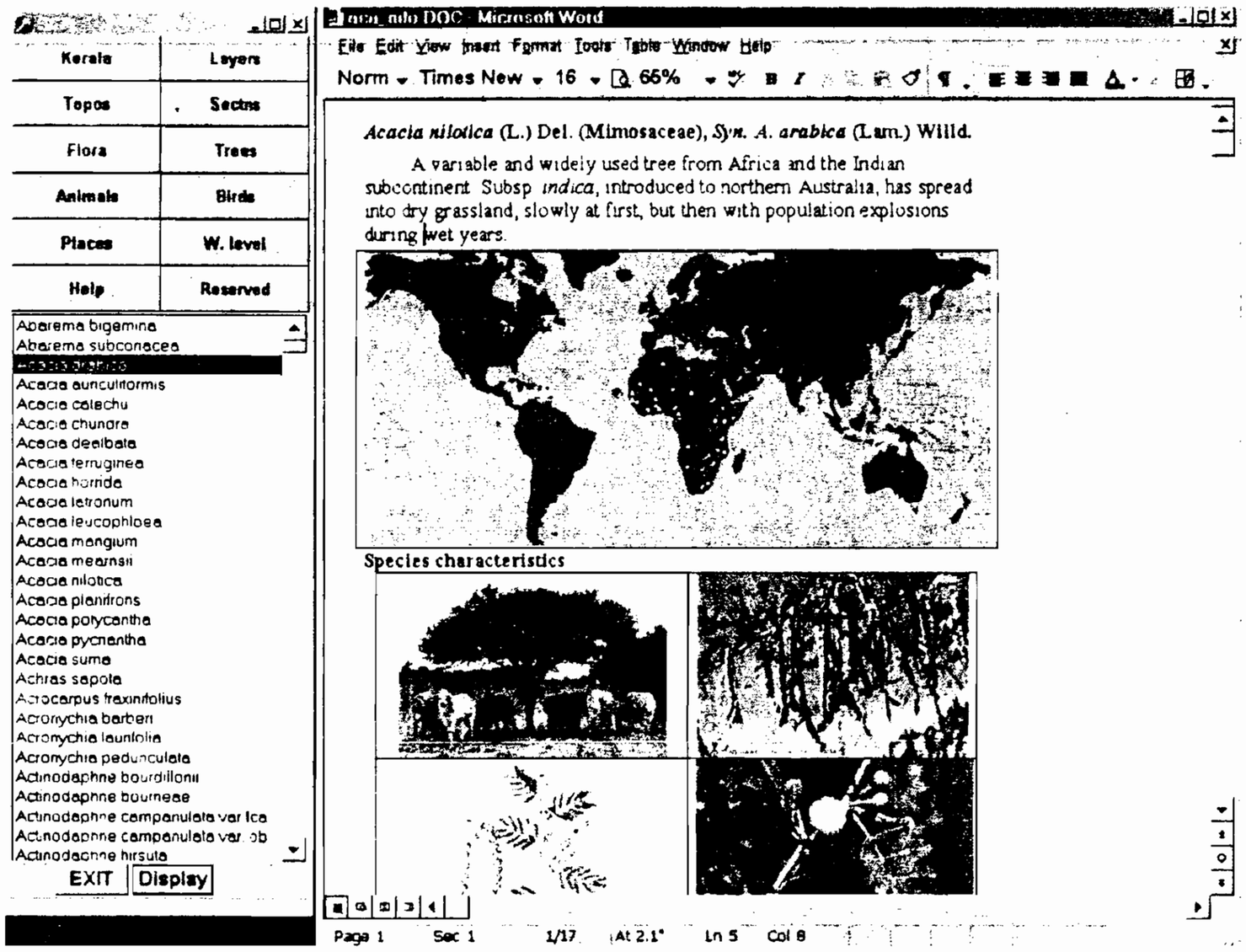

Fig. 3.8.3 Screen shot from a compilation of multiple data bases.

Several programs were developed as part of this project. It could be interesting to examine the utility of these programs and whether any of available programs could be used in their place. Some times it so happens that at the time of writing there was no alternative, but in course of time better programs or data became available.

In the case of Bibliography program, the currently available packages available are meant for library management and only a professional librarian can use them. Packages such as Reference Manager have not become popular due to some reason or other. With the availability of abstracting journals and Current Contents in electronic form the scenario may change. The current program has self contained search engine and very few commercial programs provide this facility. More over the facility for storing articles directly enable using this program for filing paper cuttings, etc. In spite of all these a program becoming popular depends on many factors. One requirement is the program to be totally intuitive to use and there should not be any bugs before release. Proper documentation is also a necessity.

The Taxa Identification program also falls in a similar niche. There is hardly any substitute now, identifying potential users and convincing them the utility of the program is the only way. Plant animal distribution can be achieved through GIS programs, but occasional users who just want their data plotted at the end of the study may find this program useful.

The program for displaying IRS images can be updated by including ability to read images from subsequent satellites. With the availability of data of equal or better quality free of cost from sources such as Global Land Cover Facility (GLCF), popularity of IRS images may decrease. For the DEM files tools for cutting, cropping and format conversion will have value for quite some time.

References

Borland (2003) Delphi User's Manual. USA. Clark University, USA.

Eastman, J.R. (2004). IDRISI Kilimanjaro Guide to GIS and Image Processing.

Saldanha, C, Rao, CK (1975). Punched card key to the dicot families of South India. Center for Taxonamic Studies, St. Joseph's College, Bangalore.

Mapinfo (2000) Mapinfo professional User's Guide. Mapinfo corporation. New York

NRSA, (1995).IRS1C Data Users hand book. National Remote Sensing Agency, Hyderabad.