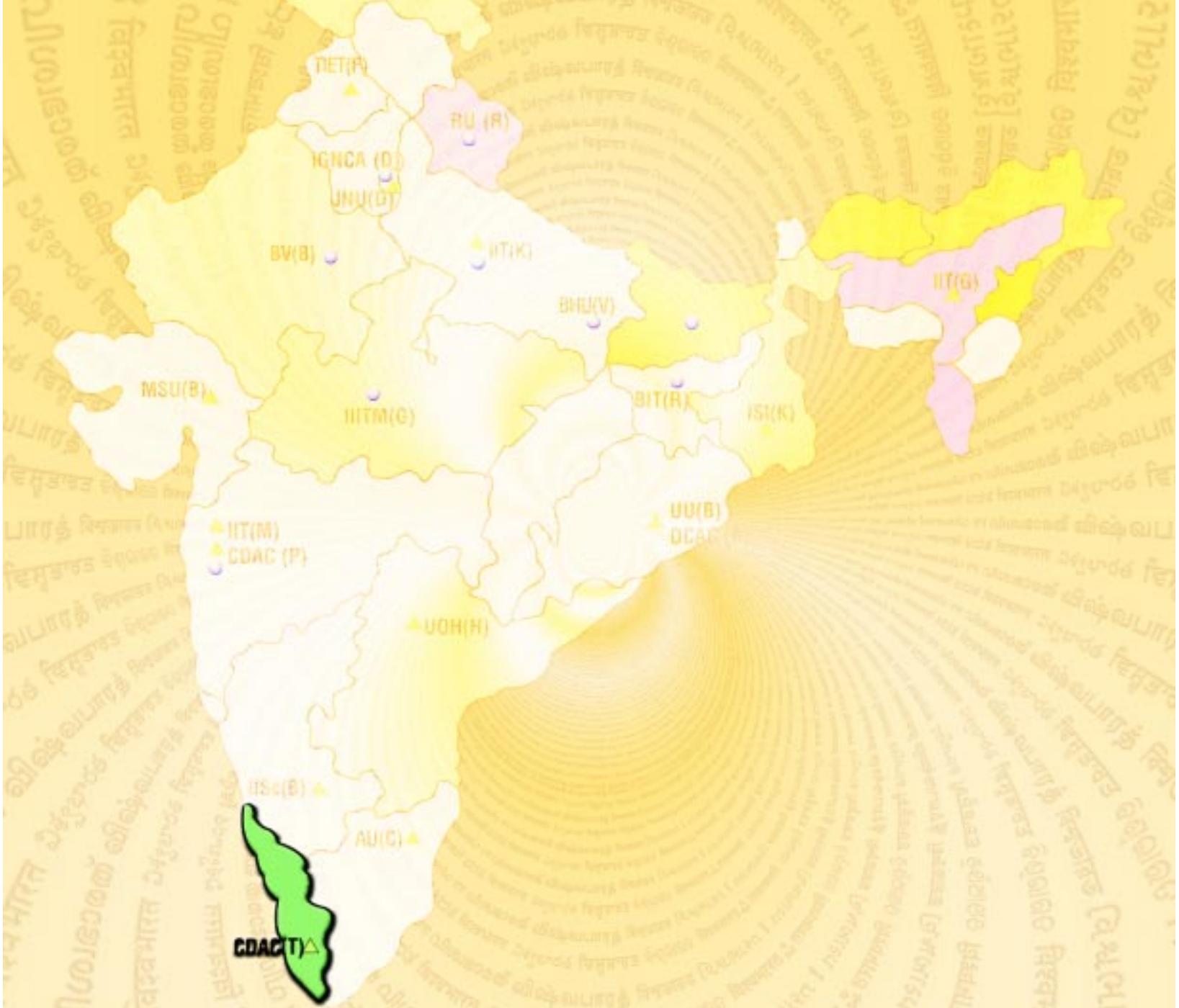


Resource Centre For  
Indian Language Technology Solutions – Malayalam  
C-DAC, Thiruvananthapuram

*Achievements*



Centre for Development of Advanced Computing  
(Formerly ER&DCI(T)), Vellayambalam,  
Thiruvananthapuram, Kerala, India.  
Tel. : 00-91-471-2723333 Extn., 243, 303  
E-mail : [contact@malayalabhasha.com](mailto:contact@malayalabhasha.com)  
Website : <http://www.ercitvm.org/hdg/mrc.htm>  
<http://www.malayalabhasha.com>



## RCILTS-Malayalam C-DAC, Thiruvananthapuram

### Introduction

C-DAC, Thiruvananthapuram, formerly ER&DCI, Thiruvananthapuram, is one of the thirteen Resource Centres (Resource Centre for Indian Language Technology Solutions) set up across the country by the Ministry of Communications and Information Technology, Govt. of India under the TDIL (Technology Development for Indian Languages) programme. These thirteen Resource Centres are aimed at taking IT to masses in their local languages and cater to all the constitutionally recognised Indian and some foreign languages. The Language of focus at C-DAC Thiruvananthapuram is Malayalam, the official language of the state of Kerala.

The main objectives of the “Resource Centre for Indian Language Technology Solutions – Malayalam” (RCILTS-Malayalam) are to build competence and expertise in the proliferation of Information Technology using Malayalam, the regional language of the state of Kerala. Development of Malayalam enabled core technologies and products would give a tremendous fillip to IT enabled services in the state. The comprehensive IT solutions developed would enable the citizens of Kerala to enhance their quality of life using the benefits of modern computer and communications technology through Malayalam. This will help them to better understand their own culture and heritage, interact with government departments and local bodies more effectively, besides obtaining a host of other advantages.

Now that the Resource Centre has completed three years of functioning, we are happy to note that we have achieved significant progress and have been able to complete development of all the expected core deliverables of the project with in the scheduled time.

We have been successful in developing a variety of tools and technologies for Malayalam computerization and taking IT to the common Malayalee in his local language.

Many of the products developed under the Resource Centre project are first of its kind and are significant for enabling Malayalam computerisation. They have got good market potential in the present scenario of computerisation and conversion of official language to Malayalam in the state of Kerala. Various Government departments have purchased our “Aksharamaala” software for Malayalam word processing. Ezuthachan, the Malayalam Tutor, also have good demand among non resident Malayalees. We have already sold 25 copies of the same and are trying for marketing tie up with some business houses.

The range of our products include:

**Knowledge Resources** like Malayalam Corpora, Trilingual (English-Hindi-Malayalam) Online Dictionary and Knowledge bases for Literature, Art and Culture of Kerala.

**Knowledge Tools** for Malayalam such as Portal, Fonts, Morphological Analyser, Spell checker, Text Editor, Search Engine and Code Converters.

**Human Machine Interface systems** comprising of Optical Character Recognition and Text to Speech Systems

**Services** like E-Commerce application and E-Mail Server in Malayalam and Language Tutors for Malayalam and English.

In addition regular training courses are being conducted as part of Language Technology Human Resource Development. Also, there is regular interaction with the Government of Kerala for providing solutions in the area of standardization, computerisation of various Government departments and converting the official language to Malayalam. We have been providing consultancy to individuals and

organisations regarding Language Technology applications. Given below is a detailed description of the products developed and the achievements of the Malayalam Resource Centre. Our Resource Centre developed the following Technologies/ Products.

### 1. Human Machine Interface Systems

#### 1.1 NAYANA™ - Optical Character Recognition System for Malayalam

**NAYANA™**  
MALAYALAM OPTICAL CHARACTER RECOGNITION SYSTEM

NAYANA is a product that enables the user to convert printed Malayalam documents to editable computer files. This system is very simple to use and requires no prior expertise.

**FEATURES**

- Nayana processes all types of printed Malayalam Documents.
- Supports TIFF and BMP image formats.
- Supports document Images with resolution 300 dpi and above.
- Detection and correction of document skew of -5° to +5°.
- The output document can be stored in both ISCI and ISPOC form.
- The output document can be saved as TXT, RTF, HTML or ACI file formats.
- User friendly interface.
- Recognition speed of 50 char /sec.

**APPLICATIONS**

- Conversion of printed documents to editable text.
- Optical Character Recognition combined with Text-To-Speech technology can be used for text reading system.

**EXPANDABILITY**

- A layout analyzer can be added to the system to reproduce the input document in its original layout.
- Can be expanded to cater to hand written and old document.

**RESOURCE CENTRE FOR INDIAN LANGUAGE TECHNOLOGY SOLUTIONS - MALAYALAM**  
ELECTRONICS ROAD CENTRE OF WSA, THIRUVANANTHAPURAM - 695012, KERALA, INDIA.  
TEL: +91-471-272910/272929/31 043, 363 748, +91-471-2725446  
Website: <http://www.malayaladirectoriescentre.org>  
e-mail: [cent@malayaladirectoriescentre.org](mailto:cent@malayaladirectoriescentre.org)

The Malayalam OCR system converts scanned images of printed Malayalam documents to editable text. It is a multi font system that works across a range of font sizes. The system has a recognition speed of fifty characters per second.

The System consists of a preprocessing module, the OCR engine and a post processing module.

The block diagram of the system is given in figure 1.

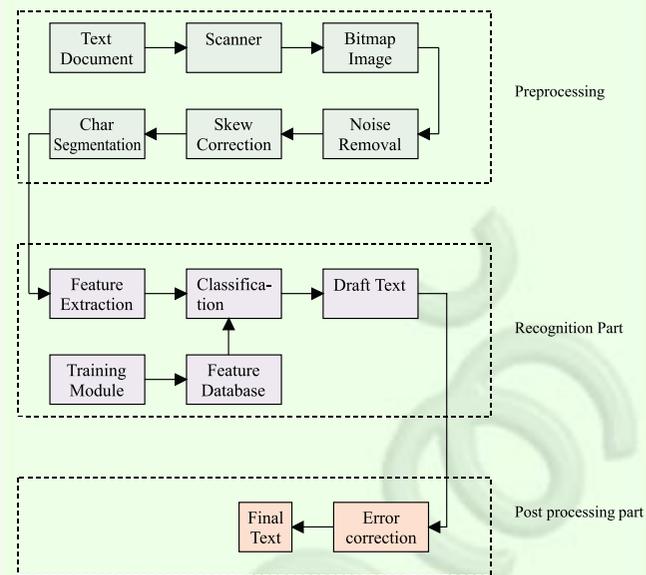


Figure. 1

The preprocessing tasks performed by the first module include noise removal, conversion of grey scale image to binary, skew detection and correction and line, word, and character segmentation. The scanned images in grey tone are converted into two-tone (binary) images using a histogram based thresholding approach (Otsu's algorithm). Skew detection is done using the Projection profile based technique. After estimating the skew angle the skew is corrected by rotating the image against the estimated skew angle.

The OCR engine (Character Recognition Module) is based on the Feature Extraction method of character recognition. Feature extraction can be considered as finding a set of vectors, which effectively represent the information content of a character. The features are selected in such a way that they help in discriminating between characters. A multistage classification procedure is used, which reduces the processing time while maintaining the accuracy.

After passing through different stages of the classifier, the character is identified and corresponding character code is assigned. A training module is incorporated in the OCR engine to recognize characters, which are different from normal characters in their shape and style (example- decorative fonts).

In the post processing module, Linguistic rules are applied to the recognised text to correct classification errors. For example, certain characters never occur at the beginning of a word and if found so, they are remapped appropriately. Similarly, dependent vowel signs can occur only with consonants or consonant conjuncts; if found along with vowels or soft consonants, they are remapped into consonants/conjuncts similar in shape to the vowel sign. Independent vowels occur only at the beginning of a word and if found anywhere else, they will be mapped into a consonant or ligature having similar shape.

#### Performance of the OCR

Developed on VC++ platform, Our Malayalam OCR runs on Windows 98/2000. It recognises 50 characters per second and gives an accuracy of 97% for good quality printed documents. The specifications and performance of the system is given below.

- Skew detection and correction : - 5 to +5 degree
- Supported image formats : BMP, TIFF
- Image scan resolution : 300dpi and above.
- Document Type : Single-font single size.
- Supporting Fonts**
- Fonts Names : CDAC Fonts (ML-TTKarthika, MLW-TTKarthika) Mathrubhumi Font, Manorama Font, Fonts used by DC Books
- Font Size : 12-20
- Font Styles : Normal, BOLD
- Supported Code format : ISCII/ISFOC

Supported output format : RTF/HTML/ACI/TXT

#### Character recognition accuracy (%)

Document Type	Good quality Paper	Bad quality Paper
Computer Printed Document	97%	94%
Magazine	92%	90%
Newspaper	85%	82%
Books	95%	93%

Table1

Extensive testing has been done on approximately 500 pages of different quality printed documents. Table1 consolidates the results of testing. The system has undergone certification testing at ETDC Chennai.

#### Applications

The Malayalam OCR can be integrated with a Malayalam Text to speech system to get a Text Reading System for the visually challenged. Other application areas include publishing sector, content creation, digital library, corpus development etc.

#### 1.2 SUBHASHINI™- Malayalam Text to Speech System (TTS)

The Malayalam Text to Speech system SUBHASHINI™ is a Windows based software, which converts Malayalam Text files into fairly intelligible speech output. The software is integrated with a text editor having both ISCII and ISFOC support. The editor supports INSCRIPT Key board layout.

The TTS is based on Speech synthesis by diaphonic concatenation and consists of the following four modules

- Diaphone Database

- Text Processing module
- Prosodic Rules Generator
- Speech Synthesiser

The Diaphone database consists of 2500 diaphones segmented from recorded words. All the commonly used allophones are also considered.

The text-processing module organizes the input sentences into manageable lists of words. It also identifies the punctuation symbols, abbreviation, acronyms and digits in the input data and tags the input data. These are then processed and converted to phonetic language – a language that the speech engine is able to recognise.

Rules for adding prosody to the speech output are generated using the Speech corpus. This includes the pitch and delay informations for different intonations.

The concatenation of diaphones corresponding to the text is done in the Synthesis module and we get speech output. We are using the MBROLA speech engine for speech synthesis.

### Applications

Text reading systems, announcement Systems, and systems providing voice interface.

### 2. Knowledge Tools

#### 2.1 NERPADAM™ - Malayalam Spell Checker

Nerpadam is a software subsystem that can be integrated with Microsoft word as a macro or the Malayalam editor stylepad developed by us, to check the spelling of words in a Malayalam text file. While running as a macro in word, it functions as an offline spell checker in the sense that one can use this software with a previously typed text file only. Both off line and online

checking are possible when it is integrated with the text editor. It generates suggestions for wrongly spelt words.

The system adapts a rule cum dictionary-based approach for spell checking. It incorporates a fully developed Morphological Analyser for Malayalam. This module splits the input word into root word, suffixes, post positions etc. and checks the validity of each using the rule database. Finally it will check the dictionary to find whether the root word is present in the dictionary. If anything goes wrong in this checking it is detected as an error and the error word is reprocessed to get 3 to 4 valid words, which are displayed as suggestion. The user can add new words into a personalised data base file, which can be added to the dictionary if required.

### Applications

All Malayalam word processing jobs.

### 2.2 AKSHARAMAALA™ Malayalam Font Package and Script Manager

The “AKSHARAMAALA” software package consists of two sub packages, “MSM” a Malayalam Script Manager and “Vahini”, a Malayalam font package. This package complies with the standard INSCRIPT keyboard layout, Phonetic Keyboard Layout and the ISFOC standard. This package enables the use of Windows based applications for Malayalam data processing. Some of the packages supported by this application are MS Office, PageMaker, Adobe Illustrator, MS Frontpage, Macromedia Dreamweaver Corel Draw, and Lotus Smart Suite. This package is intended for use under Windows 95, 98, NT, 2000, ME and XP.

The “MSM” package consists of a keyboard manager, which supports the INSCRIPT keyboard overlay and a Phonetic keyboard overlay for the entry of Malayalam characters. The Manager can optionally switch on the entry of Malayalam or English characters with the help of



a switchover key. The appropriate key combinations automatically render the Malayalam characters, such as conjuncts and soft consonants. The keyboard manager is designed to work with Malayalam ISFOC, monolingual as well as bi-lingual fonts. This manager also supports the use of Web fonts for data entry. Different options are provided in the package that turns on these features.

The “Vahini” package is a collection of Malayalam fonts that can be used for word processing, web publishing, data processing etc. and for use in MS-Windows applications. There are number of high quality True-type fonts . The fonts comply with the ISFOC standard and consist of both monolingual and bilingual fonts. A few web fonts are also provided for use in web pages. The bi-lingual fonts are those that contain both Malayalam and English characters. The webfonts are monolingual TTF fonts specially designed for the use with browsers.

We have also developed Dynamic (PFR) fonts, which can be used in web page development so that the user can view the web content without the font being actually installed in the machines. These fonts are platform independent and will work in both I.E and Netscape.

A Unicode font in open Type format is also available.

Government of Kerala is considering the AKSHARAMAALA package to teach Malayalam Data entry to the beneficiaries of “Akshaya”, a project aimed at providing Computer literacy to at least one person per family and also in the E-Governance applications of the government of Kerala.

The AKSHARAMAALA package can be used with standard content creation tools to develop contents in Malayalam.

**2.3 Text Editor:** A basic Malayalam Text Editor “STYLEPAD” is developed. It incorporates all the facilities available in Notepad together with provision to save Malayalam documents in ISCII format and read ISCII files.

**2.4 Code Converters:** Content creation in Malayalam is accomplished using different fonts by different organisations/individuals. Most of the On-line Malayalam Newspapers have their own proprietary fonts. Lack of standards in font coding makes data retrieval a difficult task. In order to alleviate this task, we have developed Font Converters for most of the commonly used Malayalam fonts. Given below is a list of the Font Converters developed by us.

MATWEB	↔	ISCII
THOOLIKA	↔	ISCII
SHREELIPI	↔	ISCII
AKRUTI	↔	ISCII
TULASI	↔	ISCII

MLBW (ISFOC) ↔ ISCII

MLW (ISFOC) ↔ ISCII

MLTT (ISFOC) ↔ ISCII

### 2.5 ANWESHANAM™ – Malayalam Web based Search Engine

“Anweshanam” is a directory based search engine, which searches for Malayalam content and information in web pages. This solution provides a Malayalam interface that helps the user to search for information quickly and easily on the web. It searches for specific Malayalam keywords and generates a list of links to web pages containing the searched information. It searches for specific content in the pages on the server. This solution was developed using Java Server pages.

The interface is in Malayalam and includes easy input facilities with keyboard driver and floating



character map. The keyboard driver supports the INSCRIPT keyboard with automatic formation of conjuncts. The application provides facilities for searching either English or Malayalam keywords. The result pages contain links to the web pages containing the searched information along with their brief description in Malayalam. A form is also provided whereby new pages and sites can be added to the database with details of the content.

The application can be expanded to function as a multilingual search engine and also as a meaning based search engine.

This application is beneficial to Multilingual portal developers and is useful for any one looking for some particular Malayalam content on the web.

This application is presently being run as a service on our website and portal, [www.malayalamresourcecentre.org](http://www.malayalamresourcecentre.org).

## 2.6 Malayalam Portal

We have designed and developed a Malayalam Web portal, [www.malayalamresourcecentre.org](http://www.malayalamresourcecentre.org) alias [www.malayalabhasha.com](http://www.malayalabhasha.com). Some of the facilities/ contents provided in the portal are:

- Malayalam version of the Constitution of India.
- A Newspaper “Pradeepam”, published from Kozhikode.
- A knowledge base of traditional Kerala Art forms and Culture (Both English and Malayalam)

Website and Portal [www.malayalamresourcecentre.org](http://www.malayalamresourcecentre.org) contains information on Malayalam language tools, resources and knowledgebases.

**CONTENTS**

- Daily News updates in Malayalam.
- Malayalam e-mail Service.
- Malayalam Search Engine.
- Malayalam version of Constitution of India.
- Multilingual Ayurvedic Classics.
- Solved S.S.L.C. Question papers in Malayalam.
- Forms commonly employed in Government.
- Knowledge base on Art forms & Culture of Kerala.
- Knowledge base on Malayalam Literature.
- Links on various Malayalam resources on the web.
- WAP enabled Tourist Information Guide.

**FEATURES**

- Easy navigational facilities.
- Designed for faster downloads.
- Easy input facilities for interactive services.
- Printer friendly format for downloadable content.

**RESOURCE CENTRE FOR INDIAN LANGUAGE TECHNOLOGY SOLUTIONS - MALAYALAM**  
 ELECTRONICS R&D CENTRE OF INDIA, THIRUVANANTHAPURAM - 695 022, KERALA, INDIA.  
 TEL: +91-471-270893/270894 (X) 243, 363 FAX: +91-471-270895  
 Website: <http://www.malayalamresourcecentre.org>  
 e-mail: [cont@malayalabhasha.com](mailto:cont@malayalabhasha.com)

**KALAKERALAM**  
 കലാകേരളം  
 KNOWLEDGE BASE OF TRADITIONAL ART FORMS & CULTURE OF KERALA

Kerala is an enchantingly beautiful, emerald green driver of land, flanked by the Western Ghats on one side, the Arabian Sea on the other, and strewn with rivers, lagoons, backwaters and rich vegetation in between. Various class and ethnic groups have influenced the culture of Kerala over a period of years. Cultural invasions led to the blending of various art forms. Koodiyattam for example is a fusion of the local Dravidian culture and some Aryan influences. Krishnattam, Kathakali, Thullal, Mohiniattam etc. are some of the local dances that arose out of the fusion. Among these, Koodiyattam, the highly stylised art form of Kerala has been recognised by the UNESCO as "a masterpiece of the oral and intangible heritage of humanity to be protected, preserved and promoted." Here we try to portray the Cultural and Classical Art Forms of Kerala.

**POTENTIAL BENEFICIARIES**

- Children of new generation.
- Research students.

**FEATURES**

- Multimedia Package.
- Large number of art forms covered.
- Knowledge base in Malayalam.
- Video clips for better understanding.
- Image gallery of traditional musical instruments.

**EXPANDABILITY**

- More research on each art form.
- e-Learning of traditional art forms.

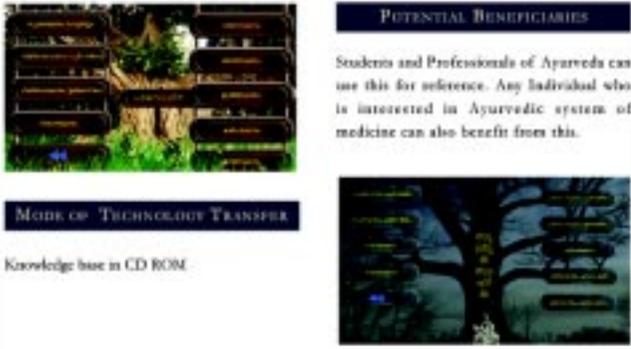
**RESOURCE CENTRE FOR INDIAN LANGUAGE TECHNOLOGY SOLUTIONS - MALAYALAM**  
 ELECTRONICS R&D CENTRE OF INDIA, THIRUVANANTHAPURAM - 695 022, KERALA, INDIA.  
 TEL: +91-471-270893/270894 (X) 243, 363 FAX: +91-471-270895  
 Website: <http://www.malayalamresourcecentre.org>  
 e-mail: [cont@malayalabhasha.com](mailto:cont@malayalabhasha.com)

- A Knowledge base of Malayalam Literature.
- Malayalam Literary classic “KRISHNAGATHA” by Cherusseri and the Grammatical classic, “KERALAPANINEEYAM” by A.R. Rajaraja Varma.




The ancient science of Ayurveda is the oldest known form of health care in the world. Often called the mother of all healing, it originated in India some 5000 or more years ago. Recently Ayurveda has been having a profound impact upon the world of health care. The potential of this ancient healing system to bring healing to those suffering with chronic diseases has been recognised worldwide. The Ayurvedic Classics Charakasamhita and Susrutasamhita contain about 10,000 shloka. The former is a major work on internal medicine and the latter describes highly developed surgery.

These classics along with the Malayalam interpretations of the shloka are made available in CD media.



**POTENTIAL BENEFICIARIES**

Students and Professionals of Ayurveda can use this for reference. Any individual who is interested in Ayurvedic system of medicine can also benefit from this.

**MODE OF TECHNOLOGY TRANSFER**

Knowledge base in CD ROM





**PRAKES**  
പ്രകൃതി  
മൂല്യ

BILINGUAL EXPERT SYSTEM FOR ESTIMATION OF PRAKRUTHI

PRAKES is a software package for estimating Prakruti (constitution) of a person as decided in Ayurvedic System of Medicine. The concept of Prakruti is very powerful in Ayurveda. The word Prakruti means nature (natural form) of the build and constitution of the human body. A balanced state of the Tridosha (Vata, Pitta and Kapha) makes a healthy and balanced person (Physically and Mentally). Since we all have different combinations of the Tridosha, Ayurveda shows that the path to optimal health is different for each person depending on their Prakruti. The diagnosis of Prakruti offers unique insights into understanding and assessing one's health. It is not merely a diagnostic device but also a guide to action for good health. The expert system package developed by ER&DCI is a comprehensive scientific tool for estimating Prakruti. It assesses the dominance of Tridosha and give advice for preventive and promotive health care.



**POTENTIAL BENEFICIARIES**

Professionals and students of Ayurveda, Primary Health Centres and any individual interested in Ayurveda.

**MODE OF TECHNOLOGY TRANSFER**

Software package with setup instructions in CD media.

**FEATURES**

- Available in English and Malayalam.
- Interactive menu-driven interface.
- Examines the presence of Akshana.
- Assesses dominance of the tridosha.
- Advice on preventive and promotive health care depending on dominance of the dosha.
- The interactions and results are recorded along with the bio data of the person for future reference.
- Hard copy print out of the results.




- Full text of Sanskrit Ayurvedic Classics Charakasamhita and Susrutasamhita (transliterated in Malayalam and ten other languages), along with their Malayalam interpretations.
- PRAKES (Prakruthi estimate) - An Interactive software package for estimating the Prakruti (constitution) of a person based on Ayurvedic concepts. Available in both English and Malayalam.
- A database of forms commonly employed by Govt. of Kerala (City Corporations, Motor Vehicles, Revenue, Civil Supplies Depts., etc.) Altogether 76 forms from different government departments put on the web.
- SSLC (Mathematics and Science) Question papers and Answers for the last five years in Malayalam.

- A knowledge base in Malayalam for rubber cultivators.
- A tourist aid package named “Explore-Kerala”. The site, <http://www.malayalamresourcecentre.org/wap/> can be accessed from WAP enabled mobile phones.
- Malayalam brochures on Cancer Awareness

In addition, some of the tools/technologies developed (Sandesam, Anweshanam, Dictionary and E commerce application) have been integrated in the portal for demonstration.

The contents of the website are continuously being upgraded and enhanced.

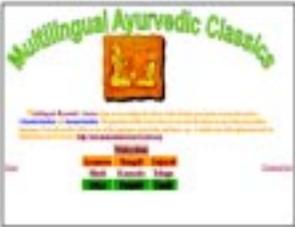
The number of visitors to the site has exceeded 20,500 since June, 2002.



## MULTILINGUAL AYURVEDIC KNOWLEDGE BASE

The ancient science of Ayurveda is the oldest known form of health care in the world. Often called the mother of all healing, it originated in India some 5000 or more years ago. Recently Ayurveda has been having a profound impact upon the world of health care. The potential of this ancient healing system to bring healing to those suffering with chronic diseases has been recognised worldwide. The Ayurvedic Classics Charakasamhita and Susrutasamhita contain about 10,000 shloka. The former is a major work on internal medicine and the latter describes highly developed surgery.

These classics along with the Malayalam interpretations of the shloka are made available in CD media. Translated versions of the shloka in all major Indian Languages is also made available.



**POTENTIAL BENEFICIARIES**

Students and Professionals of Ayurveda can use this for reference. Any Individual who is interested in Ayurvedic system of medicine can also benefit from this.



**POTENTIAL BENEFICIARIES**

- ☛ Small and medium ISPs.
- ☛ Business Organisations.
- ☛ Government Sectors.
- ☛ Multilingual Portal Developers

**MODE OF TECHNOLOGY TRANSFER**

Knowledge base in CD ROM



**RESOURCE CENTRE FOR INDIAN LANGUAGE TECHNOLOGY SOLUTIONS - MALAYALAM**

ELECTRONICS PAD CENTRE OF IICET, THIRUVANANTHAPURAM - 695022, KERALA, INDIA  
 TEL: +91-477-2726133/272222 (3) LINES, 24x7 FAX: +91-477-2723436  
 Website: <http://www.malayalamsolutionscentre.org>  
 e-mail: [contact@rciltsmala.com](mailto:contact@rciltsmala.com)



## SANDESAM™

സന്देशം  
MALAYALAM Web-based Mail Server

Web mail solution for mail server based on Qmail, Vmailmgr and Courier-IMAP running on Redhat 7.0. It is a light weight and very fast solution to provide an easy-to-use web interface in Malayalam. It is based on *maildir* structure and directly reads mail from the server's disk. This solution is developed using JSP and the Javamail API and does not use Java on the client side. This gives the user complete access to his POP3 or IMAP mail boxes via an easy-to-use web interface.



**FEATURES**

- ☛ Qmail, Vmailmgr authentication support.
- ☛ 100% maildir based.
- ☛ IMAP support.
- ☛ Basic folders support.
- ☛ User manageable folders.
- ☛ Extensive MIME support.
- ☛ Easy administration.

**POTENTIAL BENEFICIARIES**

- ☛ Small and medium ISPs.
- ☛ Business Organisations.
- ☛ Government Sectors.
- ☛ Multilingual Portal Developers



**EXPANDABILITY**

The server can be extended to support any IMAP compliant mail server running on Linux platform.

**RESOURCE CENTRE FOR INDIAN LANGUAGE TECHNOLOGY SOLUTIONS - MALAYALAM**

ELECTRONICS PAD CENTRE OF IICET, THIRUVANANTHAPURAM - 695022, KERALA, INDIA  
 TEL: +91-477-2726133/272222 (3) LINES, 24x7 FAX: +91-477-2723436  
 Website: <http://www.malayalamsolutionscentre.org>  
 e-mail: [contact@rciltsmala.com](mailto:contact@rciltsmala.com)

### 3. Services

#### 3.1 SANDESAM™ - MALAYALAM E-MAIL SERVER

Sandesam is a solution for a Web based mail service in Malayalam. The back-end mail server comprises of, or is based on the Qmail, Vmailmgr and Courier-IMAP running on Redhat Linux. The Qmail server supports the SMTP and POP3 services, the Vmailmgr program performs the management of user accounts effectively, while the Courier –Imap server supports the IMAP service. The web interface is developed using Java Server Pages and the Java mail API and does not use Java on the client side. The database server used is Postgres SQL

This is a lightweight and a fast solution to provide an easy-to-use web interface in Malayalam. The storage of mailboxes is based on *maildir* structure

and the mail is read directly from the disks. The administration of user accounts is done through the Vmailmgr program. Authentication is also done via Vmailmgr for both Qmail and Courier servers. The service gives the user complete access to his POP3 or IMAP mailboxes via an easy-to-use web interface.

Some of the facilities provided in the service are the IMAP support with user manageable folders, extensive mime support for attachments. The interface is in Malayalam and includes easy input facilities with keyboard drivers and floating character maps. The service provides facilities for sending and receiving mail in Malayalam, storage of addresses in address book with Malayalam names and description. The service also provides facilities for user configuration like changing the password, setting quota for mailboxes etc.

This solution can be expanded to support any IMAP client mail server running on Linux or Windows platform.

This solution is beneficial for Small and medium ISPs, business organizations, Government Departments Multilingual portal developers.

This solution is presently being run as a service on our website and portal, [www.malayalamresourcecentre.org](http://www.malayalamresourcecentre.org) with the domain id [sandesham.com](http://sandesham.com).

### 3.2 Malayalam E-com Application

An e-commerce application in Malayalam has been developed to help computer literates not that proficient in English to purchase goods online. This solution is developed using Java Server pages. This solution is simple using pure html and JSP and is viewable in any browser.

The web interface is in Malayalam and basically contains a window with display of products for sale with their descriptions in Malayalam. A shopping cart is provided whereby goods to be purchased can be added and removed according to will. The online bill for the cart with price details can be viewed during the process and then the purchase can be finalized by filling up and submitting an order form. The payment can be made by cheque or DD giving its details in the order form and sending the instrument by post.

The application also considers inventory details by displaying available stock, automatically updating the stock after each purchase and prompting certain functions on reaching certain limits. The facilities provided in the web interface are easy input facilities like keyboard driver and floating character map for the entry of Malayalam data and acknowledgement of the purchase order by e-mail.

The application can be extended to support secure online payment with credit cards.

The application is useful for small or medium organizations to increase their sales coverage

through the Internet. It is useful for multilingual portal developers and is beneficial to computer literates not that proficient in English but familiar with Malayalam.

The application has been presently employed in our website and portal [www.malayalamresourcecentre.org](http://www.malayalamresourcecentre.org) for the sale of the products developed at the centre.

## 4. Knowledge Resources

### 4.1 Trilingual Dictionary

An Online Trilingual (English-Hindi-Malayalam) Dictionary is developed. It Contains 50,000 plus words in each Language. English based Search and advanced search facilities are implemented. Search based on the other two languages being implemented.

**ERBOCH**  
TRILINGUAL DICTIONARY

എറവോച്ച ട്രിഡിംഗ്ല ഡിക്ഷണറി  
ERBOCH TRILINGUAL DICTIONARY

The first online trilingual dictionary for English, Hindi and Malayalam. It contains more than 50,000 words in each language. The dictionary includes idioms, glossary of foreign words, usages in English, Hindi and Malayalam, etc.

The user can input words in any of the above languages and find the meaning in the other two languages. They can also search for words based on different criteria such as parts of speech, starts with or ends with a particular string, etc.

**FEATURES**

- Formable in XML format.
- ISCII Based.
- Retrieval based on Parts Of Speech (POS)
- Word search can be made in all three languages.
- Description of every word with example.
- Advanced search facilities.
- Extremely fast processing.

**APPLICATIONS**

- Integration with any other application or web portal.
- Aid for translation-both Machine aided and manual.
- Study other languages with relative ease.

**EXPANDABILITY**

- Switching to Unicode representation.
- Other languages can be added.

**RESOURCE CENTRE FOR INDIAN LANGUAGE TECHNOLOGY SOLUTIONS - MALAYALAM**  
ELECTRONICS PWD CENTRE OF INDIA, TERUVANANTHAPURAM - 686012, KERALA, INDIA.  
TEL: +91-471-272811/272822/272833/272844 FAX: +91-471-272848  
Website: <http://www.malayalamresourcecentre.org>  
E-mail: [erboch@erboch.com](mailto:erboch@erboch.com)

The main features of the Dictionary are:

- Portable in XML format.
- ISCII Based.
- Retrieval based on Parts of Speech
- (POS)
- Word search can be made in all three languages. (Implementation in progress)
- Description of every word with example.
- Advanced search facilities.
- Extremely fast processing.

The Dictionary can be integrated with any other application or web portal. It can also be used as an aid for translation – both Machine aided and manual.

It helps the study of the concerned languages with relative ease.

## 5. Language Tutors

### 5.1 Ezhuthachan- The Malayalam Tutor

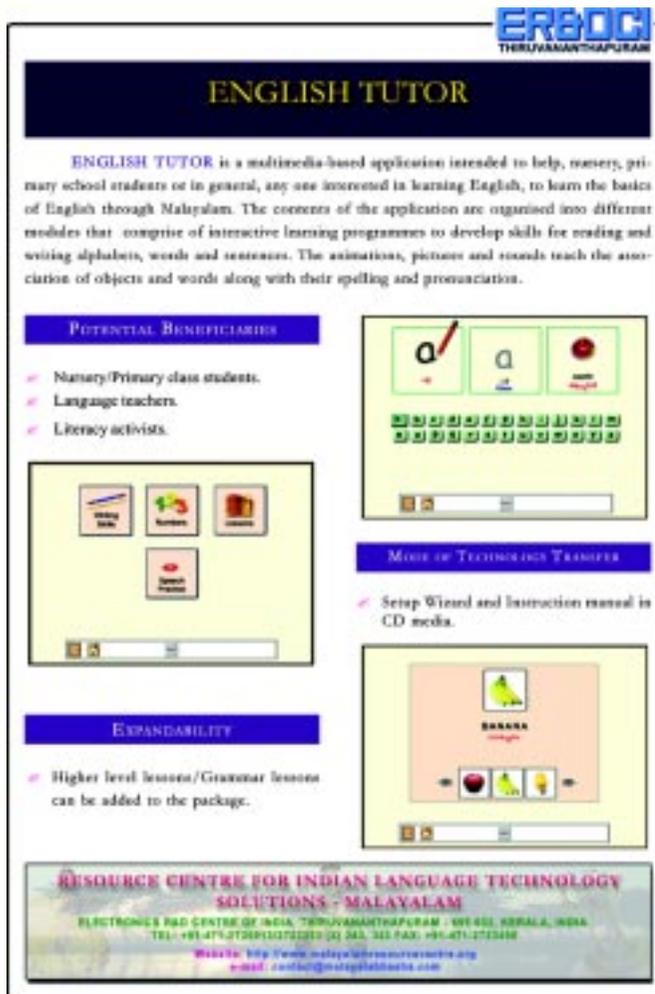
A Malayalam Tutor Package, “EZHUTHACHAN”, which is aimed at teaching Malayalam to foreigners and second generation Keralites living abroad. It is basically a multimedia package with animations, which show the method of writing the letters and sound giving the pronunciation of characters and words. A writing pad is also provided, where a shaded model of each character appears and the user can practice writing over this using the mouse. This application lists the commonly used expressions in Malayalam (words for daily use, knowing somebody, numbers, days, months, colours, animals, birds etc.), with their pronunciation and English meaning. The contents are formatted into well structured chapters. A test module is also provided at the end. The package also contains an English-Malayalam Dictionary of 2000 words.

The Honourable Minister of State for Urban Development & Poverty Alleviation, Shri. O.

Rajagopal (presently Minister of state for defence) formally released the package on 19<sup>th</sup> October, 2002. The Chief Executive of Non Resident Keralites Welfare Association (NORKWA), Shri. Satish Namboothiripad, IAS received a copy of EZHUTHACHAN from the Honourable Minister. NORKWA is considering the “EZHUTHACHAN” package for their programme of on-line teaching of Malayalam for Non resident Malayalee children. We are receiving lot of enquiries and have already sold some copies of the software.

### 5.2 English Tutor

English tutor is a multimedia-based application intended to help, nursery,/primary school students or in general, any one interested in learning English, to learn the basics of English through Malayalam. The contents of the application are organised into different modules that comprise of interactive learning programmes to develop



skills for reading and writing alphabets, words and sentences. The animations, pictures and sounds teach the association of objects and words along with their spelling and pronunciation

## 6. Other Activities

### 6.1 Providing Technology Solutions

C-DAC, Thiruvananthapuram has entered into a contract with Govt. of Kerala to be a Total Solution Provider for IT implementation in Government. The Government has already implemented “Project Grameen” for effective dissemination of IT to grass root level. C-DAC, Thiruvananthapuram together with Library Council has setup Community Information Centres in fourteen districts in Kerala. C-DAC, Thiruvananthapuram is the executing agency of the E-Governance projects of Government of Kerala.

### 6.2 Interaction with State Government

We have been active member of state-level Expert Committee for standardisation of Malayalam Keyboard and Character Encoding. The same committee has recommended the modifications to be made in the Malayalam UNICODE. The Committee has submitted its final report and the Government have approved it.

Recently Government of Kerala has launched a programme called “Akshaya” which is aimed at making at least one person per family in Kerala computer literate. C-DAC, Thiruvananthapuram is a member of the committee setup for evaluating the study material for this programme.

C-DAC, Thiruvananthapuram is Member of the General Council, IT@School Project of the Government of Kerala.

### 6.3 Training Activities

C-DAC, Thiruvananthapuram, together with Centre for Development of Imaging Technology (C-DIT) has conducted a two day work shop on Font Design in April 2002.

The ongoing computerisation of various state government departments, along with increased use of Malayalam as official language has created a need for a large number of computer literate manpower with proficiency in Malayalam data processing. A module on Malayalam word processing tools has been incorporated into the PGDCA course and other public and corporate training programmes offered by C-DAC, Thiruvananthapuram.

### 6.4 COWMAC

C-DAC, Thiruvananthapuram has formed a Consortium of Industries Working in the area of Malayalam Computing (COWMAC). The Consortium aims at mutual interaction among the participating agencies, so that standardisation in the areas of font coding, IT Vocabulary, Keyboard overlay, transliteration schemes etc. can be done effectively. It also aims at avoiding duplication of development activities by way of closer interaction between developers.

## 7. Publications

1. Malayalam Spell Checker - presented in the International Conference on "Universal Knowledge and Language" in Goa in November 2002.
2. Optical Character Recognition System for Printed Malayalam Documents and
3. Text to Speech System for Malayalam – presented in the SAP workshop at the Centre for Applied Linguistics and Translation Studies(CALTS) in the University of Hyderabad, in March 2003.
4. Text Reading System for Malayalam - Selected for presentation in the International Conference, "Information Technology : Prospects and Challenges in the 21<sup>st</sup> Century" in Kathmandu, Nepal in May, 2003.

We have also prepared the final proposal for modifications to be made in the next version of Malayalam Unicode and the Malayalam Design guide for TDIL.

## 8. Expertise Gained

The Malayalam resource Centre consists of a core team of Design engineers, programmers and linguists proficient in natural Language processing. We have gained expertise in the areas of Image Processing, Speech Synthesis, Font Design, Phonology of Malayalam language and Morphological Analysis of Malayalam language.

We have also built up technical capability in the use of Data base packages like MS Access , XML, Postgre SQL Scripts, Languages and tools : HTML, Java Server Pages, Java , and JavaScript, VC++, C++, VB, HTML, DHTML, Diaphone Studio, Macromedia Flash, C-DAC Iplugin & Leap Office, Macromedia Fontographer and Bitstream Webfont wizard 1.0

Expertise in configuring server side soft wares are also available at the Centre.

Web Servers : Linux Apache Server, Tomcat server

Mail Server : Qmail, Send mail

## 9. Future Plans

We plan to take up development of the following products as second phase of the Resource Centre Project.

- Online Character Recognition System for Malayalam
- Lexical Resources for Machine Translation
- Speech to Text
- Malayalam Word Net and
- Porting of the various applications developed to Linux platform.

## 10. The Team Members

Ravindra Kumar R	ravi@erdcitvm.org
Sulochana K.G	sulochana@erdcitvm.org
Jithesh .K	jithesh@erdcitvm.org
Jose Stephen	jose_stephen@erdcitvm.org
Santhosh Varghese	santhosh@erdcitvm.org
Shaji V.N	shaj@erdcitvm.org
Vipin C Das	vipincdas@erdcitvm.org
Keith Fernandes	keith@erdcitvm.org
Sanker C. S	sankercs@rediffmail.com
Sunil Babu	sunilbabu_ktr@yahoo.com
Hari Kumar	karamanahari@rediffmail.com
Neena M.S	neenams@rediffmail.com
Praveen V.L	praveenvl@erdcitvm.org
Mithra R.S.	shammymithra@rediffmail.com
Praveen Kumar	praveenvijay@rediffmail.com
Hari .K	harichithira@yahoo.com
Dr. V.R.Prabhodhachandran Nayar	pnayar@email.com
Dr. Usha Nambudripad	shanambudripad@yahoo.co.in

*Courtesy: Prof. Ravindra Kumar  
C-DAC, Vellayambalam*

*Thiruvananthapuram-695 033  
(RCILTS for Malayalam)*

*Tel: 00-91-471-2723333, 2725897, 2726718*

*E-mail : ravi@erdcitvm.org*