

languages application use binary sort option provided by the database. Since the data through GIST SDK controls is directly stored in ISCII, there is no need for any conversion of data to sort it, and

- In case the data is in more than one Indian language, Indian languages are sorted on the top and then in English.

Language technology is continuously evolving. While the available tools can add value to the available solutions, better solutions for Indian language searching, and sorting will emerge, it is hoped, in the near future.

BIBLIOGRAPHY

1. <http://www.unicode.org>
2. <http://www.cdac.in>
3. <http://egranthalaya.kar.nic.in>
4. <http://www.microsoft.com/msdn>
5. <http://www.kagapa.org>

**INFORMATION RETRIEVAL STRATEGIES IN INDIAN LANGUAGES:
A GUIDELINE***

M. Chetana+ and P. Sunangala Devi+

Information retrieval in Indian languages is a crucial issue. The library of the Central Institute of Indian Languages handles materials in several languages and scripts. In the traditional card catalogue, in order to maintain uniformity, the bibliographical details were rendered in Roman script. In automating the library catalogue and other housekeeping activities of the library a challenging factor was to select a suitable software that would work with all Indian languages and scripts. This paper presents an overview of the key issues in information retrieval in Indian languages and guidelines to retrieve the needed information in any Indian language based on the experience at CIIL Library in using VIRTUA – a library automation software package, with UNICODE support.

1 INTRODUCTION

Information retrieval is the art and science of searching for information in documents or searching within databases, whether relational stand alone databases or hypertext networked databases such as the Internet or intranets, for text, sound, images or data. India is rich with 1652 spoken languages of which 22 are scheduled languages. Documents are published in all the scheduled languages. Developments in information technology influence the handling of Indian languages. In the traditional card catalogue, in order to maintain uniformity, the bibliographical details were rendered in Roman script. In automating the library catalogue and other activities of the library a challenging factor was to select a suitable software that would work with all Indian languages.

2 CIIL LIBRARY

The library selected for the present study is the Central Institute of Indian Languages Library (CIIL Library). Its future plan is to make the library an advanced research centre in language and linguistic studies. The library collection focuses on linguistics and books in all Indian languages, literature, folklore etc. are available. The institute has academic users of national and international repute

* Based on a paper to the National Seminar on Information/Knowledge Organization in the Humanities, 9-11 August 2005, Bangalore

+ *Resource Persons*, Central Institute of Indian Language, Manasgangotri, Mysore 570 006.
E-mail: Chetana_inadhu@yahoo.com; Sunabegur@gmail.com

and the collection of the library is highly specialized in these areas. The library collection of CIIL and its regional language centres exceeds 2 lakh books and more than 60% of the total collection is in Indian languages. After testing the performance of various library management software, and their capability to manage Indian languages, the VIRTUA Management System developed by VTLS Inc. was selected. VTLS stands for Virginia Tech Library System now known as Visionary Technology in Library Solutions, from Virginia USA. This has a branch at Noida (UP), India. VTLS provides the experience, scope, and vision to take today's libraries and information centers into tomorrow's possibilities.

Hardware

Sun Fire-280 servers

Solaris platform 9

Software

Virtua software, version 43

Windows XP operating system

Virtua is an integrated user-friendly, multi-user and comprehensive library automation system certified by ISO 9001 having innovative technology and global perspective.

3 FEATURES OF VIRTUA

- To support the full MARC (Machine Readable Cataloging) format,
- To implement the coded holdings standard for serials,
- To implement linked authority files, and
- To provide UNICODE support for all Indian and foreign languages.

Virtua is a three tier client/server architecture consisting of:

- Database warehousing,
- UNICODE language support, and
- Network optimized application.

To handle three aspects, namely:

- Data management,
- Software development, and
- Network delivery.

Data management relates to preparation of worksheets using the MARC format, software support refers to UNICODE support. Virtua software has inbuilt support for UNICODE. The data can be entered in the original language. Sorting and indexing can be done in the original language which is mandatory. Network relates to networking all the regional centres.

VTLS software comprises two main aspects: a) Virtua, and b) iPortal.

Virtua is for the library staff to input the data such as, work covering the entire library house-keeping activities like:

- Acquisition
- Cataloguing
- Circulation
- Serials control
- OPAC

Since the objective of the present study is information retrieval in Indian languages, only cataloguing and OPAC are discussed.

4 OBJECTIVES

Information retrieval in Indian languages is a crucial issue. Though in a few library management packages bibliographic data in Indian languages could be entered using Leap Office, but the out put will not be in Indian languages. This article presents an overview of the key issues in information retrieval in Indian languages and guidelines to retrieve the needed information in any Indian language irrespective of discipline. The objectives of the present article are to provide guidelines for:

- Browse searching?
- Keyword searching,
- Use of Boolean operators/Boolean search,
- Expert searching,
- Category searching,
- Truncation searching,
- Parentheses searching,
- Phrase searching.

- First field operator,
- Complete field operator,
- Proximity searching, and
- Heading Keyword Searching.

5 CATALOGUE

The main tool for information retrieval is the catalogue of documents. Technically it is known as bibliographic control. In the digital environment this function is executed using tailor-made worksheets in the international standard MARC format. In this format each of the fields, such as language, author, title etc., have fixed tags. The worksheet can be modified according to the library's needs. In addition to the defined MARC tags, all the tags after 900 are for local variations. Many of the documents do have accompanying materials such as CD-ROM, Maps etc., an online link that is located in 856 tag and 880 tag is used to input cataloging details in the original language.

Main function in cataloguing is the retrospective conversion work. The retrospective conversion of bibliographic data is the process of transferring the entire cataloguing/bibliographic data to computerized form using an international standard MARC21 format as mentioned above. A tailor-made worksheet with MARC21 tags representing each field is prepared. This MARC21 tags are unique. For example, tag 100 represents "Author" irrespective of libraries. The worksheet was prepared to suit the needs of the CIIL library. The data entered in the worksheet were transferred to the Virtua software. The newly received books are entered in the acquisition stage itself and the bibliographic information is fed and automatically added to the database.

6 OPAC (ON LINE PUBLIC ACCESS CATALOG)

OPAC refers to online access to the bibliographical records of the library's holdings with the access points being the same as those in a card catalogue. But permits use of more information retrieval strategies than the card catalogue.

6.1 Information Retrieval Strategies

To retrieve the information from the database the search strategies mentioned in section 4 are used.

Before working with OPAC, the language of the interface used for searching can be changed. Once a language is selected from the language list, all buttons, prompts and search windows are changed to the selected language. Changing the language selection displays all text-including headers, footers, button labels, field labels and search field list in that language.

6.2 Browse Searching

In browse search, when the user enters a keyword for search from a selected category, the system retrieves all the authority headings, controlled vocabulary and cross-reference structure in the database that match or are near-matches to the search terms. A user can narrow the search from a broad topic to specific topic. Basic search allows searching fields such as author, title, subject, course ID and call number.

6.3 Search Strings

- In author-wise searching entry element precedes the other components of the name:
Ex: Tagore, Rabindranath
- In title-wise search stop words like - a, an, the, to, etc, are not used as the first word of the title
- In subject search type, the concept terms appear as in the subject heading field 650 tag or LCSH (Library of Congress Subject Heading)
- Course ID search uses the official instructor name and number
- Call number search allows search by Classification Number including the local call number
- In the browse search 4 steps to be followed:
 1. Enter the search term (Tagore, Rabindranath),
 2. Select search category (author),
 3. Choose database and location (CIIL), and
 4. Then press the search button.

For example: tagore, rabindranath

Search result

- 8 Tagore, Rabindranath
- 213 Tagore, Rabindranath 1861-1941
- Tagore, Rabindranath Sir 1861-1941
- 2 Tagore, Ravindranath
- 2 + Tagore, Ravindranath 1861-1941
- 1 Tagore, Sandip Kumar
- 6 + Tagore, Saumyendranath
- 4 Tagore, Sunil
- 1 Tagore, Surendranath
- 2 Tagore, Upendra

This search output displays the headings that closely match the search term Tagore, Rabindranath, followed by other closely matching terms alphabetically irrespective of language. The number of records in the database associated with each term is listed under hit count and cross reference is also displayed. If the author has a pseudonym or an acronym, the controlled field has facility to provide "See" and "See also" references respectively which will work on the linkage to related synonyms. For the convenience of users "See" entry is represented by default minus mark (-) and 'See also' entry is by plus (+) mark.

6.4 Keyword Search

Keyword search is search for bibliographic records that contain one or more keywords in an indexed searchable field. In this search, queries can use any term that is in the index for a given category such as author, title, subject, ISBN, ISSN, LCCN, anywhere and language. Language search has to be followed by the code that has been adopted by the software. The language and their codes. For example:

Language	Code	Language	Code
Arabic	ara	English	eng
Assamese	asm	Gujarathi	guj
Bengali	ben	Hindi	hin
Dogri	doi	Kannada	kan

6.5 Boolean Operators

Boolean search logic is used in most of the public access systems to specify combinations of terms to be linked to synonyms and related terms in controlled

vocabulary and with spelling variants and in natural language searches. Terms can be combined using Boolean operators 'AND(&)' 'OR(+)' 'NOT(-)' to retrieve documents.

In many instances, a search retrieves far too long too many records. The search can be narrowed down by combining search terms with the 'AND' operators.

The 'AND' operator finds records that match both terms in the query. If any of the search terms is not contained in a records, that records does not appear in the results list. For example, If a user wants to view 'ಸಂಸ್ಕೃತ ಕವಿ ಚರಿತೆ' written by 'ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್' concept terms can be combined using Boolean c 'AND'.

For example: 'ಸಂಸ್ಕೃತ ಕವಿ ಚರಿತೆ' AND 'ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್'

Title = 'ಸಂಸ್ಕೃತ ಕವಿ ಚರಿತೆ'(smskruta kavi charitre)

Author = 'ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್' (Srinivasa Iyengar)

Search result

Author	Srinivasa Iyengar
Author	ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್
Title	Sanskruta kavicharithre/ Srinivasa Iyengar
Title	ಸಂಸ್ಕೃತ ಕವಿ ಚರಿತೆ/ ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್
Publisher	ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್, 1986.

Sometimes a browse search retrieves a disappointing result. The query may not locate any matching records, the list of results may contain too few items and may fail to retrieve the required information.. In this case, the query may need to be broadened using Boolean OR. The records retrieved will contain one or all of the concepts. For example, if a users want to view ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್'s books or ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್ the Boolean OR may be used to expand or broaden the search by including synonyms and related terms.

For example: ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್ OR ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್

Title = ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್ (Thukkural)

Author = ಶ್ರೀನಿವಾಸ ಅಯ್ಯಂಗಾರ್ (Thiruvalluvar)

Author Thiruvalluvar
 Author திருவள்ளுவர்
 Title Tirukkural/ Thiruvalluvar
 திருக்குறள்/ திருவள்ளுவர்
 வி. வி. அழகர்.
 Publisher திருச்சிராப்பள்ளி: ஸ்ரீராம சிறுவர்கள், 1982.

This search will retrieve records on *Tirukkural* irrespective of language and author or Thiruvalluvar irrespective of language and title. It also retrieves records *Tirukkural* written by Thiruvalluvar. It is used to combine words or sets that deal with a single search concept.

To exclude irrelevant records (that is, records containing concepts not of interest to the user) from a search, users can use the 'NOT' modifier as part of the query. For example, a reader looking for records about Tamil grammar and a search for the term grammar retrieves records relating to Tamil phonetics as well, those records can be excluded by forming the following query:

For example: Grammar AND tam NOT phonetics

Subject = Grammar

Language = Tamil

Subject = phonetics

Author Rajagopalacharya, K

Author ராஜகோபாலாச்சாரியார், க

Title Ilakkana vilakkam: punariyal / K.

Rajagopalacharyar

Title இலக்கண விலக்கம்: புனரியல் / கே.

ராஜகோபாலாச்சாரியார் அழகுநெய்யல்,

Publisher சென்னை: எல்.நி. பிரதீபம், 1982.

Subject Tamil language

Subject Tamil language Grammar

This search will retrieve only the records 'grammar in Tamil language'. But it will not retrieve any records containing subjects like phonetics even if it is also in Tamil language.

6.6 Expert Search

Expert search allows users to enter precision search queries. A precision search is a type of keyword search which specifies search categories manually and links operators in the text of the search query. It also allows experienced users to build sophisticated searches using a combination of search categories, keywords, parentheses and Boolean operators (AND, OR, NOT and NEAR) or symbols (&, +, _ and &x) etc.

6.7 Search Commands

- Category Searching,
- Parentheses Searching,
- Wildcards Searching,
- Phrase Searching,
- First Field Operator, and
- Complete Field Operator.

6.7.1. Category Search

Search category codes are defined in the keyword index configuration of the cataloguing parameters. Category codes specify categories within precision search strings.

Search Category Codes

Code	Category	Code	Category
a:	author	pp:	place of publication
pa:	personal name	nt:	notes
anc:	author name, corporate author	g:	general
t:	title	aw:	anywhere (all indexed fields)
s:	subject	nc:	nature of contents
p:	publisher	tes:	set

For example: t: *புராணங்கள்*

title=*புராணங்கள் (Brajbhaskha)*

If the search is for title 'Brajbhaskha' there is within the search string the category code 't' which is always followed by a colon (:). There is no space

between category code ':', colon (:) and the keyword 'Brajhasha'. If category code is not mentioned, it will search all categories (field) indexed in a keyword search.

Search Result

Author Verma, Dharendra
Author वर्मा, धरेंद्र
Title Brajhasha / Dherendra Verma
Title ब्रजशा / धरेंद्र वर्मा
Publisher Allahabad: Hindustani Academy, 1954.

In this case the retrieved record title has only 'Brajhasha' irrespective of the language.

6.7.2. Parentheses Search

Parentheses can be used anywhere within a query string to group a set of operators and search terms. Search strings within the parentheses will be processed first and then those results will be combined with other terms.

For example: s:novel & (aw:kannada + aw:tamil)

Subject = novel

anywhere = Kannada bandopadhyaya, tarasankara

anywhere = Tamil

Search Result

Author Jegasiripyan
Author ஜெகசிறியன்
Title Aravathu Thaagam/ Jegasiripian
Title ஆறாவது தாகம்/ ஜெகசிறியன்
Publisher சென்னை: சாரணி பதிப்பகம், 1977.
Subject Literature Tamil
Subject Novel

Author Bandopadhyaya, Tarasankara 1898-1971
Author ಬಂಡೋಪಧ್ಯಾಯ, ತರಾಸಂಕರ
Title Ivanalla / Tarasankara Bandyopadhyaya ; translator, B. Puttaswamaiah
Title ಇವನಲ್ಲ/ ತರಾಸಂಕರ ಬಂಡೋಪಧ್ಯಾಯ ; ಅನುವಾದಕ, ಬಿ. ಪುಟ್ಟಸ್ವಾಮಯ್ಯ
Publisher ಬೆಂಗಳೂರು: ಹೊಸತನ ಸಾಹಿತ್ಯ ಪ್ರಕಾಶನ, 1967.
Subject Kannada literature
Subject Novel Social novel
Subject Novels

This search retrieves all records containing novel as a subject and have the term Kannada and Tamil in anywhere (category).

6.7.3. Wild Card Search

Using wild cards with search terms can specify left, right or medial truncation. There are two types of truncation marks.

Asterisk(*)

Question (?)

Truncation is indicated by placing an asterisk (*) and question (?) sign immediately after the last root character.

The asterisk is useful for matching terms with the same root or prefix. For example, when users do not know the correct spelling or variations of spellings or search for all variations of a certain terms, truncation mark can be used.

For example: kav*

This query will retrieve records that match 'kav*', as well as records that match terms like kavi, kavana, kavite, kavigalu, etc. Once users have found the correct term they can submit a new query to limit the search to the records of their interest.

The question mark is useful to substitute for characters in a particular position in a term. For example, if a user is searching for a record for a document by or on

someone named Krishnamurthi but is not sure if the name is spelled Krishnamurthi or Krishnamoorti, then this query can be:

Krishnam?????

The search will retrieve records that contain either Krishnamurthi or Krishnamoorti

6.7.4. Phrase Search

Phrase search category query can be applied to all keywords within the double quotation mark.

For example: "ब्राह्मण व्याकरण की रूप रेखा"

It will retrieve records that contain "Brajhasha vyakaran ki ruprekha" anywhere or irrespective of the category in the records. If users want to use category code, they can use with double quotation mark. Ex t: "Brajhasha vyakaran ki ruprekha". This search will list records which contain the phrase "Brajhasha vyakaran ki ruprekha" in the title only.

6.7.5. First Field Operator

If, for example, a document is by more than one author and so described in the catalogue record, e.g. Kuvempu and Masi Venkatesh Iyengar, users have to search Kuvempu as a first field operator. Then they can specify the first in field operator by inserting =: after the category code.

For example: a=:Kuvempu

The query will retrieve all the records where Kuvempu appears as the first or at the beginning of the author category. If Masi Venkatesh Iyengar is searched as a first field operator, it will not retrieve because it is not the beginning in the author category.

6.7.6. Complete Field Operator

If a user wants to view a Telugu book on 'అకాశా దీపం' but if the author of this book is not known, then 'Aakasa deepam' can be specified in complete field operator search query.

For example: t^:'అకాశా దీపం'

It will retrieve all the records where 'Aakasa deepam' appears as the complete field operator.

6.7.7. Heading Keyword Search

Heading Keyword search allows user to enter a keyword query to search for the terms anywhere in the authority heading index. A heading keyword search searches on a single index of authority index. And search always returns a list of authority headings. Users can use wild cards search. But precision search is not allowed in heading keywords search.

In the search expression 'women AND Kannada literature', one can be specify data format by all words or phrase or exact match.

For example: Women AND Kannada literature

Subject = women (all words)

Subject = Kannada literature (all words)

It will retrieve records containing the term 'Women and Kannada literature' in the subject category.

6.7.8. Limiting Search Filters

The iPortal interface allows refining searches by limiting search filters. They provide six types of filters by nature of contents, format, language, location, place of publication, year of publication. Click on the set filter in the toolbar and select parameter before clicking on the search button.

6.8 Saving Search Strategy

Saving searches allows storing search sets that have been created during iPortal search session. Users can then logoff and search at a later time without having to retype each search set manually. Users can save a search set temporarily or permanently.

6.8.1. Saving and Retrieving Search History

Click the Save Search History button on the Main Search Page to store the sets that have been created during iPortal session.

In the Save Current Search page, type a short, one-word name to describe search history in the Name text entry box. Click the Temporary button to save

search history for 24 hours. After this time it will automatically be erased. Click the permanent button to save search for an indefinite period of time or until one deletes it.

One can include a comment by typing it in the comment box; use feature to describe the contents and purpose of the particular search. Click the Save Search button at the bottom of the page to store search history. To re-run the search history, any workstation may be used; it must login with the same user ID. Click the Run Search History button on the Main Search Page and follow the prompts.

6.9 Printing, Emailing or Saving Search Results

Readers have the option of saving, e-mailing and printing records in full, brief, item, or MARC format for the duration of the iPortal session.

7 CONCLUSION

Some important aspects of the information retrieval using a powerful library management software has been discussed in relation to Indian languages. From broad search to pinpoint search is made possible here in the original language. Knowing English is not mandatory.

BIBLIOGRAPHY

1. Eftimiadis, E.N. Online public access catalogues: characteristics of the literature. *Journal of Information Science*. Vol. 16; 1990; p107-112.
2. Krishnamurthy, M. Online Public Access Catalog through Libsys: an experience. In *DRTC Annual Seminar on Electronic Sources of Information*. 1-3 March 2000.
3. Roose, Tina. Online catalogs: Making them better reference tools. *Library Journal*. Vol. 113(20); 1998; p76-77.
4. *VTLS training guide 2002*. VTLS Inc., Noida, 2002.
5. <http://www.infosocous.com>

INFORMATION/KNOWLEDGE ORGANIZATION IN A FOLK MUSEUM*

N. Rajashekara+

The study focuses on the folk collection of one of the well-known folklore museum affiliated to the University of Mysore. The collection of the museum comprises of mementos preserved over generations by professional collectors or curators. The present study recommends classification methods in organizing the artifacts and suggests some of the recent techniques in the digital era that could be applied to folk museums in general. Also stresses the need of library environment in the museum for easy and time saving retrieval of information hidden therein.

1 INTRODUCTION

Folk museum is a place where the collections of articles used by our ancestors, are preserved and placed in a systematic order to serve the needs of researchers. It is considered as a real study research center. Folk museums were first established in Stockholm in 1893.

There is a similarity between library and museum, both are meant for storing or collecting or housing of materials. In ancient period books were stored or housed in a safe place. Aim of the library at that time was to preserve the collection for future use. We can call museum as a "Pre-Paper library". After the invention of printing technology in the 15th century by Johannes Gutenberg, library collections grew and the resources were made available to users. By adopting this new technology library has grown into an important institution, an Information or Knowledge resource centre. In the same way we can also apply the new digital technology to museums and the collected material can be digitized and utilized for research purposes.

Understanding, appreciating and sharing another culture's folklore transcends race, colour, class, and creed more effectively than any other single aspect of our lives and, as an element of our past and present society, it is something we can all relate to. Its value is no less than any other part of our history and heritage and as such, it must be documented and preserved as a legacy for our future.

* Based on a paper to the National Seminar on Information/Knowledge Organization in the Humanities, 9-11 August 2005, Bangalore.

+ *Documentation Assistant*, Central Institute of Indian Languages, Manasagangotri, Mysore 570 006. E-mail: rajju2722@gmail.com