INFORMATION SYSTEMS IN
MALAYALAM
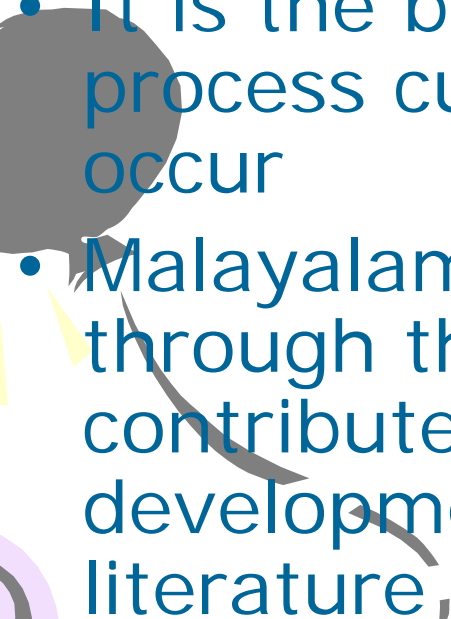
Dr. R. Raman Nair

University Librarian,
Mahatma Gandhi University, Kottayam -686 560
ramannair_r@yahoo.com

Language and Script

- Knowledge accumulated in a society is recorded in its language
- It is the base on which development process cultural, economic, political etc occur
- Malayalam language and script developed through thousands of years and has contributed like any other language to the development of Indian culture and literature,
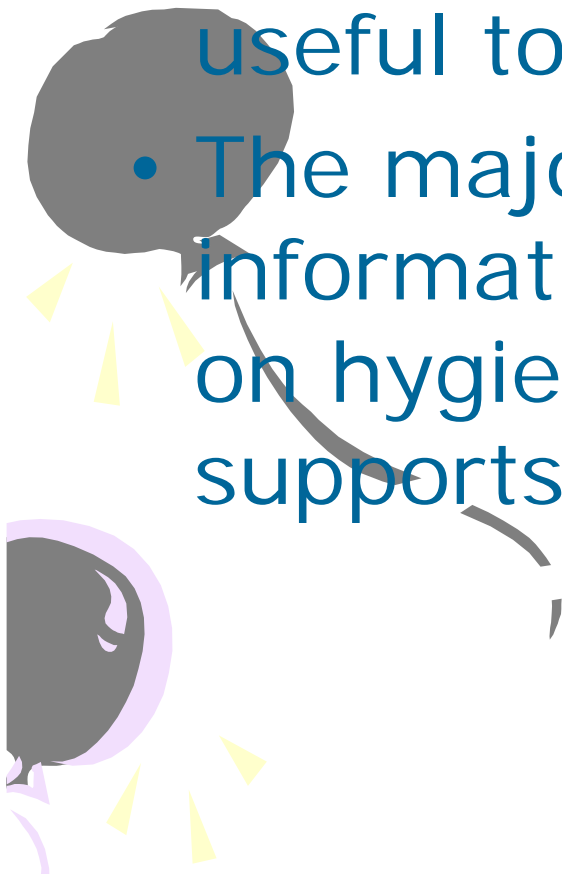
ICT/Language and Empowerment

- The existing Knowledge/Information Resource of the society is the capital for development
- So it needs to be equitably distributed
- ICT can contribute immensely to make it accessible to every one and hence it can empower and emancipate the economically and socially backward people
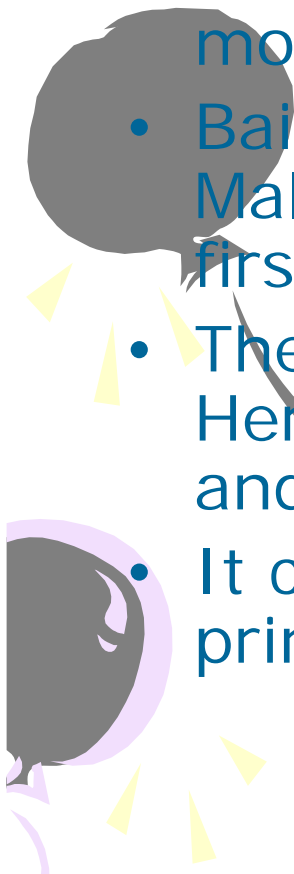- This needs ICT to use the language/script of the people

Information for the People

- Information systems should be in the language/script of the people to be useful to them

- The majority of people require vital information for their day to day life on hygiene, farming, subsidies, supports, educational facilities etc .

Malayalam Language/Script

- Got final shape four centuries back at the time Tunjath Ezhuttachan.
- 1821 Benjamin Bailey designed the first set of movable types of Malayalam.
- Bailey preserved unique characteristics of Malayalam lipi and conjuncts in designing the first font.
- The Malayalam types and characters he (and late Herman Gundert) designed was more scientific and systematic
- It continued without major modifications in printing and writing for one and half centuries.

Up to Typewriters in 1970

- Malayalam Language/script not modified to suit the needs of the machines

- Technology was developed to suit the specificities of our language and script.

- Even after typewriters came and till recently the original Malayalam lipi stood as our script; lipi of the people
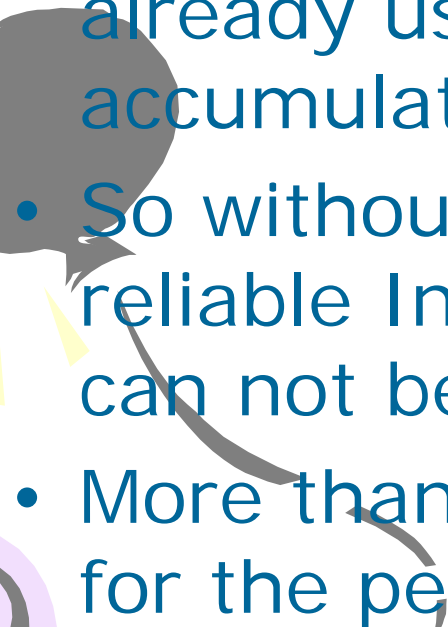
Existing Knowledge

- Society is fast becoming digitally networked which makes information Systems very important in day to day life of the people.

- Our recorded knowledge up to the present, 95% percent of the existing documents, millions of pages of information is in in our old script.

Local language/Script

- Information systems in Malayalam should have the capacity to process old script already used for recording the ocean of accumulated knowledge.

- So without the power to use old script reliable Information Systems in Malayalam can not be build up.

- More than that Information system meant for the people is to be in their language.

Systems Without Information

- Kerala has invested millions for computers, cables, LANs, WANs and V-Sats in the name of Information Infrastructure development.
- Established Internet Kiosks up to Panchayat level
- We claim these are going to make our information accessible to the un-reached.

What Villager Will Get

- Will one approaching the Internet Kiosk at a Panchayat office get information relevant to his needs?
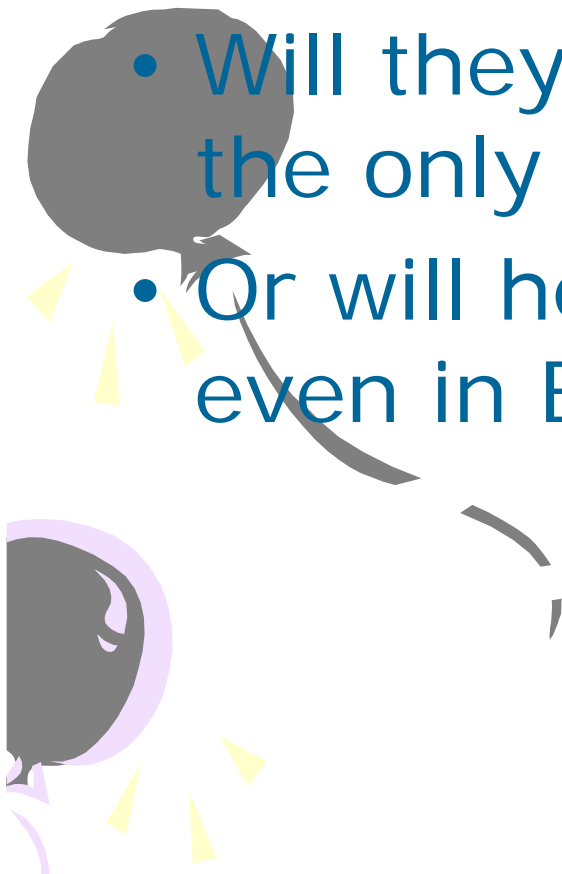
Villager's Needs

- higher education options available for one's daughter completing +2,
- Gazetteer of the region,
- the preventive measures for Denki, pesticides to control Mandari or Manjalippu disease of the coconut palms
- the support available from government to safe guard life and properties if Mullapperiyar Dam fails
- Information for farming and animal rearing

Will the Villager Get

- If he succeeds in getting any information
- Will they be in his Mother tongue, the only language he knows?
- Or will he get relevant information even in English

What He Can Access

- Within seconds at the Kiosk in the Panchayat he will get
- Any government document of USA,
- The educational facilities in UK,
- Any book from Library of Congress
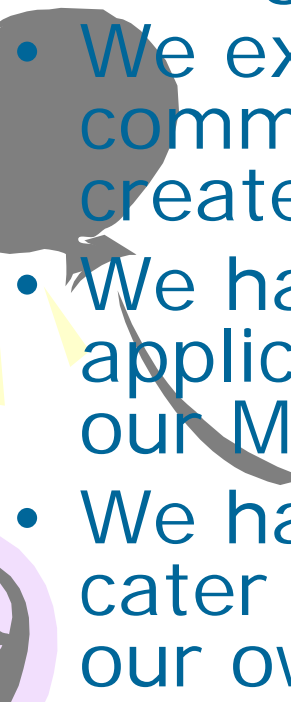- Market information of grains from a Japanese village in Japanese language

Where We Went Wrong

- Is this the way a Kerala villager should be equipped with ICT to get the information vital for his life and work, development?
- Where we went wrong?

We Disregarded Information

- Development of Information Infrastructures through two decades disregarded its 'information' part
- We expected computer and communication stacks to automatically create Information Systems
- We have never cared to develop applications for managing our content in our Mother tongue/Our script
- We have not developed content that can cater the requirement of our people, in our own language/script.

Lack of Language Technology

- Computer technology still could not process Malayalam
- 1970s New Lipi was developed to suite English typewriters
- Late 1980s when computers came language scholars thought computer to be a new kind of typewriter & not considered reviving original script
- Never cared to explore the functionality and aesthetics of the Original script
- Adapted the deformed alphabet 'New' Lipi' for computers also; leading to total chaos and confusion.
- 1999 Rachana Akshara Vedi proved the viability Original Scripts to solve problems in computer processing of Malayalam.
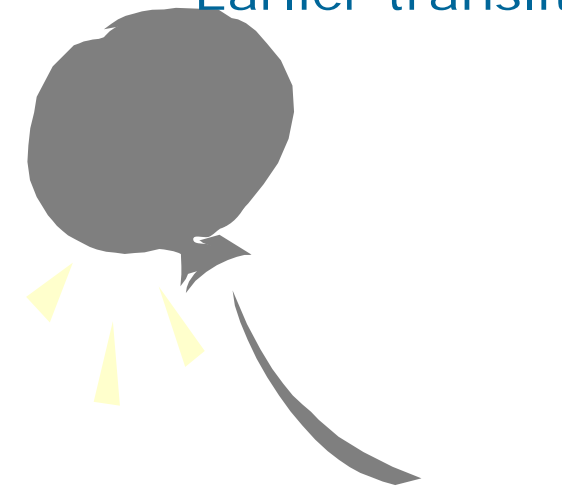
Granthasooji- Sri. K.M. Govi

- Earliest Information Systems in Indian languages were developed by librarians using traditional methods.
- K.M Govi's documentation of our literary heritage has no parallel in any Indian language.
- *Malayalam Granthasoochi* published by Sahithya Akademi in print runs into more than 10 volumes/10,000 pages.
- Complete information system of books ever published in Malayalam
- In no other Indian language such an initiative was ever completed. Institutions spending crores of rupees and years, could never achieve such a system.
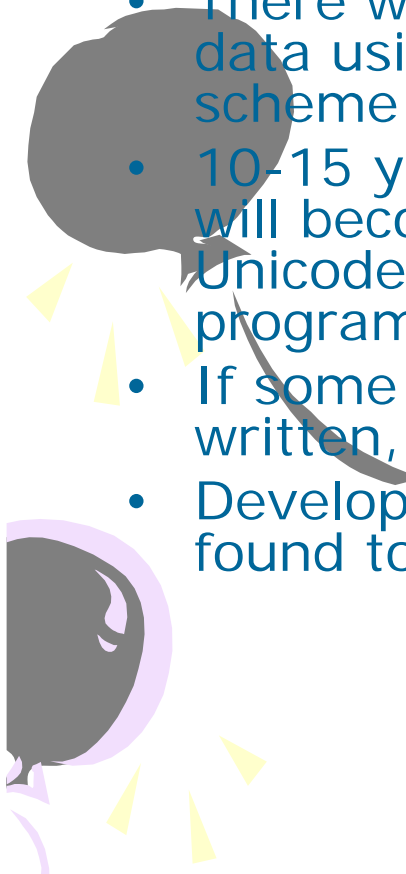
Early Computer Application

- In ICT systems library professionals are first to seriously encounter the difficulties in using scripts of languages other than English for indexing
- Earlier transliteration method was applied for decades.

Problems of Transliterated Data

- There was no standardized transliteration scheme. Entering data using one scheme and searching using different scheme results failure in retrieval.
- 10-15 years of transliterated data in computerized libraries will become useless and will need to be re-entered once Unicode based Malayalam/Multilingual database program becomes available
- If some standard not followed a transliterated word can be written, read and interpreted in dozen ways.
- Development of Information Systems using our script was found to be the only solution

Delay for Systems in Malayalam

- the official language scholars considered computer a new kind of typewriter
- For script technology development considered only DTP works and not digital applications.
- Never explored the functionality of Original script
- Result total chaos and confusion. .

Rachana Akshara Vedi

- It is in such a situation that those from library field like K.H. Hussain took initiatives for developing Malayalam language Technology that can be used in DBMS

- Rachana Akshara Vedi led by Mr. R. Chitrajakumar proved the power of Old Script for digital applications in 1999

DOS and ASCII Age

- Late 1980s: C-DAC designed Gist cards which are extra hardware components for Indian language processing.
- It makes Malayalam databases capable of search and retrieval using Malayalam script.
- Thrissur Public Library uses it
- is one such library. Store inventory control at DC Books, Kottayam had developed and used a database program using Gist and dBase2, processing all their published Malayalam titles using Malayalam script
- Late 1990s Rachana provided Malayalam Language Technology that can be used DBMS/Search Mechanisms

MS Windows and ASCII Age

- 2001: St. Mary's College, Thrissur developed its catalogue in MISIS based on new lipi, First attempt in Malayalam to localize CDS\ISIS

- 2001: Kerala Sahitya Akademy started Computerization of Grandhasoochi; 50,000 records Programmed in PowerBuilder; New Lipi of ISM- Gist fonts like Revathi and Kartika; not Published

- 2002: Malayala Sahithya Parmparyam of Kerala Sahitya Akademy

- 2003: Brennen College published its catalogue of 21000 records in CD with MISIS using Malayalam Old Script.

- 2004: Vimala College: Developed the Malayalam book's database using MISIS for internal use in library

Unicode Age

- Research and development on Unicode complaint script by Rachana Aksharavedi
- 2005, State Central library, Trivandrum developed an in-depth cataloguing of 800 old rare books in Malayalam and English using Nitya Archive.
- The first Unicode based Multilingual DBMS in Kerala. Used original/old script applying Rachana font
- 2005 DC Books Kottayam developed catalogue for their collection in Linux using Rachana
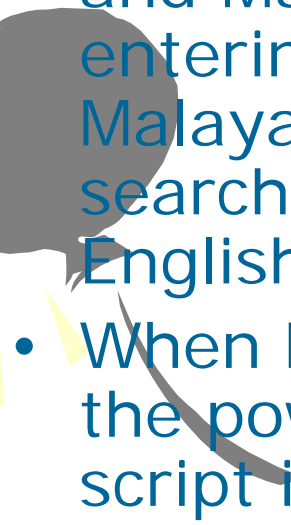
Centre for South Indian Studies (CSIS)

- Kerala Reference Library' a DL consisting of selected authoritative rare and antique books on Kerala's history and culture in Malayalam and English.
- Will be an authentic and reliable reference collection on Kerala affordable to more than 6000 rural libraries /schools/colleges of Kerala.
- Malayala Granthasooji; a catalogue of rare and antique Malayalam books / banned books of erstwhile Travancore and Cochin kingdoms etc available in collections of public and private organizations/individuals in India and abroad,.
- Both Unicode based Multilingual DBMS.

Why our script?

- Once Unicode takes a final decision on CHILLUS and Malayalam Code page is fixed for ever, the entering and rendering of Malayalam words using Malayalam script becomes standardized and search and retrieval becomes as natural as English

- When Malayalam language technology which has the power to process data in original Malayalam script in computers become available why should we go for non standards scripts

Essential Information Systems in Malayalam (General)

- Library Databases: Catalogue of books in many languages using Unicode based Multilingual DBMS
- Lexicological Information System -Multilingual
- Land/Revenue Records
- Right to Information: related systems and services

Essential Information Systems in Malayalam (Indigenous)

- Indigenous Knowledge - Nattarivu, Kattarivu, Kadalrivu etc.
- Indigenous Medicines –Otttamooli, Siddha, Unani, Ayurveda, Visha chikitsa
- Folk Arts
- Lexicological Information System -Multilingual
- Agriculture/Farm Information System

Research & Devt at C-DAC

- Nayana - Malayalam OCR System.
- Aksharamaala - Malayalam Font /Script Manager.
- Ezhuthachan - The Malayalam Tutor.
- Nerpadam - Malayalam Spell Checker.
- Subhashini - Malayalam TTS System.
- Aachaaryan - The English Tutor.
- Prakes - Prakruti Estimation Software.
- Sandesam - Malayalam Web-based Mail Server.
- Anweshanam - Malayalam Web Search Engine
- Kalakeralam - Knowledge base of Kerala.

Other Institutions/Projects

- Centre for Development of Imaging Technology (C-DIT) Trivandrum
- Rachana Aksharavedi
- Centre for Informatics Research and Development
- Centre for Agricultural Informatics
- Malayalam Lexicon Dept, University of Kerala
- Information Kerala Mission
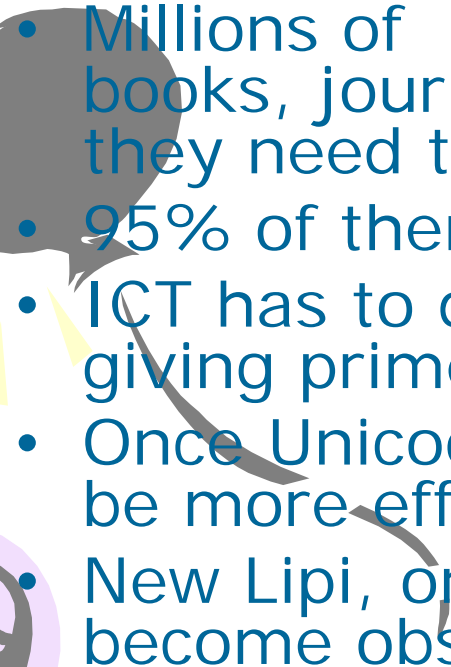- Encyclopedic Publications

Advanced Research at C-DAC

- **Malayalam** Optical Character Recognition (OCR) system. The presence of two different scripts (Old Script and New Script) and a large number of characters (including conjuncts) makes the Malayalam OCR a complex

- **Web Malayalam:** to solve the problems in the representation of phonetic symbols of Malayalam in html page.

- **Malayalam Spell Checker: to** check the spelling of the words in a Malayalam text file, validate them and in case of error, list out the right spelling in the form of suggestions.
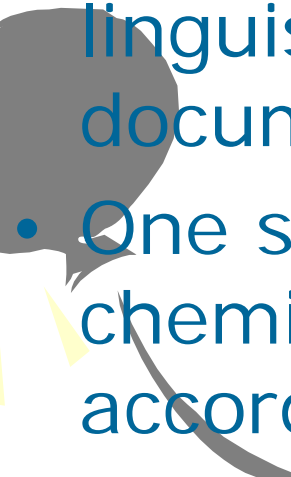
Original Script Should Survive

- We are moving to a less paper society
- Digital archiving will be major activity of Govt and organizations in the coming decades.
- Millions of pages like note files, revenue records, books, journals, etc are in Malayalam script and they need to be digitized and indexed.
- 95% of them are in Old/original script.
- ICT has to develop and use OCR spell check etc giving prime focus to Original script.
- Once Unicode standard is finalized, Old script can be more effectively used in computers
- New Lipi, only a subset of Original script, may become obsolete and will not stand.
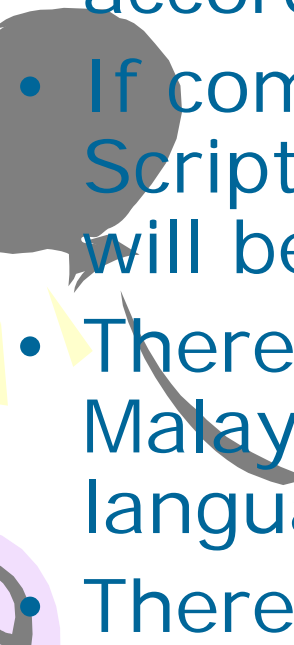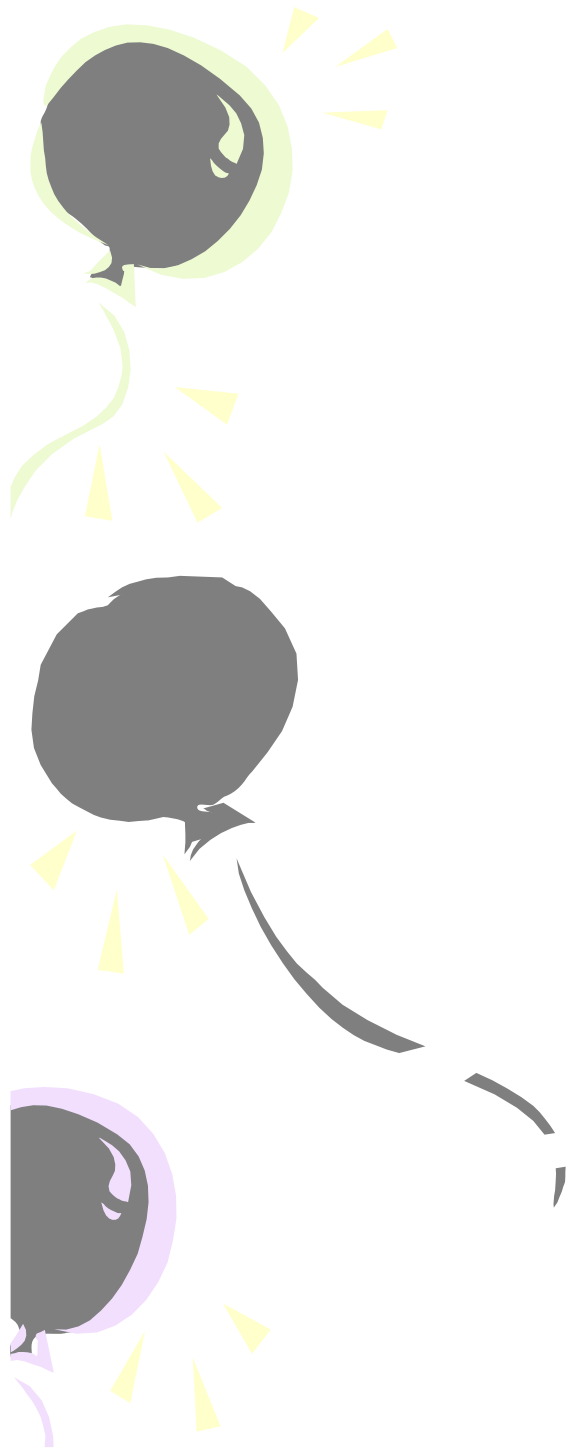
Tool Should Not Decide Master

- Though logical, computer is only a tool.
- Gives unprecedented capability/speed to linguists, historians, mathematicians, documentalists, etc. to do their work.
- One should not change his language or chemistry/physics/mathementics to fit it according to the needs of a machine
- Computer technology should be developed to do its master's work and not vice versa

Lipi: Our Heritage/Science

- For all Indian languages computer technology has developed programmes in accordance with their original script

- If computers attempt to deform Tamil Script which is like Mother for them; their will be hundreds of athamahuthis

- There is no other group of people like the Malayali who has no love for their lipi, language, culture.

- There lies our problem

Thanks