## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Introduction</strong></td>
<td>5</td>
</tr>
<tr>
<td>1.1</td>
<td>Character encoding</td>
<td>6</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Character encoding process</td>
<td>6</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Character Vs Glyph</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>ASCII</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>UNICODE</td>
<td>8</td>
</tr>
<tr>
<td>1.4</td>
<td>National Initiatives</td>
<td>11</td>
</tr>
<tr>
<td>1.5</td>
<td>Malayalam keyboard</td>
<td>11</td>
</tr>
<tr>
<td>2.</td>
<td>Malayalam character set</td>
<td>13</td>
</tr>
<tr>
<td>3.</td>
<td>Malayalam keyboard layout</td>
<td>16</td>
</tr>
<tr>
<td>3.1</td>
<td>Modifications in Inscript layout</td>
<td>18</td>
</tr>
<tr>
<td>3.2</td>
<td>Typing sequence</td>
<td>19</td>
</tr>
<tr>
<td>3.3</td>
<td>Conjuncts</td>
<td>23</td>
</tr>
<tr>
<td>3.4</td>
<td>Intra glyph positioning</td>
<td>26</td>
</tr>
<tr>
<td>4.</td>
<td>Malayalam character encoding</td>
<td>28</td>
</tr>
<tr>
<td>4.1</td>
<td>Modifications required in Unicode</td>
<td>29</td>
</tr>
<tr>
<td>4.2</td>
<td>Modifications required in ISCII</td>
<td>33</td>
</tr>
<tr>
<td>5.</td>
<td>Lexicographic ordering of Malayalam characters</td>
<td>37</td>
</tr>
<tr>
<td>5.1</td>
<td>Searching and sorting</td>
<td>37</td>
</tr>
<tr>
<td>6.</td>
<td>Other recommendations</td>
<td>43</td>
</tr>
<tr>
<td>6.1</td>
<td>Visual representation of characters</td>
<td>43</td>
</tr>
<tr>
<td>6.2</td>
<td>Malayalam Research Centre</td>
<td>43</td>
</tr>
<tr>
<td>6.3</td>
<td>Official Malayalam dictionary</td>
<td>44</td>
</tr>
<tr>
<td>6.4</td>
<td>Government initiatives required</td>
<td>44</td>
</tr>
</tbody>
</table>
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCII</td>
<td>American Standard Code for Information Interchange</td>
</tr>
<tr>
<td>BIS</td>
<td>Bureau of Indian Standards</td>
</tr>
<tr>
<td>CDAC</td>
<td>Centre for Development of Advanced Computing</td>
</tr>
<tr>
<td>DBMS</td>
<td>Database Management System</td>
</tr>
<tr>
<td>EDCDIC</td>
<td>Extended Binary Coded Decimal Interchange Code</td>
</tr>
<tr>
<td>HTML</td>
<td>Hyper Text Markup Language</td>
</tr>
<tr>
<td>ISCII</td>
<td>Indian Script Code for Information Interchange</td>
</tr>
<tr>
<td>ISO</td>
<td>International Standards Organisation</td>
</tr>
<tr>
<td>MIT</td>
<td>Ministry of Information Technology</td>
</tr>
<tr>
<td>OCR</td>
<td>Optical Character Recognition System</td>
</tr>
<tr>
<td>OS</td>
<td>Operating System</td>
</tr>
<tr>
<td>TDIL</td>
<td>Technology Development in Indian Languages</td>
</tr>
<tr>
<td>UTF</td>
<td>Unicode Transformation Format</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
</tbody>
</table>
Terminology

Alphabet: A set of letters used in writing.

Aspirated consonant: A consonant which is pronounced with an extra puff of air coming out at the time of release of oral obstruction. This has a sound of an extra ‘h’.

Basic alphabet: The minimal set of letters which can be used for uniquely encoding every word of a language. Bit: Binary digit. It can have only two values: 0 and 1.

Byte: Group of eight bits (Also referred to as octet).

Character: A symbol which can represent a letter, a numeral, a punctuation mark, a special symbol or even a control function.

Character code: Position in the code table of the character.

Character set: A set of characters grouped together for a purpose, like that of representing a script.

Code table: A table showing the positions allotted to individual characters from a character set.

Conjunct: A letter which is a combination of two or more letters.

Consonant: A letter representing a speech sound in which breath is at least partly obstructed, and which has to be combined with a vowel to form a syllable.

Display rendering: The process by which a string of characters is displayed (or printed).

Font: A collection of glyph images that share the basic design.

Glyph: A recognizable abstract graphic symbol.

Letter: A character representing one or more of the simple or compound sounds used in speech. It can be any of the alphabetic symbols.

Nasal consonant: A consonant pronounced with the breath passing through the nose.

Phonetic alphabet: An alphabet which has direct correspondence between letters and sounds.

Pure consonant: A consonant which does not have any vowel implicitly associated with it.

Script: A distinctive and complete set of characters used for the written form of language.
Syllable: A unit of pronunciation uttered without interruption, forming whole or part of a word, and usually having one vowel sound optionally surrounded by one or more consonants.

Vowel: A letter representing a speech sound made with the vibration of vocal cords, but without audible obstruction.

Vowel sign: A graphic character associated with a letter, to indicate a vowel to be associated with that character.
1. Introduction

Information Technology has become so pervasive that it is difficult to identify any area of human concern, which is not influenced by it. One of the major areas where the advancement of Information Technology is extensively used is text processing, storage and distribution. Preparing and processing of documents, including desk top publishing, is increasingly being done on computers and in the days to come, documents prepared by hand will be very rare except those prepared for very personal use.

Gone are the days where the documents were tied down to the computer in which it was prepared. With the advent of Internet and World Wide Web, people can access any document in any other computer anywhere in the world. Documents are, in general, meant for somebody else to read. If a second person has to read a document, then it has to be written in a format unambiguously understood by him. That is, there has to be a unique way of preparing the document, which is known both to the creator and the user of the text. Thus, if there is a common format in which documents are prepared, anybody who adhere to the standard will be able to read the document in any other computer.

Most of the information available on the web today is in English. There is clearly defined standard in English for inputting text through computer keyboard, for storing and retrieving data and in rendering (displaying) the text on the screen or paper. All the software developers, computer developers and users adhere to this standard. So, as far as English is concerned there is no need to worry about who created the document in what software package and what platform etc. Receiving and interpreting data in English, from whichever source it has come, is straightforward, thanks to the standard keyboard and character encoding accepted by all.

If the fruits of the breath-taking developments in Information Technology are to reach the common people in Kerala, the computer applications should be available in Malayalam, or at least with front end in Malayalam. We need to have a lot of applications in Malayalam – be it text processing, publishing, database applications, web applications or anything else – which can be ported to any other platform and which can be executed in any platform without any patches. This calls for two minimum requirements to be achieved - standardization of Malayalam keyboard layout and Malayalam character encoding. In the absence of these standards, the application developers will use their own schemes for keyboard layout and character encoding with the result that one document/application developed with one package cannot be read/executed in another platform/environment.

Today, there are many software packages available in Malayalam, mainly for word processing. The developers of these software packages have used different keyboard layouts and different character encoding schemes. As a result of this, a document prepared using one software package cannot be read using another package. Malayalam is also not supported in most of the widely used platforms, operating systems or application packages, because there is no standardised keyboard layout and character encoding. For the same reason, Malayalam content in the World Wide Web is also very few. The current situation is that one has to
download and install several Malayalam fonts, each with its unique mapping scheme, to be able to read Malayalam contents available on the Internet.

Government of Kerala has initiated various schemes aimed at bringing the Government to the people. Computerisation efforts of various Government departments and local bodies are in full swing. Content generation in Malayalam is a main component of all these projects. This makes it all the more important and urgent to define a standard keyboard layout and character encoding scheme for Malayalam, which is accepted by all.

On this background, Government of Kerala constituted a Committee as per GO (Rt) No.93/2000/ITD dated 2.6.2000 to recommend to the Government a Malayalam computer keyboard layout and Malayalam character encoding standard. The Committee was formed with Shri P.Govinda Pillai, Chairman, CDIT as the Chairman and Smt Aruna Sundararajan, Secretary, Department of Information Technology, Government of Kerala as Convener. Subsequently, some more experts on IT and linguistics members were co-opted to the Committee. The names of the members of the Committee are given in Annexure 1.

1.1 Character encoding

When a word processor user presses a key labeled “A” on the keyboard, he expects the character “A” to appear on the screen. He also expects “A” to appear in paper when he prints the text. He expects the character to be stored as “A” when he saves the text, and somebody else using a different computer to see this character as “A”, when he sends this text to that computer (over network).

Computer can operate only on numbers. So, any other input to the computer, including text, has to be converted to numbers before computer processing. When a word processor user types a key in the computer keyboard, the computer’s system software receives a message that the user pressed a key (or a combination of keys) for “A”, which it encodes as a number. The word processor stores this number in memory, and also passes it on to the display software responsible for putting the character on the screen. The display software, which may be a window manager or part of the word processor itself, uses the number as an index to find the image of “A”, which it draws on the monitor screen. The process continues as the user types in more characters.

To be of any use in computers, especially in computer communications and in particular on the World wide Web, characters must be encoded. In fact, much of the information processed by computers over the past few decades has been encoded text, exceptions being images, video and numeric data. Encoding can be loosely defined as mappings between the character sequences that users manipulate and the sequences of bits that computers manipulate.

1.1.1 Character encoding process

A character set encoding has to satisfy the primary requirement, that of the unambiguous representation of the content of the written text. To make it possible to successfully encode, process and interpret text a character set must:
• Define the smallest useful elements of text to be encoded. The units of encoding, the characters, are pragmatically chosen as appropriate to express text and allow various text processes. This set is called a repertoire.

• Assign a unique code to each element of the repertoire.

• Provide basic rules for encoding and interpreting text so that programs can successfully read and process text.

In some cases, the whole encoding process can be collapsed to a single step, a trivial one-to-one mapping from characters to bytes.

Because encoded text cannot be interpreted and processed without knowing the encoding, it is vitally important that the character encoding is known at all times and places where text is exchanged or stored. Mandating a unique encoding has strong virtues of simplicity, efficiency and robustness. Since the unique encoding is known implicitly from usage of such protocol or data format, the protocol or data format does not have to carry character-encoding tags.

1.1.2 Character Vs glyph

A writing system’s alphabet, numbers, punctuation and other writing marks consist of characters. A character is the symbolic representation of an element of a writing system - a letter, symbol or number. As soon as we write a character, it is no longer abstract but concrete. The exact shape by which the character is represented on screen and paper is called a glyph. They are the components used to generate the visible representation of characters. Glyphs need not correspond one-to-one with characters. A single character can be represented by multiple glyphs and a single glyph can represent multiple characters. Just like characters are the basic unit of organization of encoded text, glyphs are the basic unit of organization of the visual rendering text.

The character sets that are stored in the computer are independent of the display format. The character sets and the font/glyph are defined separately and there need not be a one-to-one relationship between them.

A font is a collection of glyphs, all of similar design, that constitute one way to represent the characters of the language. Different fonts for the same language will typically have different glyphs to represent the same character. The code positions of the glyphs are not important, as are positions of characters in a characters set. For example, Times New Roman font may have the general style on Macintosh or IBM-compatible PC, but the code positions of the underlying characters are significantly different on these systems after the first 128 characters. The character set differences are taken care of by the underlying system.

But in the Internet environment, this may not be applicable always. HTML encoded documents may be created on any computer system and read by any other. It is in this cross
platform environment that differences between the basic character sets of computers show through.

Most HTML content does not identify the encoding used in creating the document, nor do most HTTP servers transmit this information when it is available. As a result, browsers need to either guess at the encoding used or use a default value. Even if the encoding is specified and transmitted, browsers are limited in what they can do to present the document correctly. Encoding translation tables and suitable font resources are necessary for proper interpretation and display of documents.

Unambiguous communication of text is much simpler when some recognized standard is followed within a document. If the character encoding and interpreting scheme is unique and widely accepted, then this scheme can be made available in any computer and any document which uses this scheme can be correctly read and processed.

1.2 ASCII

ASCII is the most common format used for encoding text files in computers and on the Internet. The most common keyboard used is QWERTY keyboard, which most of us are familiar with. The ASCII-QWERTY combination has been the de-facto standard so far. In ASCII a 7-bit number represents each of alphabetic and numeric characters or special characters like punctuation marks. Thus a total of 128 (= 2^7) characters are defined. Capital letter A has a code value 65, B has a code value 66, numeric 1 has a code value of 49 and so on.

ASCII encoding scheme lets a computer store a document as a series of numbers and also lets the computer share such documents with other computers that use ASCII system of coding. These stored characters can be processed using any text editor, word processor and other applications including database and web. They can be also be sent through any email system on the Internet.

Since it can define a maximum of only 128 characters, the 7-bit ASCII standard severely limits the repertoire of characters available for multilingual applications. EDCDIC (Extended Binary Coded Interchange Code) is an 8-bit character set, which can represent 256 characters. In some encoding schemes for other languages, the first 128 bits of 8 bit coded characters are used for English characters and the next 128 characters are used for the other language, so that both English and another language can be encoded in the same scheme.

1.3 UNICODE

The Unicode Standard is the universal character-encoding standard used for representation of text for computer processing. It is fully compatible with the International Standard ISO/IEC 10646-1; 1993, and contains all the same characters and encoding points as ISO/IEC 10646.
Unicode provides a consistent way of encoding multilingual plain text and brings order to a chaotic state of affairs that has made it difficult to exchange text files internationally.

The Unicode Consortium is an organization of major computer corporations, software developers, database vendors, international agencies and various user groups and was formed in 1991. Ministry of Information Technology, Government of India is a full member of Unicode Consortium. Unicode and International Standards Organisation (ISO) joined hands to bring out the international character code standard in 1992, known as Unicode standard Version 1.0. Unicode version 1.0 used somewhat different names for some characters than ISO 10646. In Unicode version 2.0, the names were made the same as in ISO 10646. Version 3.0 was published in February 2000. Unicode is now a 16-bit form of the larger ISO10646, 32-bit standard.

The design of Unicode is based on the simplicity and consistency of ASCII, but goes far beyond ASCII's limited ability to encode only the Latin alphabet. The Unicode Standard provides the capacity to encode all of the characters used for the written languages of the world. It uses a 16-bit encoding that provides code points for more than 65,000 characters. To keep character coding simple and efficient, the Unicode Standard assigns each character a unique 16-bit value, and does not use complex modes or escape codes.

While 65,000 characters are sufficient for encoding most of the many thousands of characters used in major languages of the world, the Unicode standard and ISO 10646 provide an extension mechanism called UTF-16 that allows for encoding as many as a million more characters, without use of escape codes. This is sufficient for all known character encoding requirements, including full coverage of all historic scripts of the world.

The Unicode Standard defines codes for characters used in the major languages written today. Scripts include Latin, Greek, Cyrillic, Armenian, Hebrew, Arabic, Devanagari, Bengali, Gurmukhi, Gujarati, Oriya, Tamil, Telugu, Kannada, Malayalam, Thai, Lao, Georgian, Tibetan, Japanese Kana, the complete set of modern KoreanHangul, and a unified set of Chinese/Japanese/Korean (CJK) ideographs. Many more scripts and characters are to be added shortly, including Ethiopic, Canadian Syllabics, Cherokee, additional rare ideographs, Sinhala, Syriac, Burmese, Khmer, and Braille.

In all, the Unicode Standard, Version 3.0 provides codes for 49,194 characters from the world's alphabets, ideograph sets, and symbol collections. These all fit into the first 64K characters, an area of the code space that is called basic multilingual plane, or BMP for short. Unicode also reserves some code values for private use, which the software and hardware developers can assign internally for their own characters and symbols.

The Unicode standard directly addresses only the encoding and decoding of the text elements, that is, it defines how characters are interpreted. The character identified by a Unicode code value is an abstract entity, such as “MALAYALAM CONSONANT PA “. The mark made on the screen or paper, called a glyph, is the visual representation of the character. Unicode standard does not define glyph images, that is how the character should appear on the screen or paper, which is the responsibility of the hardware or software rendering engine.
The Unicode Standard defines three encoding forms that allow the same data to be transmitted in a byte, word or double word oriented format (i.e. in 8, 16 or 32-bits per code unit). All three encoding forms encode the same common character repertoire and can be efficiently transformed into one another without loss of data. The Unicode Consortium fully endorses the use of any of these encoding forms as a conformant way of implementing the Unicode Standard.

UTF-8 is popular for HTML and similar protocols. UTF-8 is a way of transforming all Unicode characters into a variable length encoding of bytes. It has the advantages that the Unicode characters corresponding to the familiar ASCII set have the same byte values as ASCII, and that Unicode characters transformed into UTF-8 can be used with much existing software without extensive software rewrites.

UTF-16 is popular in many environments that need to balance efficient access to characters with economical use of storage. It is reasonably compact and all the heavily used characters fit into a single 16-bit code unit, while all other characters are accessible via pairs of 16-bit code units.

UTF-32 is popular where memory space is no concern, but fixed width, single code unit access to characters is desired. Each Unicode character is encoded in a single 32-bit code unit when using UTF-32.

All three encoding forms need at most 4 bytes (or 32-bits) of data for each character.

Code spaces U+0900 to U+0D7F is allotted to Indian scripts in Unicode. The allotment of code space for Indian scripts in Unicode is as follows:

<table>
<thead>
<tr>
<th>Language</th>
<th>Code space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devanagari</td>
<td>U+0900 to U+097F</td>
</tr>
<tr>
<td>Bengali</td>
<td>U+0980 to U+09FF</td>
</tr>
<tr>
<td>Gurumukh</td>
<td>U+0A00 to U+0A7F</td>
</tr>
<tr>
<td>Gujarati</td>
<td>U+0A80 to U+0AFF</td>
</tr>
<tr>
<td>Oriya</td>
<td>U+0B00 to U+0B7F</td>
</tr>
<tr>
<td>Tamil</td>
<td>U+0B80 to U+0BFF</td>
</tr>
<tr>
<td>Telugu</td>
<td>U+0C00 to U+0C7F</td>
</tr>
<tr>
<td>Kannada</td>
<td>U+0C80 to U+0CFF</td>
</tr>
<tr>
<td>Malayalam</td>
<td>U+0D00 to U+0D7F</td>
</tr>
</tbody>
</table>

Malayalam is allotted 128 character positions, code space from U+0D00 to U+0D7F. The table showing the Malayalam character codes of Unicode is given in section 4.1.

Unicode has been hailed by many in the computing communities as an ideal solution to the problems of multiplatform internationalization. It is destined to replace ASCII and other single and multibyte character sets currently in existence. Majority of the of the software...
developers the world over have declared conformance to Unicode. They include IBM, Microsoft, Oracle, Sybase, Unisys, Apple, Bell Labs, Compaq, GNU/Linux, Sun, SCO, Hewlett Packard, Netscape, Ericsson and Novell. More and more applications are becoming Unicode compliant. It is expected that Unicode will become the de facto standard in the multilingual world, especially with the spread of Internet.

1.4 National Initiatives

The demand for computer systems capable of providing input/output facilities in Indian languages started in 1970s. Department of Electronics (now Ministry of Information Technology), Government of India took the initiative in 1983 for standardization of Indian language keyboard and character encoding. The code thus arrived at was ISCII-83, an 8-bit code, which complied with ISO 8-bit recommendations. ISCII stands for Indian Script Code for Information Interchange.

The fact that all Indian scripts, with the exception of Perso-Arabic scripts, have evolved from the ancient Brahmi script and they have a common phonetic structure, formed the basis of ISCII. A common platform for all Indian script was envisaged and DOE recommended the 8-bit ISCII code in 1986. ISCII was tried out in various implementations and refined in 1986 and again in 1988.

ISCII is defined in such a way that all Indian languages can use a single character encoding scheme. 8 bits used for ISCII coding can represent a maximum of 256 characters. Out of this, Roman characters and punctuation marks as defined in ASCII is allotted the first 128 character slots. The next 128 character positions are allotted for the Indian languages. The standard provides a unique code for the vowel, consonant ad modifier characters of Indian languages and leaves the screen rendering process to the hardware or software.

The revision of this standard in 1991 made the code chart more compact. ISCII was accepted as the encoding standard for Indian languages in 1991 by Bureau of Indian Standards (BIS) as IS 13194:1991. ISCII standard was reaffirmed by BIS in 1997.

Department of Electronics also defined a keyboard layout called Inscript for all Indian languages in 1986. Inscript keyboard is based on phonetic layout and it matches with ISCII in terms of the characters provided on the keyboard and the codes provided for characters. Bureau of Indian Standards accepted Inscript as National standard along with ISCII.

1.5 Malayalam Keyboard

Malayalam script contains hundreds of characters, including conjuncts. It is widely accepted that as the number of characters become large, it becomes unwieldy and inefficient for typing and printing. After detailed study, discussions and deliberations, characters of Malayalam script were modified in 1968 to make it easier to handle in typewriters. Along with that, Malayalam typewriter keyboard layout was also standardized.
The manufacturers of the typewriter keyboards added their own modifications to this standard. The typewriter keyboard layout was designed for mechanical typewriters and the criteria is slightly different for computer keyboards.
2. Malayalam character set

The first step in evolving a keyboard layout as well as character-encoding scheme is to define the character set in Malayalam. The character set shall contain all the smallest useful elements of text to be encoded. The characters are to be pragmatically chosen as appropriate to express text and allow various text processes.

State Institute of Languages, Thiruvananthapuram has conducted a detailed study and recently has arrived at the basic character elements required for Malayalam. The basic character elements have been finalised on the basis of the following works:

- The discussions and debates by linguistic experts in Malayalam
- The workshop organised by State Institute of Languages and Central Institute of Indian Languages (Mysore) from 17th to 27th of August 1998, with the support of experts in Department of Official Languages, Kerala Sahitya Academy, NCERT, Linguistics Department (University of Kerala) and Sarva Vijnana Kosam.
- Seminar organised by Lexicon Department (University of Kerala), at Senate Hall on 19th and 20th of March 1999.
- Combined discussions of the Committees for Malayalam script standardisation, under the Chairmanship of Sri P.Govinda Pillai and Dr.V.R.Prabhodhachandran Nair.
- Evaluation of the feedbacks received on the recommendations of the Committee.

The decisions evolved after the series of these activities are given in the Style Book (Malayalam Achatiyum Ezhuthum) published by the State Institute of Languages. The Committee accepted the basic Malayalam character set as given in the Style Book.

Thus, Malayalam keyboard shall have the following characters:

**Vowels (13)**

<table>
<thead>
<tr>
<th>്</th>
<th>൏</th>
<th>ൔ</th>
<th>ൌ</th>
<th>൐</th>
<th>൑</th>
<th>൒</th>
<th>ൔ</th>
</tr>
</thead>
<tbody>
<tr>
<td>്</td>
<td>൏</td>
<td>ൔ</td>
<td>ൌ</td>
<td>൐</td>
<td>൑</td>
<td>൒</td>
<td>ൔ</td>
</tr>
</tbody>
</table>

**Consonants (36)**

<table>
<thead>
<tr>
<th>്</th>
<th>൏</th>
<th>ൔ</th>
<th>ൌ</th>
<th>൐</th>
<th>൑</th>
<th>൒</th>
<th>ൔ</th>
</tr>
</thead>
<tbody>
<tr>
<td>്</td>
<td>൏</td>
<td>ൔ</td>
<td>ൌ</td>
<td>൐</td>
<td>൑</td>
<td>൒</td>
<td>ൔ</td>
</tr>
</tbody>
</table>

State Institute of Languages, Thiruvananthapuram has conducted a detailed study and recently has arrived at the basic character elements required for Malayalam. The basic character elements have been finalised on the basis of the following works:

- The discussions and debates by linguistic experts in Malayalam
- The workshop organised by State Institute of Languages and Central Institute of Indian Languages (Mysore) from 17th to 27th of August 1998, with the support of experts in Department of Official Languages, Kerala Sahitya Academy, NCERT, Linguistics Department (University of Kerala) and Sarva Vijnana Kosam.
- Seminar organised by Lexicon Department (University of Kerala), at Senate Hall on 19th and 20th of March 1999.
- Combined discussions of the Committees for Malayalam script standardisation, under the Chairmanship of Sri P.Govinda Pillai and Dr.V.R.Prabhodhachandran Nair.
- Evaluation of the feedbacks received on the recommendations of the Committee.

The decisions evolved after the series of these activities are given in the Style Book (Malayalam Achatiyum Ezhuthum) published by the State Institute of Languages. The Committee accepted the basic Malayalam character set as given in the Style Book.

Thus, Malayalam keyboard shall have the following characters:

**Vowels (13)**

<table>
<thead>
<tr>
<th>്</th>
<th>൏</th>
<th>ൔ</th>
<th>ൌ</th>
<th>൐</th>
<th>൑</th>
<th>൒</th>
<th>ൔ</th>
</tr>
</thead>
<tbody>
<tr>
<td>്</td>
<td>൏</td>
<td>ൔ</td>
<td>ൌ</td>
<td>൐</td>
<td>൑</td>
<td>൒</td>
<td>ൔ</td>
</tr>
</tbody>
</table>

**Consonants (36)**

<table>
<thead>
<tr>
<th>്</th>
<th>൏</th>
<th>ൔ</th>
<th>ൌ</th>
<th>൐</th>
<th>൑</th>
<th>൒</th>
<th>ൔ</th>
</tr>
</thead>
<tbody>
<tr>
<td>്</td>
<td>൏</td>
<td>ൔ</td>
<td>ൌ</td>
<td>൐</td>
<td>൑</td>
<td>൒</td>
<td>ൔ</td>
</tr>
</tbody>
</table>
Anuswaram, Visargam, Chandrakkala (3)

\[ \text{Anuswaram, Visargam, Chandrakkala (3)} \]

\[ \text{chillu (5)} \]

<table>
<thead>
<tr>
<th>( \text{chillu} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{chillu} )</td>
</tr>
<tr>
<td>( \text{chillu} )</td>
</tr>
<tr>
<td>( \text{chillu} )</td>
</tr>
<tr>
<td>( \text{chillu} )</td>
</tr>
<tr>
<td>( \text{chillu} )</td>
</tr>
<tr>
<td>( \text{chillu} )</td>
</tr>
</tbody>
</table>

Consonant signs (4)

\[ \text{Consonant signs (4)} \]

Thus there are 73 basic characters required in Malayalam keyboard.

The character set consist of the following character types:

- vowels
- consonants
- Anuswaram, Visargam and Chandrakkala
- chillu
- consonant signs
- vowel signs

Consonants

The consonant letter represent a single consonantal sound, but also an inherent vowel. In the presence of a dependent vowel, however, the inherent vowel associated with the consonant letter is overridden with the dependent vowel.

Vowels

The independent vowel letters stand on their own. The writing system treats independent vowels as (Consonant and Vowel) syllables, in which the consonant is null. The independent vowels are used to write syllables which start with a vowel.

Vowel signs

The dependent vowels do not stand on their own, but they are depicted in combination with a consonant or consonant cluster. Explicit appearance of a dependent vowel in a syllable overrides the inherent vowel of a single consonant character. The positioning of the dependent vowel during rendering may be to the left, to the right, or both to the left and right of the consonant or conjunct, as shown below, depending on the vowel sign being attached.
Anuswaram ( الأوروبي ) is the pure consonant form of the character. Anuswaram cannot stand independently, but is always attached to a vowel, consonant, or a conjunct.

Ex: ആ ഇ ഓ ക

Visargam ( ഇ ) is the pure consonant form of the character. Visargam cannot stand independently, but is always attached to a vowel, consonant, or conjunct.

Ex: ആ ഇ ഓ ക

Chandrakkala ( വ ) is interpreted depending on its position in a word. If chandrakkala is the last character in a word, it is treated as the half-vowel of ഓ.

Ex: ആ ഓ ഓ ക

If chandrakkala is not the last character of a word (it is followed by a consonant or conjunct), it is treated as the vowel omission sign. This serves to cancel the inherent vowel of the consonant to which it is applied. It functions as a combining character (to form a conjunct).

Ex: ആ ഓ ഓ ക

<table>
<thead>
<tr>
<th>Vowel Sign</th>
<th>Attached to Left/Right of the Consonant/Conjunct</th>
<th>Example (Vowel Sign Attached to ഓ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ഓ</td>
<td>right</td>
<td>ഓം</td>
</tr>
<tr>
<td>ഓ</td>
<td>right</td>
<td>ഈ</td>
</tr>
<tr>
<td>ഓ</td>
<td>right</td>
<td>ഓങ</td>
</tr>
<tr>
<td>ഓ</td>
<td>right</td>
<td>ഓങ</td>
</tr>
<tr>
<td>ഓ</td>
<td>left</td>
<td>ഓങ</td>
</tr>
<tr>
<td>ഓ</td>
<td>left</td>
<td>ഓങ</td>
</tr>
<tr>
<td>ഓ</td>
<td>left and right</td>
<td>ഓങ</td>
</tr>
<tr>
<td>ഓ</td>
<td>left and right</td>
<td>ഓങ</td>
</tr>
<tr>
<td>ഓ</td>
<td>right</td>
<td>ഓങ</td>
</tr>
</tbody>
</table>
3. Malayalam keyboard layout

As shown above there are 73 basic characters that should be provided in Malayalam keyboard. This is after leaving the still larger number of conjuncts and Malayalam numerals. Since the numerals and the punctuations are required in Malayalam also, it is clear that in order to accommodate all the Malayalam characters in the keyboard, there has to be some compromises.

Any keyboard layout which cannot be mapped to the standard QWERTY keyboard will not be commercially sustainable. The QWERTY keyboard has provision for 26 keys for English letters (same keys used for lower case (normal position) and upper case (shift position)), 10 keys for numerals and 11 keys for punctuation marks. Thus, altogether there are 47 keys, excluding the control keys.

The optimum Malayalam keyboard layout should be such that

- It should be capable of accommodating all Malayalam characters being used
- It should be usable in all computing platforms available
- The typing efficiency should be high

The two keyboard standards available in Malayalam, which are not proprietary, are the Inscript layout and Typewriter layout.

The typewriter keyboard has been designed for the mechanical medium and is not very efficient for computer keyboards. The keyboard design does not take advantages of the flexibility of computers in rendering the conjunct letters, even if the conjuncts require more than one key-press. Also, the typing sequence of words in the typewriter keyboard is based on the actual appearance of the characters in the word. This is not very efficient, if any processing on the text, like sorting, searching, spell-check, grammar-check etc are to be done on the text.

The Inscript overlay can be used in any QWERTY keyboard. Malayalam script legends are shown on the right hand side of the key and English legends, on the left hand side. English or Malayalam overlay can be selected through the Scroll Lock key (each key press toggles the overlay) or by software selection. The keyboard is based on phonetic layout. Consonants and vowels are given keystrokes and the consonant-vowel combinations are formed by typing the consonant key followed by the vowel key. Similarly, conjuncts are formed by typing the constituent consonants with a chandrakkala typed in between.

Inscript overlay has taken into account the logical structure of script alphabet, derived from the phonetic properties. The vowels have been laid out on the left hand side of the keyboard and the consonants on the right hand side.
The vowels are allotted the shift positions of the corresponding vowel signs. This is due to the fact that the frequency of occurrence of vowel signs is more than the corresponding vowels in Malayalam. The vowel omission sign, chandrakkala, is given in the un-shifted position of A, since A does not have a corresponding vowel sign. Alternate hand action occurs while typing conjuncts, since chandrakkala, is in the left hand side and most of the consonants are on the right hand side of the overlay (chandrakkala is used as the link to form conjuncts and is to be typed in between the constituent consonants of the conjunct).

On the consonant side, all the primary characters of the five vargs (?,?,?,?,?) are included in the home row. The aspirated consonants are kept in the shift positions of the corresponding unaspirated counterparts.

In the Inscript scheme, vowel sign is to be typed after the consonant/conjunct to override the inherent A vowel of the consonant/conjunct. In the case of hand-written text, the vowel signs are to be put before the consonants in some cases (for example, in the case of ?, ? , ? signs), and in some other cases part of the vowel signs are to be put before the consonant and part after the consonant. (For example, in the case of ?, ?). For the processing of the document like sorting, searching, spell check etc. it is better that the vowel signs are typed after the consonants, since these processes are based on the construct of the words/syllables. The syllables are, in turn, built around consonants/conjuncts and modified by vowel signs.

Thus, the options available for Malayalam keyboard standard are:

- Accept Inscript layout for Malayalam keyboard, recommended by Department of Electronics, Government of India (now Ministry of Information Technology) and accepted by Bureau of Indian Standards.

- Accept Inscript layout as the base and then work on the modifications required in this specification, so as to make it a more acceptable one.

- Accept the typewriter keyboard layout

- Go for a fresh standard, based on the frequency study of occurrence of characters in Malayalam documents and the study of the existing keyboard schemes.

- Accept more than one keyboard layout standard.

The Committee decided to go for the second option, that is, accept the Inscript keyboard as the standard and address the shortcomings of the layout, for the following reasons:

- It is reported that more than 90% of the word processing packages used in Malayalam uses Inscript keyboard. So, the acceptability of a totally new standard may not be very high and may take time, even if the typing efficiency is more with the new layout. Also, the numerous documents already generated using Inscript will have to be converted to the new standard.
• The frequency study conducted by the Kerala State Institute of Languages has found that the layout of Inscript keyboard has been done scientifically. The keys of most frequent use are allotted strong key positions and those of less frequent use need two key presses (shift + character key).

• Bureau of Indian Standards and Government of India has accepted ISCII and Inscript as the keyboard standard for all Indian languages including Malayalam, as per IS 13194:1991.

• Ministry of Information Technology has initiated various projects for popularisation of local language contents in Computer and Internet. Most of these initiatives will be based on ISCII and Inscript standards.

• Software developers have developed and are developing various products based on Inscript/ISCII standards in Malayalam. Malayalam keyboards with Inscript layout has come into the market.

• As mentioned earlier, the typewriter keyboard was designed for mechanical medium and is not particularly suitable for computer operations, which has flexibility to modify the visual representation based on the keystroke combinations. The typing sequence in Typewriter keyboard is not very efficient as far as the processing of text is concerned.

• It is highly advisable that we have Malayalam keyboards available in the market from various manufacturers. The commercial viability of this increases substantially, if we have a unique keyboard layout. Going for multiple keyboard layouts will create confusion for those who have to type in various computers/ various application packages using different layouts.

3.1 Modifications required in Inscript keyboard layout

There are some shortcomings in the Inscript layout, and the Committee recommends the following modifications in Inscript keyboard layout for making it a standard keyboard layout.

1. Inclusion of chillu

In the Inscript layout, chillu characters are not provided on the keyboard, in spite of the fact that they are extensively used in Malayalam. Users have to type three keys to get chillu. It is recommended that five chillu characters be given place on the keyboard. The keys suggested, based on the frequency of occurrence of the chillu and the availability of keys, are as shown below:

\[ \text{̄, } \text{̄, } \text{̄, } \text{̄, } \text{̄} \]

The shifted position of \( \ddot{v} \) (accent), \( \ddot{.} \) (decimal point), and \( \ddot{\text{\textbackslash}} \) (backslash).
2. Inclusion of consonant signs.

As mentioned earlier, the character set of Malayalam contains four consonant signs, of which the frequency of occurrence of three (ŋ l ɭ) is high. So, it is recommended that these consonant signs be provided keys in the keyboard. Out of these four, ɭ is already provided key in the Inscript keyboard (Shift position of 3). Based on the frequency of occurrence of these consonant signs and the availability of keys, the suggested keys for the consonant signs are:

ŋ shift position of 4
ɭ shift position of 6
ɭ shift position of 5

3. It should be possible to type the vowel signs and consonant signs like

० १ २ ३ ४ ५ ६ ७ ८ ९ ə ɛ ɪ ɔ ɚ without being attached to consonants and they should also be displayed as such. The word processing software may indicate (on the screen) that there is a spelling mistake in those cases where these signs are typed without attaching to a consonant/ conjuncts.

4. NUKTA is not used in Malayalam and can be deleted from the keyboard overlay.

5. £ has been given a key in the Inscript layout. Since £ is not a basic character, it may be deleted from the keyboard overlay.

The layout of the recommended Malayalam keyboard layout is given in figure 1.

3.2 Typing sequence

There has to be a unique typing sequence for each of the Malayalam word. For typing a word, it has to be broken into a sequence of separate keystrokes.

a) vowels and consonants

The vowels and consonants are to be typed in the sequence of pronunciation

<table>
<thead>
<tr>
<th>word</th>
<th>keystroke sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ഭക്തി</td>
<td>ഭ ് ൈ ൊ</td>
</tr>
<tr>
<td>കലക്ഷണം</td>
<td>ക ല</td>
</tr>
<tr>
<td>ശ്രീ</td>
<td>ശ ര</td>
</tr>
<tr>
<td>പാ</td>
<td>പ ഓ</td>
</tr>
<tr>
<td>പി</td>
<td>പ ര</td>
</tr>
<tr>
<td>ഗോ</td>
<td>ഗു ഓ</td>
</tr>
<tr>
<td>ശുബ്ദം</td>
<td>ശ വു</td>
</tr>
</tbody>
</table>
b) vowel signs attached to consonants

The vowel sign to be attached to a consonant is to be typed immediately after the consonant/conjunct.

<table>
<thead>
<tr>
<th>Consonant</th>
<th>Vowel Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>മ യ മ</td>
<td>എ യ യ</td>
</tr>
<tr>
<td>ത ത ത</td>
<td>ല ല ല</td>
</tr>
<tr>
<td>ന ന ന</td>
<td>ത ത ത</td>
</tr>
<tr>
<td>ക ക ക</td>
<td>പ പ പ</td>
</tr>
<tr>
<td>കം കം കം</td>
<td>ത ത ത</td>
</tr>
<tr>
<td>തം തം തം</td>
<td>ത ത ത</td>
</tr>
<tr>
<td>ര ര ര</td>
<td>സ സ സ</td>
</tr>
<tr>
<td>തര തര തര</td>
<td>ത ത ത</td>
</tr>
<tr>
<td>നർ നർ നർ</td>
<td>ത ത ത</td>
</tr>
<tr>
<td>സാംതി</td>
<td>ത ത ത</td>
</tr>
</tbody>
</table>

(c) anuswaram

Anuswaram to be attached to a consonant or vowel is to be typed immediately after the consonant/ vowel

<table>
<thead>
<tr>
<th>Consonant</th>
<th>Anuswaram</th>
</tr>
</thead>
<tbody>
<tr>
<td>മ യ മ</td>
<td>യ യ</td>
</tr>
<tr>
<td>ത ത ത</td>
<td>യ യ</td>
</tr>
<tr>
<td>ന ന ന</td>
<td>യ യ</td>
</tr>
<tr>
<td>ക ക ക</td>
<td>യ യ</td>
</tr>
<tr>
<td>കം കം കം</td>
<td>യ യ</td>
</tr>
<tr>
<td>തം തം തം</td>
<td>യ യ</td>
</tr>
<tr>
<td>ര ര ര</td>
<td>യ യ</td>
</tr>
<tr>
<td>തര തര തര</td>
<td>യ യ</td>
</tr>
<tr>
<td>നർ നർ നർ</td>
<td>യ യ</td>
</tr>
<tr>
<td>ചാലി</td>
<td>യ യ</td>
</tr>
</tbody>
</table>

If vowel sign is present, anuswaram will follow it (as is the order in which the syllable is pronounced)

<table>
<thead>
<tr>
<th>Consonant</th>
<th>Anuswaram</th>
</tr>
</thead>
<tbody>
<tr>
<td>മ യ മ</td>
<td>യ യ</td>
</tr>
<tr>
<td>ത ത ത</td>
<td>യ യ</td>
</tr>
<tr>
<td>ന ന ന</td>
<td>യ യ</td>
</tr>
<tr>
<td>ക ക ക</td>
<td>യ യ</td>
</tr>
<tr>
<td>കം കം കം</td>
<td>യ യ</td>
</tr>
<tr>
<td>തം തം തം</td>
<td>യ യ</td>
</tr>
<tr>
<td>ര ര ര</td>
<td>യ യ</td>
</tr>
<tr>
<td>തര തര തര</td>
<td>യ യ</td>
</tr>
<tr>
<td>നർ നർ നർ</td>
<td>യ യ</td>
</tr>
<tr>
<td>ചാലി</td>
<td>യ യ</td>
</tr>
</tbody>
</table>

(d) visargam

Visargam is used to indicate an aspiration sound (h) and is to be typed immediately after the consonant/ vowel to which it is attached. If vowel sign is present, anuswaram will follow it.

<table>
<thead>
<tr>
<th>Consonant</th>
<th>Visargam</th>
</tr>
</thead>
<tbody>
<tr>
<td>മ യ മ</td>
<td>ഹ</td>
</tr>
<tr>
<td>ത ത ത</td>
<td>ഹ</td>
</tr>
<tr>
<td>ന ന ന</td>
<td>ഹ</td>
</tr>
<tr>
<td>ക ക ക</td>
<td>ഹ</td>
</tr>
<tr>
<td>കം കം കം</td>
<td>ഹ</td>
</tr>
<tr>
<td>തം തം തം</td>
<td>ഹ</td>
</tr>
<tr>
<td>ര ര ര</td>
<td>ഹ</td>
</tr>
<tr>
<td>തര തര തര</td>
<td>ഹ</td>
</tr>
<tr>
<td>നർ നർ നർ</td>
<td>ഹ</td>
</tr>
</tbody>
</table>

(e) chandrakkala

Chandrakkala is typed after a consonant/ conjunct to indicate missing of the inherent അ vowel in the consonant.

<table>
<thead>
<tr>
<th>Consonant</th>
<th>Chandrakkala</th>
</tr>
</thead>
<tbody>
<tr>
<td>ക ക ക</td>
<td>അ് അ്</td>
</tr>
<tr>
<td>പ പ പ</td>
<td>അ് അ്</td>
</tr>
<tr>
<td>ക ക ക</td>
<td>അ് അ്</td>
</tr>
</tbody>
</table>
Chandrakkala cannot be attached to vowels, vowel signs, anuswaram and visargam. 

\[\text{ granularity } \text{ granularity } \text{ granularity } \text{ granularity} \] are not valid syllables.

\( f \) Consonant sign \( \zeta \)

Consonant sign \( \zeta \) is attached to consonants/conjuncts as follows. The vowel signs, if present, are to be typed after \( \zeta \)

\begin{align*}
\text{ kvw } & \text{ kwv } I x v \\
\text{ hyXymkw } & \text{ hyXymkw } h y x m k w \\
\text{ Atymyw } & \text{ Atymyw } A t y m w
\end{align*}

Typing \( \omega \) after consonant and chandrakkala (which is the sequence used in Inscript keyboard overlay to get \( \zeta \)) will result in the consonants being separated by chandrakkala.

\begin{align*}
\text{ kvw } & \text{ kwv } I x v \\
\text{ hyXymkw } & \text{ hyXymkw } h y x m k w \\
\text{ Atymyw } & \text{ Atymyw } A t y m w
\end{align*}

\( g \) Consonant sign \( \iota \)

Consonant sign \( \iota \) is attached to consonants/conjuncts as follows. Vowel sign, if present, is to be typed after \( \iota \)

\begin{align*}
\text{ kzw } & \text{ kzw } I x w \\
\text{ izmkw } & \text{ izmkw } i z m k w
\end{align*}

Typing \( \omega \) after consonant and chandrakkala (which is the sequence used in Inscript keyboard overlay to get \( \iota \)) will result in the consonants being separated by chandrakkala.

\begin{align*}
\text{ kzw } & \text{ kzw } I x w \\
\text{ izmkw } & \text{ izmkw } i z m k w
\end{align*}

\( h \) Consonant sign \( \zeta \)

Consonant sign \( \zeta \) is attached to consonants/conjuncts as follows. Vowel sign, if present, is to be typed after \( \zeta \)

\begin{align*}
\text{ kzw } & \text{ kzw } I x w \\
\text{ izmkw } & \text{ izmkw } i z m k w
\end{align*}

The attachment of \( \zeta \) to consonant/ conjuncts in most of the cases results in appending \( \circ \) sound to the pure consonant/ conjunct, as shown above. But in some cases, \( \zeta \) sign denotes the attachment of \( \circ \) sound to the pure consonant/conjunct.

Typing \( \omega \) after consonant and chandrakkala (which is the sequence used in Inscript keyboard overlay to get \( \zeta \)) will result in the consonants being separated by chandrakkala.
Typing ɔ after consonant and chandrakkala will result in the consonants being separated by chandrakkala.

<table>
<thead>
<tr>
<th>Malayalam</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>ു</td>
<td>õ o 0 m `</td>
</tr>
<tr>
<td>ു</td>
<td>õ o 0 m `</td>
</tr>
</tbody>
</table>

i) Consonant sign

Consonant sign is attached to consonants/ conjuncts as shown. Vowel sign, if present, is to be typed after

<table>
<thead>
<tr>
<th>Malayalam</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>തുംപ</td>
<td>õ o 0 m `</td>
</tr>
<tr>
<td>തുംപ</td>
<td>õ o 0 m `</td>
</tr>
</tbody>
</table>

The attachment of  to consonant/ conjuncts in most of the cases results in appending ̃ sound to the pure consonant/ conjunct, as shown above. But in some cases, ̃ sign denotes the attachment of ̃ sound to the pure consonant/conjunct.

Typing  in after consonant and chandrakkala (which is the sequence used in Inscript keyboard overlay to get ) or typing  in after consonant and chandrakkala will result in the consonants being separated by chandrakkala.

<table>
<thead>
<tr>
<th>Malayalam</th>
<th>Transliteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>തുംപ്പ</td>
<td>õ o 0 m `</td>
</tr>
<tr>
<td>തുംപ്പ</td>
<td>õ o 0 m `</td>
</tr>
</tbody>
</table>

3.3 Con juncts

Malayalam has a large number of consonant conjunct forms, which serve as orthographic abbreviations of two or more adjacent letter forms. A consonant cluster is depicted with a conjunct glyph, if such a glyph is available in the current glyph. In the absence of a conjunct glyph, the conjunct is depicted with the nominal consonant forms with chandrakkala in between.

The split up (and the key sequence) of the conjuncts need to be unambiguously defined for the following reasons:

1. The key combinations and the order in which they are to be typed to get a conjunct letter is a basic need for the user.
2. The conjunct letters will be stored in computer as the constituent characters. So, for the conjunct letter to be read from the memory and displayed (or processed) in the same way as was typed in, the constituent character elements and the order of their typing in need to be known.

3. The sort order also is dependent on the split up of the conjuncts (and hence the order of typing of the constituent characters).

In most of the cases the split up and hence the typing sequence is straight forward, but in a few cases, there has not been unanimity among the Malayalam language experts on the split-up.

The conjuncts commonly used in Malayalam and their key sequence are shown below:

\[\begin{array}{c|c}
\text{conjunct} & \text{key sequence} \\
\hline
\text{क} & क
\end{array}\]

\[\begin{array}{c|c}
\text{conjunct} & \text{key sequence} \\
\hline
\text{ल} & ल
\end{array}\]

\[\begin{array}{c|c}
\text{conjunct} & \text{key sequence} \\
\hline
\text{म} & म
\end{array}\]

\[\begin{array}{c|c}
\text{conjunct} & \text{key sequence} \\
\hline
\text{न} & न
\end{array}\]

\[\begin{array}{c|c}
\text{conjunct} & \text{key sequence} \\
\hline
\text{व} & व
\end{array}\]

\[\begin{array}{c|c}
\text{क} & क
\end{array}\]

\[\begin{array}{c|c}
\text{व} & व
\end{array}\]

\[\begin{array}{c|c}
\text{क} & क
\end{array}\]

\[\begin{array}{c|c}
\text{व} & व
\end{array}\]

\[\begin{array}{c|c}
\text{क} & क
\end{array}\]

\[\begin{array}{c|c}
\text{व} & व
\end{array}\]

\[\begin{array}{c|c}
\text{क} & क
\end{array}\]

\[\begin{array}{c|c}
\text{व} & व
\end{array}\]

\[\begin{array}{c|c}
\text{क} & क
\end{array}\]

\[\begin{array}{c|c}
\text{व} & व
\end{array}\]

\[\begin{array}{c|c}
\text{क} & क
\end{array}\]

\[\begin{array}{c|c}
\text{व} & व
\end{array}\]

\[\begin{array}{c|c}
\text{क} & क
\end{array}\]

\[\begin{array}{c|c}
\text{व} & व
\end{array}\]
The following points may be noted:

1. The split up of $\frac{1}{2}$ is given as $\text{av } [\cdot v ]$, whereas the split up used generally has been $\text{av } [\cdot v ]$. This is because $\frac{1}{2}$ is nowhere used to represent $\text{av } [\cdot v ]$ in Malayalam text, but to represent $\text{av } [\cdot v ]$. The typing sequence involving $\frac{1}{2}$ followed by chandrakkala and then a consonant will result in the consonants separated by chandrakkala.

2. The key combination required to get $\text{sl } ^{3}d$ shall be $\text{sl } ^{3}d$ followed by $\text{chandrakkala } (\text{av } [\cdot v ])$ followed by $\text{d}$.

3. $\text{sl } ^{3}d$ and $\text{sl } ^{3}d$ can be split in two ways as below.
For uniqueness in the split up, the first of these are accepted, since the user will find it comfortable to map these split up with the visual representation of the conjuncts.

4. Conjuncts can also be shown with chandrakkala in between the constituent consonants, without formation of the conjunct. This can be achieved by typing chandrakkala twice.

5. Wherever there is no glyph available for the conjuncts, the constituent consonants can be displayed with chandrakkala in between.

6. Typing sequence of some of the conjuncts with more than two consonants:

3.4 Intra glyph positioning of vowel signs

When the conjunct letter is presented as a single physical entity (the constituent consonants joined together), the attachment of vowel signs to the conjunct letter is similar to attaching vowel signs to consonants. But when the conjunct letter does not have a glyph of its own, it is presented as consonants separated by chandrakkala. In such cases, confusion may arise on the position of the vowel sign with reference to the constituent consonants of the conjunct letter. The simple rule in this regard is that the vowel sign is to be attached with the second consonant of the conjunct, which is in agreement with the pronunciation of the syllable, Some of the examples are shown below, which shows the positioning of the vowel signs as well as the key sequence for the words formed.
<table>
<thead>
<tr>
<th>Malayalam</th>
<th>Roman</th>
</tr>
</thead>
<tbody>
<tr>
<td>ഗംഗാദേവി</td>
<td>gngaadevi</td>
</tr>
<tr>
<td>സിന്ധൂദാനി</td>
<td>sindhoonadi</td>
</tr>
<tr>
<td>പ്രിയാമിനി</td>
<td>priyaminii</td>
</tr>
<tr>
<td>അവലോട്ടിയിൽ</td>
<td>avalottiel</td>
</tr>
<tr>
<td>ചിലയാഴ്ചിൽ</td>
<td>chilaazhiil</td>
</tr>
</tbody>
</table>
4. Malayalam character encoding

The purpose of encoding is to represent the text elements in a language in a unique way so that applications could implement a variety of text processes in the desired language. The different operations performed on text in a computer, including input, rendering (display), searching and sorting, have different preferences for the manner in which the text is encoded. Choosing one encoding scheme may make one operation more efficient, but it may make other operations less efficient. The only possible solution is to have a compromise that can adequately satisfy all text processing needs, even though this may be result in lesser efficiency.

The character encoding scheme for Malayalam should be such that:

- It should accommodate all Malayalam characters being used
- Malayalam characters should be recognized and represented uniformly in any system
- It should be usable in all widely used platforms, without any patches

There are two encoding standards for Malayalam in use today - Unicode and ISCII. Unicode is the standard developed by Unicode Consortium in consultation with ISO and caters for all languages in the world. ISCII is the standard developed by Department of Electronics, Government of India and accepted by BIS. ISCII is an 8-bit bilingual coding standard, in which 128 code spaces are available for English and the remaining 128 code spaces being multiplexed by the Indian languages.

The options available for a standard character encoding in Malayalam are:

- Accept ISCII with the necessary modifications
- Accept Unicode with necessary modifications
- Develop a new encoding standard

The Committee recommends Unicode to be the standard for character encoding in Malayalam. The shortcomings in the allotment of character codes for Malayalam can be taken up with Unicode consortium through Ministry of Information Technology, Government of India, which is a full member of the consortium. This recommendation is based on the following reasons:

- It is predicted that majority of the content available on the web and other electronic media is going to be non-English, within a short span of time. This means that content available on the web will be in various languages and it is necessary to identify, read and process correctly the content in a particular language, including Malayalam. No coding scheme available today anywhere in the world has the capability to do so, except Unicode. It is the only standard which caters for the multilingual world. It has provided for unique codes for all the characters of all the major languages used today. Unicode is destined to replace ASCII and other single and multibyte character sets currently in existence.
• Most of the major software developers the world over have committed support to Unicode. They include IBM, Microsoft, Oracle, Sybase, Unisys, Apple, Bell Labs, Compaq, GNU/Linux, Sun, SCO, Hewlett Packard, Netscape, Ericsson and Novell. More and more operating systems, Database Management Systems (DBMS), web and other applications are becoming Unicode compliant. It is expected that Unicode will become the de facto standard in the multilingual world, especially with the spread of Internet.

However, not many applications are available in Malayalam, which conforms to Unicode. Also, Unicode requires 16 bits per character against 8 bits of ISCII. Till the time by which the storage space and the bandwidth ceases to be a constraint, this is a factor to be considered. So, as an interim standard, it is recommended that ISCII be accepted for character encoding in Malayalam, with the necessary modifications incorporated. It is suggested that the situation be reviewed after one year, to take a decision on the changeover to Unicode standard. CDAC has assured the Committee that their Malayalam software products will be compatible with Unicode within a year.

4.1 Modifications required to be carried out in the Unicode representation of Malayalam characters

1. The following chillu characters are to be included in the Malayalam character set:

<table>
<thead>
<tr>
<th>Suggested code position</th>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D58</td>
<td>���</td>
<td>Malayalam Letter L</td>
</tr>
<tr>
<td>0D59</td>
<td>���</td>
<td>Malayalam Letter LL</td>
</tr>
<tr>
<td>0D5A</td>
<td>���</td>
<td>Malayalam Letter RR</td>
</tr>
<tr>
<td>0D5B</td>
<td>���</td>
<td>Malayalam Letter NN</td>
</tr>
<tr>
<td>0D5C</td>
<td>���</td>
<td>Malayalam Letter N</td>
</tr>
</tbody>
</table>

2. The following consonant signs are to be included in the Malayalam character set:

<table>
<thead>
<tr>
<th>Suggested code position</th>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D3A</td>
<td>ළ</td>
<td>Malayalam Consonant Sign RRA</td>
</tr>
<tr>
<td>0D3B</td>
<td>ළ</td>
<td>Malayalam Consonant Sign YA</td>
</tr>
<tr>
<td>0D3C</td>
<td>ළ</td>
<td>Malayalam Consonant Sign VA</td>
</tr>
<tr>
<td>0D3D</td>
<td>ළ</td>
<td>Malayalam Consonant Sign LLA</td>
</tr>
</tbody>
</table>

3. The following characters which are now provided in Unicode may be deleted, since these are no more used in Malayalam:
Malayalam keyboard layout and character encoding

**Remarks**

- Malayalam Letter Vocalic L (0D0C)
- Malayalam Letter Vocalic LL (0D61)
- Malayalam Letter Vocalic RR (0D60)
- Malayalam Vowel Sign AU (0D4C) Malayalam Vowel Sign AU is ू (0D57), which has been wrongly named as Malayalam AU Length Mark.

- Malayalam Digit ZERO (0D66)
- Malayalam Digit ONE (0D67)
- Malayalam Digit TWO (0D68)
- Malayalam Digit THREE (0D69)
- Malayalam Digit FOUR (0D6A)
- Malayalam Digit FIVE (0D6B)
- Malayalam Digit SIX (0D6C)
- Malayalam Digit SEVEN (0D6D)
- Malayalam Digit EIGHT (0D6E)
- Malayalam Digit NINE (0D6F)

The description of some of the characters need to be corrected as mentioned in detail below.

The Unicode code table for Malayalam is given below. The suggested modifications (to be taken up with Unicode) are shown in the table.

<table>
<thead>
<tr>
<th>Code</th>
<th>Character</th>
<th>Description</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D02</td>
<td>ॐ</td>
<td>Malayalam sign Anuswara</td>
<td></td>
</tr>
<tr>
<td>0D03</td>
<td>ॐ</td>
<td>Malayalam sign Visarga</td>
<td></td>
</tr>
</tbody>
</table>

**Various signs**

<table>
<thead>
<tr>
<th>Code</th>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D05</td>
<td>ᠠ</td>
<td>Malayalam letter A</td>
</tr>
<tr>
<td>0D06</td>
<td>trì</td>
<td>Malayalam letter AA</td>
</tr>
<tr>
<td>0D07</td>
<td>_VOICE</td>
<td>Malayalam letter I</td>
</tr>
<tr>
<td>0D08</td>
<td>ˠI</td>
<td>Malayalam letter II</td>
</tr>
<tr>
<td>0D09</td>
<td>sI</td>
<td>Malayalam letter U</td>
</tr>
<tr>
<td>0D0A</td>
<td>.splice</td>
<td>Malayalam letter UU</td>
</tr>
<tr>
<td>0D0B</td>
<td>gł</td>
<td>Malayalam letter vocalic R</td>
</tr>
<tr>
<td>0D0C</td>
<td>眙</td>
<td>Malayalam letter vocalic L</td>
</tr>
<tr>
<td>0D0E</td>
<td>.Emit</td>
<td>Malayalam letter E</td>
</tr>
<tr>
<td>0D0F</td>
<td>throat.</td>
<td>Malayalam letter EE</td>
</tr>
<tr>
<td>0D10</td>
<td>ខ</td>
<td>Malayalam letter AI</td>
</tr>
<tr>
<td>0D12</td>
<td>���</td>
<td>Malayalam letter O</td>
</tr>
<tr>
<td>0D13</td>
<td>���</td>
<td>Malayalam letter OO</td>
</tr>
<tr>
<td>0D14</td>
<td>���</td>
<td>Malayalam letter AU</td>
</tr>
</tbody>
</table>
### Malayalam keyboard layout and character encoding

#### Consonants

<table>
<thead>
<tr>
<th>Code</th>
<th>Malayalam letter</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D15</td>
<td>ஞ (KA)</td>
<td>Malayalam letter KA</td>
</tr>
<tr>
<td>0D16</td>
<td>அ (KHA)</td>
<td>Malayalam letter KHA</td>
</tr>
<tr>
<td>0D17</td>
<td>ஆ (GA)</td>
<td>Malayalam letter GA</td>
</tr>
<tr>
<td>0D18</td>
<td>இ (GHA)</td>
<td>Malayalam letter GHA</td>
</tr>
<tr>
<td>0D19</td>
<td>ஈ (NGA)</td>
<td>Malayalam letter NGA</td>
</tr>
<tr>
<td>0D1A</td>
<td>உ (CHA)</td>
<td>Malayalam letter CHA</td>
</tr>
<tr>
<td>0D1B</td>
<td>ஊ (CHHA)</td>
<td>Malayalam letter CHHA</td>
</tr>
<tr>
<td>0D1C</td>
<td>஋ (JA)</td>
<td>Malayalam letter JA</td>
</tr>
<tr>
<td>0D1D</td>
<td>஌ (JHA)</td>
<td>Malayalam letter JHA</td>
</tr>
<tr>
<td>0D1E</td>
<td>஍ (NHA)</td>
<td>Malayalam letter NHA</td>
</tr>
<tr>
<td>0D1F</td>
<td>எ (TA)</td>
<td>Malayalam letter TA</td>
</tr>
<tr>
<td>0D20</td>
<td>ஐ (TTA)</td>
<td>Malayalam letter TTA</td>
</tr>
<tr>
<td>0D21</td>
<td>ஒ (DHA)</td>
<td>Malayalam letter (hard) DA</td>
</tr>
<tr>
<td>0D22</td>
<td>ஓ (DDHA)</td>
<td>Malayalam letter (hard) DDA</td>
</tr>
<tr>
<td>0D23</td>
<td>ஔ (NA)</td>
<td>Malayalam letter (hard) NA</td>
</tr>
<tr>
<td>0D24</td>
<td>க (THA)</td>
<td>Malayalam letter THA</td>
</tr>
<tr>
<td>0D25</td>
<td>க (TTHA)</td>
<td>Malayalam letter TTHA</td>
</tr>
<tr>
<td>0D26</td>
<td>஖ (DA)</td>
<td>Malayalam letter (soft) DA</td>
</tr>
<tr>
<td>0D27</td>
<td>ச (DHA)</td>
<td>Malayalam letter (soft) DDA</td>
</tr>
<tr>
<td>0D28</td>
<td>ங (NA)</td>
<td>Malayalam letter NA</td>
</tr>
<tr>
<td>0D2A</td>
<td>ங (PA)</td>
<td>Malayalam letter PA</td>
</tr>
<tr>
<td>0D2B</td>
<td>ச (PHA)</td>
<td>Malayalam letter PHA</td>
</tr>
<tr>
<td>0D2C</td>
<td>ஞ (BA)</td>
<td>Malayalam letter BA</td>
</tr>
<tr>
<td>0D2D</td>
<td>ஞ (BHA)</td>
<td>Malayalam letter BHA</td>
</tr>
<tr>
<td>0D2E</td>
<td>ஞ (MA)</td>
<td>Malayalam letter MA</td>
</tr>
<tr>
<td>0D2F</td>
<td>ஞ (YA)</td>
<td>Malayalam letter YA</td>
</tr>
<tr>
<td>0D30</td>
<td>ஞ (RA)</td>
<td>Malayalam letter RA</td>
</tr>
<tr>
<td>0D31</td>
<td>ஆ (RRA)</td>
<td>Malayalam letter RRA</td>
</tr>
<tr>
<td>0D32</td>
<td>ஆ (LA)</td>
<td>Malayalam letter LA</td>
</tr>
<tr>
<td>0D33</td>
<td>ஆ (LLA)</td>
<td>Malayalam letter LLA</td>
</tr>
<tr>
<td>0D34</td>
<td>ஞ (ZHA)</td>
<td>Malayalam letter ZHA</td>
</tr>
</tbody>
</table>

Specified as CA in standard.  
To be modified as CHA.  
Specified as CHA in standard.  
To be modified as CHHA.  
Specified as NYA in standard.  
To be modified as NHA.  
Specified as TTA in standard.  
To be modified as TA.  
Specified as TTHA in standard.  
To be modified as TTA.  
Specified as DDA in standard.  
To be modified as (hard) DA.  
Specified as DDHA in standard.  
To be modified as (hard) DDA.  
Specified as NNA in standard.  
To be modified as (hard) NA.  
Specified as TA in standard.  
To be modified as THA.  
Specified as THA in standard.  
To be modified as TTHA.  
Specified as DA in standard.  
To be modified as (soft) DA.  
Specified as DHA in standard.  
To be modified as (soft) DDA.  
Specified as LLLA in standard.  
To be modified as ZHA.
Malayalam keyboard layout and character encoding

<table>
<thead>
<tr>
<th>Code</th>
<th>Symbol</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D35</td>
<td>⵲</td>
<td>Malayalam letter VA</td>
<td>Specified as SHA in standard.</td>
</tr>
<tr>
<td>0D36</td>
<td>⵳</td>
<td>Malayalam letter (soft) SHA</td>
<td>To be modified as (soft) SHA</td>
</tr>
<tr>
<td>0D37</td>
<td>⵳</td>
<td>Malayalam letter SHA</td>
<td>Specified as SSA in standard.</td>
</tr>
<tr>
<td>0D38</td>
<td>⵲</td>
<td>Malayalam letter SA</td>
<td>To be modified as SHA</td>
</tr>
<tr>
<td>0D39</td>
<td>⵲</td>
<td>Malayalam letter HA</td>
<td></td>
</tr>
</tbody>
</table>

**Consonant signs**

<table>
<thead>
<tr>
<th>Code</th>
<th>Symbol</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D3A</td>
<td>⵲</td>
<td>Malayalam consonant sign RRA</td>
<td>To be added</td>
</tr>
<tr>
<td>0D3B</td>
<td>⵲</td>
<td>Malayalam consonant sign YA</td>
<td>To be added</td>
</tr>
<tr>
<td>0D3C</td>
<td>⵲</td>
<td>Malayalam consonant sign VA</td>
<td>To be added</td>
</tr>
<tr>
<td>0D3D</td>
<td>⵲</td>
<td>Malayalam consonant sign LLA</td>
<td>To be added</td>
</tr>
</tbody>
</table>

**Dependent vowel signs**

<table>
<thead>
<tr>
<th>Code</th>
<th>Symbol</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D3E</td>
<td>⵲</td>
<td>Malayalam vowel sign AA</td>
<td></td>
</tr>
<tr>
<td>0D3F</td>
<td>⵲</td>
<td>Malayalam vowel sign I</td>
<td></td>
</tr>
<tr>
<td>0D40</td>
<td>⵲</td>
<td>Malayalam vowel sign II</td>
<td></td>
</tr>
<tr>
<td>0D41</td>
<td>⵲</td>
<td>Malayalam vowel sign U</td>
<td></td>
</tr>
<tr>
<td>0D42</td>
<td>⵲</td>
<td>Malayalam vowel sign UU</td>
<td></td>
</tr>
<tr>
<td>0D43</td>
<td>⵲</td>
<td>Malayalam vowel sign vocalic R</td>
<td></td>
</tr>
<tr>
<td>0D46</td>
<td>⵲</td>
<td>Malayalam vowel sign E</td>
<td></td>
</tr>
<tr>
<td>0D47</td>
<td>⵲</td>
<td>Malayalam vowel sign EE</td>
<td></td>
</tr>
<tr>
<td>0D48</td>
<td>⵲</td>
<td>Malayalam vowel sign AI</td>
<td></td>
</tr>
<tr>
<td>0D4A</td>
<td>⵲ ⵲</td>
<td>Malayalam vowel sign O</td>
<td></td>
</tr>
<tr>
<td>0D4B</td>
<td>⵲ ⵲</td>
<td>Malayalam vowel sign OO</td>
<td></td>
</tr>
<tr>
<td>0D4C</td>
<td>⵲ ⵲</td>
<td>Malayalam vowel sign AU</td>
<td>To be deleted. Malayalam vowel sign AU is ⵲ (0D57)</td>
</tr>
</tbody>
</table>

**Various signs**

<table>
<thead>
<tr>
<th>Code</th>
<th>Symbol</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D4D</td>
<td>⵲</td>
<td>Malayalam sign Chandrakkala</td>
<td>Specified as Virama in standard. To be corrected as chandrakkala</td>
</tr>
<tr>
<td>0D57</td>
<td>⵲</td>
<td>Malayalam vowel sign AU</td>
<td>Specified as AU length mark. To be corrected as vowel sign AU</td>
</tr>
</tbody>
</table>

**Chillu**

<table>
<thead>
<tr>
<th>Code</th>
<th>Symbol</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D58</td>
<td>⵲</td>
<td>Malayalam letter L</td>
<td>To be added</td>
</tr>
<tr>
<td>0D59</td>
<td>⵲</td>
<td>Malayalam letter LL</td>
<td>To be added</td>
</tr>
<tr>
<td>0D5A</td>
<td>⵲</td>
<td>Malayalam letter RR</td>
<td>To be added</td>
</tr>
<tr>
<td>0D5B</td>
<td>⵲</td>
<td>Malayalam letter NN</td>
<td>To be added</td>
</tr>
<tr>
<td>0D5C</td>
<td>⵲</td>
<td>Malayalam letter N</td>
<td>To be added</td>
</tr>
</tbody>
</table>
4.2 Modifications required in ISCII

The modifications required in ISCII for the character set of Malayalam are:

1) Include chillu characters in ISCII with the codes as shown

<table>
<thead>
<tr>
<th>Code</th>
<th>Decimal</th>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EB</td>
<td>235</td>
<td>૗</td>
<td>Chillu N</td>
</tr>
<tr>
<td>EC</td>
<td>236</td>
<td>૖</td>
<td>Chillu L</td>
</tr>
<tr>
<td>ED</td>
<td>237</td>
<td>૕</td>
<td>Chillu LL</td>
</tr>
<tr>
<td>EE</td>
<td>238</td>
<td>૗</td>
<td>Chillu R</td>
</tr>
<tr>
<td>FB</td>
<td>251</td>
<td>૗</td>
<td>Chillu NN</td>
</tr>
</tbody>
</table>

2) Include consonant signs and assign codes as shown below

<table>
<thead>
<tr>
<th>Code</th>
<th>Decimal</th>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
<td>206</td>
<td>𐤇</td>
<td>Consonant sign YA</td>
</tr>
<tr>
<td>FC</td>
<td>252</td>
<td>𐤋</td>
<td>Consonant sign VA</td>
</tr>
<tr>
<td>FD</td>
<td>253</td>
<td>𐤑</td>
<td>Consonant sign RA</td>
</tr>
<tr>
<td>FE</td>
<td>254</td>
<td>𐤒</td>
<td>Consonant sign hard LA</td>
</tr>
</tbody>
</table>

3) EXT is not required in Malayalam. Hence the code for that (F0 hex) may be deleted.

The ISCII character codes with the recommended modifications is given below
<table>
<thead>
<tr>
<th>Position</th>
<th>Hex</th>
<th>Decimal</th>
<th>Character</th>
<th>Name</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>161</td>
<td>161</td>
<td>အ</td>
<td>Anuswara</td>
<td>Reserved</td>
</tr>
<tr>
<td>A2</td>
<td>162</td>
<td>162</td>
<td>အ</td>
<td>Visarga</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>163</td>
<td>163</td>
<td>အ</td>
<td>Vowel A</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>164</td>
<td>164</td>
<td>အ</td>
<td>Vowel AA</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>165</td>
<td>165</td>
<td>အ</td>
<td>Vowel I</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>166</td>
<td>166</td>
<td>အ</td>
<td>Vowel U</td>
<td></td>
</tr>
<tr>
<td>A7</td>
<td>167</td>
<td>167</td>
<td>အ</td>
<td>Vowel UU</td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>168</td>
<td>168</td>
<td>အ</td>
<td>Vowel R</td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>169</td>
<td>169</td>
<td>အ</td>
<td>Vowel EE</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>170</td>
<td>170</td>
<td>အ</td>
<td>Vowel AI</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>171</td>
<td>171</td>
<td>အ</td>
<td>Vowel O</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>172</td>
<td>172</td>
<td>အ</td>
<td>Vowel OO</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>173</td>
<td>173</td>
<td>အ</td>
<td>Vowel AU</td>
<td></td>
</tr>
<tr>
<td>AE</td>
<td>174</td>
<td>174</td>
<td>အ</td>
<td>Consonant KA</td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>175</td>
<td>175</td>
<td>အ</td>
<td>Consonant KHA</td>
<td></td>
</tr>
<tr>
<td>B0</td>
<td>176</td>
<td>176</td>
<td>အ</td>
<td>Consonant GA</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>177</td>
<td>177</td>
<td>အ</td>
<td>Consonant GHA</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>178</td>
<td>178</td>
<td>အ</td>
<td>Consonant NGA</td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>179</td>
<td>179</td>
<td>အ</td>
<td>Consonant CHA</td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>180</td>
<td>180</td>
<td>အ</td>
<td>Consonant CHHA</td>
<td></td>
</tr>
<tr>
<td>B5</td>
<td>181</td>
<td>181</td>
<td>အ</td>
<td>Consonant JA</td>
<td></td>
</tr>
<tr>
<td>B6</td>
<td>182</td>
<td>182</td>
<td>အ</td>
<td>Consonant JHA</td>
<td></td>
</tr>
<tr>
<td>B7</td>
<td>183</td>
<td>183</td>
<td>အ</td>
<td>Consonant NHA</td>
<td></td>
</tr>
<tr>
<td>B8</td>
<td>184</td>
<td>184</td>
<td>အ</td>
<td>Consonant TA</td>
<td></td>
</tr>
<tr>
<td>B9</td>
<td>185</td>
<td>185</td>
<td>အ</td>
<td>Consonant TTA</td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>186</td>
<td>186</td>
<td>အ</td>
<td>Consonant hard DA</td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>187</td>
<td>187</td>
<td>အ</td>
<td>Consonant hard DDA</td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>188</td>
<td>188</td>
<td>အ</td>
<td>Consonant hard NA</td>
<td></td>
</tr>
<tr>
<td>BD</td>
<td>189</td>
<td>189</td>
<td>အ</td>
<td>Consonant TTHA</td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>190</td>
<td>190</td>
<td>အ</td>
<td>Consonant soft DA</td>
<td></td>
</tr>
<tr>
<td>BF</td>
<td>191</td>
<td>191</td>
<td>အ</td>
<td>Consonant soft DDA</td>
<td></td>
</tr>
<tr>
<td>C0</td>
<td>192</td>
<td>192</td>
<td>အ</td>
<td>Consonant soft NA</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>193</td>
<td>193</td>
<td>အ</td>
<td>Consonant PA</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>194</td>
<td>194</td>
<td>အ</td>
<td>Consonant PHA</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>195</td>
<td>195</td>
<td>အ</td>
<td>Consonant BA</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>196</td>
<td>196</td>
<td>အ</td>
<td>Consonant BHA</td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Key</td>
<td>Character</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC 204</td>
<td>ർ</td>
<td>Consonant MA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD 205</td>
<td>ྷ</td>
<td>Consonant YA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE 206</td>
<td>൷</td>
<td>Consonant sign YA Addition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF 207</td>
<td>൹</td>
<td>Consonant RA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D0 208</td>
<td>โ</td>
<td>Consonant RRA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1 209</td>
<td>ใ</td>
<td>Consonant LA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2 210</td>
<td>ไ</td>
<td>Consonant hard LA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3 211</td>
<td>ๅ</td>
<td>Consonant ZHA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D4 212</td>
<td>ๆ</td>
<td>Consonant VA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D5 213</td>
<td>็</td>
<td>Consonant soft SHA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D6 214</td>
<td>่</td>
<td>Consonant SHA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D7 215</td>
<td>้</td>
<td>Consonant SA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8 216</td>
<td>๊</td>
<td>Consonant HA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9 217</td>
<td>INV</td>
<td>Inverse key See note 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DA 218</td>
<td>ັ</td>
<td>Vowel sign AA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB 219</td>
<td>າ</td>
<td>Vowel sign I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DC 220</td>
<td>ຳ</td>
<td>Vowel sign II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DD 221</td>
<td>ິ</td>
<td>Vowel sign U</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE 222</td>
<td>ີ</td>
<td>Vowel sign UU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF 223</td>
<td>ຶ</td>
<td>Vowel sign RR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E0 224</td>
<td>ື</td>
<td>Vowel sign E</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1 225</td>
<td>ຸ</td>
<td>Vowel sign EE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2 226</td>
<td>ູ</td>
<td>Vowel sign AI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E3 227</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E4 228</td>
<td>຺</td>
<td>Vowel sign O</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5 229</td>
<td>ົ</td>
<td>Vowel sign OO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6 230</td>
<td>ຼ</td>
<td>Vowel sign AU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E7 231</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E8 232</td>
<td>ຽ</td>
<td>Chandrakkala</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E9 233</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EA 234</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EB 235</td>
<td>຾</td>
<td>Chillu N Addition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC 236</td>
<td>຿</td>
<td>Chillu L Addition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED 237</td>
<td>຾</td>
<td>Chillu LL Addition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE 238</td>
<td>຿</td>
<td>Chillu R Addition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF 239</td>
<td>ATR</td>
<td>Attribute code See Note1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 240</td>
<td>Delete</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 241</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2 242</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3 243</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F4 244</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F5 245</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F6 246</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F7 247</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F8 248</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F9 249</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Malayalam keyboard layout and character encoding

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FA</td>
<td>250</td>
<td>Reserved</td>
</tr>
<tr>
<td>FB</td>
<td>251</td>
<td>φñδ</td>
</tr>
<tr>
<td>FC</td>
<td>252</td>
<td>Consonant sign VA</td>
</tr>
<tr>
<td>FD</td>
<td>253</td>
<td>Consonant sign RA</td>
</tr>
<tr>
<td>FE</td>
<td>254</td>
<td>Consonant sign hard LA</td>
</tr>
</tbody>
</table>

**Note 1**: Attribute code (ATR) defines a font attribute applicable to the following characters. The mechanism is meant for use in the medium where alternative font selection mechanism is not available. Details are given in Annexure E of the BIS document on ISCII.

**Note 2**: INVERSE key (code: D9) is used to modify visual representations of glyph sets from the normal display. Refer BIS document on ISCII and CDAC keyboard manual for details.
5. Lexicographic ordering of Malayalam characters

5.1 Searching and Sorting

Two utilities of main major use in any document, database or Internet applications are Searching and Sorting. Any character encoding scheme should address both these utilities.

Searching can be for a character, word or combination of words. Searching becomes complicated when the same word can be written in different ways, as is practiced in Malayalam. Many initiatives have been taken at the Government and other levels to standardize the script in Malayalam. One of the major initiatives taken by the Government of Kerala recently is known as Malayalathanima.

Malayalathanima organized various seminars and workshops to arrive at a suitable standard for writing and printing of Malayalam script. The activities were coordinated by State Institute of Languages and the expertise of Central Institute of Indian Languages, Department of Official Languages (Kerala), Kerala Sahitya Academy, NCERT, Press Academy, Department of Linguistics (University of Kerala) and various other organizations were made use of in the process. The State Institute of Languages came out with a recommendation, a style book, which was evolved from the seminars and workshops. These recommendations were discussed and debated in various forums and some of the suggestions were also included in the latest edition of the book.

Sorting is arranging words in the lexicographic ordering of characters. Many times the linear order based on the codes of the characters cannot be used as the basis for sorting.

Sorting is one of the major utilities of any database applications. Be it the telephone directory, the voter's list, dictionary, encyclopedia or the index of books, people expect the ordering of names and other words to be ordered in a unique way, so that it is easy to find out a name or a word from a list or a dictionary.

In Malayalam, different authors have used different ordering schemes for the lexicographic ordering of words and there has not been a clearly defined standard for the ordering. Thus, it is highly essential that a unique lexicographic ordering scheme be specified at the earliest.

The basic factors which were weighed to arrive at the lexicographic ordering are:

- Some studies have been made in this direction by eminent scholars and their inputs have been made use of.
- The ordering scheme is totally based on the script and the pronunciation of words will not have any role on the scheme.
- The scheme is independent of the coding scheme being used.
The following order of characters is recommended for the lexicographic ordering of Malayalam words.

A complete listing of the characters arranged in lexicographic order is given below. It is possible that some of the conjunct letters shown are not used in Malayalam. But they have been shown for the completeness of the ordering scheme and to avoid any ambiguity in deciding the sorting order.

Some of the examples shown may be English words written in Malayalam script. This have been done in cases where the point is clearly explained by the given examples.
Malayalam keyboard layout and character encoding

시스템

한글

수상한

저작권 정보

Report of the Committee
May 2001
39/48
The following points may be noted in deciding the lexicographic ordering of Malayalam words:

1. The chillu characters are treated as the pure consonants of the corresponding consonants. 
   \((^N - n; \; ^T - M; \; ^S - m; \; ^S - s; \; ^T - T; \; ^S - L)\)

2. Anuswaram \( (a) \) is treated as the pure consonant of \( a \).

3. Visargam \( (i) \) is treated as the pure consonant of \( a \).

3. Space, punctuation marks takes precedence over (shall come before) any other characters. Pairs of words arranged in the sort order are given below.

   - \( \text{കല്ലെ} - \text{കല്ലെ} \)
   - \( \text{കാന്തി} - \text{കാന്തി} \)
   - \( \text{കരുവിനം} - \text{കരുവിനം} \)

4. Conjunct letter shall be treated as a single entity for deciding the lexicographic order.

   Ex: \( \text{ആട്ടാ} - \text{ആട്ടാ} \)
   Here the comparison is between \( ആണ് \) and \( മും (മിക) \).

5. Visargam \( (i) \) is treated as the pure consonant of \( a \) and is treated as a separate character.

   Ex: \( \text{ആല്ലീ} - \text{ആല്ലീ} \)
   Here, the comparison is between \( ഒയ്മ \) and \( മി \).

6. While comparing the conjunct letters for deciding the order, the character components of
   the conjunct letter shall be considered in the order of formation of the conjunct letter.

   Ex: \( \text{ആട്ടാ} - \text{ആട്ടാ} \)
   Here, the comparison is between \( (a) \) and \( (a) \). \( (a) \) is formed by characters \( a ; \; O \) in that order and \( (a) \) is formed by characters \( a ; \; O \) in that order. Since \( a \) takes precedence over \( O \), \( (a) \) comes before \( (a) \).

   \( \text{ഞണ്ണാട്ടി} - \text{ഞണ്ണാട്ടി} \)
   Here the comparison is between \( (a) \) and \( (a) \). \( (a) \) is formed by characters \( a ; \; O \) in that order and \( (a) \) is formed by characters \( a ; \; O \) in that order. Since \( a \) takes precedence over \( O \), \( (a) \) comes before \( (a) \).

7. The lexicographic order is decided exclusively based on the script. The pronunciation is not considered.
Ex: ക്കാരണം - ക്കാരണം (even though ഒ is pronounced as ഒ)

8. Only the chandrakkala at the end of a word will be treated as ബന്ധിപ്പിക്കുന്ന. Chandrakkala coming within a word (followed by other character(s) of the word) denotes a conjunct letter formed by the character(s) preceding and following the chandrakkala.

Ex: അക്കാരണം - അക്കാരണം (chandrakkala treated as ബന്ധിപ്പിക്കുന്ന.

7. Chillaksharas and Anuswaram will take precedence over the corresponding pure consonants. That is, ഒ will come before ഒ, എ before ഒ, എ before എ, എ before എ, എ before ആ and ഒ before ആ.

Ex: ക്കാരണം - ക്കാരണം (chandrakkala treated as ബന്ധിപ്പിക്കുന്ന.

8. Chillu and Anuswaram will take precedence over the corresponding consonants. That is, ഒ will come before ഒ, എ before ഒ, എ before ഒ, എ before ആ and ഒ before ആ.

Ex: ഓ - ഓ

9. Chillu and anuswaram appearing within a word (between other characters) is treated combined with the following character, as a conjunct letter.

Ex: അക്കാരണം - അക്കാരണം (chandrakkala treated as conjunct letter ബന്ധിപ്പിക്കുന്ന)

The examples shown below illustrate the guidelines for arranging the words in lexicographic order.
Ipam³ þ Ipw³ (w³ treated as the conjunct av³)

NtImZcw þ N{Iam³Wn (³ takes precedence over \, where ³ is the last character of a word.)

cwKw þ cwKame (w³ takes precedence over a, where w is the last character of a word)

A\p{IaWnIþ A\p{Iaw (W³ takes precedence over w³)

A´cw þ A³hÀ (³h³ treated as conjunct letter \v h³)

Atâm¬ þ A\zbw (â³ is treated as conjunct letter \vd, which takes precedence over \v h³)

AXn\p{Ipsda þ AXnsâbpw (â³ treated as conjunct letter \vd)

BÂþ Beþ Bev þ BÂacw (Âa³ treated as conjunct letter eva³)

A\´³ þ A \´\mcmbW³ þ A\´³\]nÅ (³\]³ treated as conjunct letter \v\])

AXv{Io¨nsb þ A¯ÀknwKv þ A¯kmen
(XvIvdo (XvIvdo takes precedence to XvX³, and Àkn³ takes precedence to km³)

AZaym³ þ AZwk þ AZm¯p (avb³ takes precedence to avk³ and Z³ comes before Zm³)

\]Zva\m`³ þ \]Zvaw (w³ is treated as pure consonant of a³)
6. Other recommendations

6.1 Visual representation of characters

Malayalam characters can combine or change shape, depending on their context. A character's appearance may be affected by its ordering with respect to other characters, the font used to render the character, and the application and system environment.

The coding schemes we have recommended, both ISCII (with modifications) and Unicode (with modifications) do not specify the visual representations of characters. They completely de-link the codes from the displayed fonts. The same text can be seen in different font styles by using a different font composition routine. A word can be displayed in a variety of styles depending on the conjunct repertoire used.

These standards define how the characters are interpreted, not how glyphs are rendered. The software or hardware rendering engine of a computer is responsible for the appearance of a character on the screen. The standards do not specify the shape, size or orientation of the characters on the screen or on the paper.

6.2 Malayalam Research Centre

Malayalam. Documents of the future will be generated and distributed through Internet and computer networks. We need to have a lot of contents in Malayalam to be generated and distributed through Internet and other electronic media, which will be useful to the common man. We also need to have the tools and technologies available in Malayalam, which will aid the generation and distribution of documents as well as other applications in Malayalam. The Committee recommends to the Government to set up a Malayalam Research Centre with the following broad objectives:

- To be a repository of Malayalam language tools
- To develop tools and technologies in Malayalam
- To exploit the emerging technologies for use in Malayalam computing
- Assist Government in technology dissemination

The Centre can take up activities like

- Development OS support and liaisoning with OS developers
- Development of database support and liaisoning with DBMS developers
- Develop resources like
  - Corpora and lexical resources
  - Malayalam fonts, to be distributed freely
  - Malayalam based AI tools
  - Electronic dictionary, Thesaurus, encyclopedia
  - Malayalam language website
• Development of tools for content generation
  ▪ Malayalam learning tools
  ▪ Curriculum based educational software and games in Malayalam
  ▪ Web-enabled and multimedia content generation
  ▪ Development/ enhancement of Multimedia authoring tools in Malayalam
  ▪ Malayalam searching and sorting engine
  ▪ Spell-checker and grammar-checker

• Development of web-centric applications
  ▪ Malayalam browser
  ▪ Malayalam email
  ▪ Malayalam search engine
  ▪ Web publishing tools in Malayalam

• Malayalam Interface tools
  ▪ OCR
  ▪ Text to Speech synthesis system
  ▪ Speech to text conversion system
  ▪ Translation support systems, including Brail
  ▪ Malayalam interfaces including voice

6.3 Official Malayalam Dictionary

The Committee recommends to the Government to publish an official Malayalam dictionary. The dictionary shall be prepared in such a way as to be used as a reference for the linguists, the software developers and the common man alike. The dictionary shall aid the standardising process of the script of Malayalam words and words will be arranged in the order mentioned in this recommendation. This will help in the software development activities like electronic dictionary, thesaurus, sorting engine, search engine, spell check etc.

Flexibility in the way of writing the scripts of Malayalam words will be one of the major stumbling blocks in the development of tools like search engine, spell check, thesaurus etc. Efforts are required to standardize the scripts of Malayalam words. Use of Style book (Malayalam Achatiyum Ezhuthum - oru style pusthakam) brought out by State Institute of Languages to be encouraged.

6.4 Government initiatives required for implementation of the standard

If the standard recommended in this document is to be widely used, Government may implement the following plan of action.

• Official declaration of the standard to be published
• A committee to be formed to evaluate any feedback and amend the standard, if required
• BIS to be approached to amend the keyboard and character encoding standards in Malayalam
• CDAC to be approached to make amendments in the applications developed by them
• Liaise with MIT closely on the issues related to Unicode
• Take initiative for the development of at least one Malayalam font, which will work in all major platforms, especially Windows, Unix, Linux and Mac systems and distribute it freely on Internet.
• All software developers in Malayalam to be intimated about the new standard.
• Make it mandatory for all the Malayalam software to conform to the standard within six months.
• Any training in Malayalam software (DTP) to be recognized only if the package conforms to the standard.
• Government to procure only those Malayalam software packages, which conforms to the standard.
Reference

1. Indian Standard - Indian Script Code for Information Interchange (ISCII), Bureau of Indian Standards, Delhi, 1991


3. Malayalam Achatiyum Ezhuthum - oru style pusthakam, Kerala Bhasha Institute, 1999


5. Unicode Phase1-The Review: http://www.lib.berkeley.edu/SSEAL/SouthAsia/Review.html
Annexure 1

Members of the Committee for standardization of Malayalam keyboard layout and character encoding.

1. Shri P.Govinda Pillai Chairman
   Chairman, CDIT
   Thiruvananthapuram

2. Smt Aruna Sundararajan Convener
   Secretary, Dept of IT
   Thiruvananthapuram

3. Dr.S.P.Mudur Member
   Associate Director, National Centre for Software Technology
   Gulmohar Cross Road No 9
   Juhu, Mumbai – 400 049

4. Dr.M.R.Thampan Member
   Director, State Institute of Languages
   Nalanda, Thiruvananthapuram

5. Dr A.P.Andrews Kutty Member
   Head of the Department, Department of Linguistics
   University Of Kerala,
   Thiruvananthapuram

6. Shri Satish Babu Member
   Chairman, Computer Society of India
   Thiruvananthapuram Chapter

7. Ms Kala Sriram Member
   Team Coordinator, GIST
   CDAC, Ganesh Khind
   Pune – 411 007

8. Ezhumattoor Rajaraja Varma Member
   Dept of Official Languages
   Govt of Kerala, Thiruvananthapuram

9. Shri Sasi P.M Member
   Mission Coordinator
   IT Mission Group
   Thiruvananthapuram
Members co-opted to the Committee

1. Shri L. Natarajan  
   Dept of Official Languages  
   Thiruvananthapuram

2. Prof. Prabhodhachandran Nair  
   Former Head of the Dept (Linguistics)  
   University of Kerala

3. Shri R. Raveendra Kumar  
   Additional Director  
   ER&DCI, Thiruvanathapuram

4. Shri P. V. Unnikrishnan  
   Executive Director  
   Information Kerala Mission Group on IT  
   Thiruvananthapuram

5. Prof. K. S. Narayana Pillai  
   Language Expert