SINHALA CHARACTER CODE FOR INFORMATION INTERCHANGE
PART 1 : COLLATION SEQUENCE

SRI LANKA STANDARDS INSTITUTION

Sri Lanka Standard
SINHALA CHARACTER CODE FOR INFORMATION INTERCHANGE
PART 1 : COLLATION SEQUENCE

SLS 1134 Part 1 : 2006
(Incorporating Amendment No.1)

Gr. 4

SRI LANKA STANDARDS INSTITUTION
No. 17, Victoria Place
Elvitigala Mawatha
Colombo 08
SRI LANKA.

Sri Lanka Standards are subject to periodical revision in order to accommodate the progress made by industry. Suggestions for improvement will be recorded and brought to the notice of the Committees to which the revisions are entrusted.

This standard does not purport to include all the necessary provisions of a contract.

Sri Lanka Standard
SINHALA CHARACTER CODE FOR INFORMATION INTERCHANGE
PART 1 : COLLATION SEQUENCE

FOREWORD

This standard was approved by the Sectoral Committee on Information Technology and was authorized for adoption and publication as a Sri Lanka Standard by the Council of the Sri Lanka Standards Institution on 2006-10-25.

This standard should be read in conjunction with **SLS1134** : Sinhala character code for information interchange.

Sinhala collation is based on the order of Indic letters derived from Sanskrit, but has evolved its own conventions over the years. The dictionaries and other reference works which have been published since the 19th Century AD agree on these basic Sinhala collation sequence, but disagree in details.

A collation sequence is needed to set up data bases and other collections of information in Sinhala. Consequently the Information and Communication Technology Agency of Sri Lanka (ICTA) requested the University of Colombo School of Computing (UCSC) to recommend a suitable collation algorithm. This standard was developed based on the recommendations made by the UCSC, as modified by the ICTA.

In the preparation of this standard, the assistance provided by ICTA is gratefully acknowledged.

1. SCOPE

This standard prescribes the collation sequence for arranging a list of words or phrases in the Sinhala language.

2. DEFINITIONS

2.1 collation: Collation is the process of arranging a list of words or phrases in a standard order.

3. COLLATION SEQUENCE

This standard defines a dictionary collation sequence for use in compiling dictionaries and other scholarly works and a simple collation sequence for use in data processing and other activities on lists of personal and other names.

3.1 Dictionary collation sequence

The Dictionary collation is the canonical collation order, and should be used when correct collation, based on the linguistic derivation of Sinhala, is required, e.g for a dictionary. This collation is recommended for use in scholarly and academic activities.

3.2 Simple collation sequence

The Simple collation should be used for preparing lists of names, places etc, and would produce identical results as the dictionary collation sequence when collating personal name, place names and other common data. This collation is recommended for use in collating lists of names, locations, etc.

The two collations will produce different results only between words with the letter ඍ or ඎ, and the letter ඈ in a given position.

The Simple collation is both easier to implement, thus encouraging vendors to support Sinhala in their products, and produces a result which will not confuse a person who is not aware of the subtleties of the language.

It is recommended that implementations which only support a single collation sequence support the Simple collation sequence.

4 COLLATION ALGORITHM

The collation algorithm is based on syllables. A syllable may consist of a:
   i) vowel, e.g. අ, ඔෟ,
   ii) pure consonant, e.g. ක්, ර්,
   iii) consonant combined with a vowel, e.g. ක, කා, කැ, කෑ, ... කො, ... ග, ගා,...
   iv) semi-consonant, e.g. ං, ඃ

4.1 The Dictionary collation algorithm

The text to be collated is first decomposed using the following rules:

   a) Conjunct letters (බැඳි අකුරු) are decomposed into the equivalent <pure consonant, consonant-with-vowel> sequence e.g. ක්ෂේ->ක්ෂේ.
   b) Touching letters are decomposed into the equivalent <pure consonant, consonant-with-vowel> sequence e.g. ස්ව -> ස්ව, ම්ම -> ම්ම.
   c) The yansaya and rakaransaya are decomposed into their equivalent forms e.g:ක්‍ය -> ක්ය and ක්‍ර-> ක්ර.
   d) The repaya is decomposed in its equivalent form e.g.ර්ම ->ර්ම.
   e) The letter ඈ is decomposed as follows:
        ඈ->�්ඇ.
   Thus, ඈන is collated as being equivalent to ඇ්ඇන.

NOTE : *If words with both the composed and decomposed form are collated, the decomposed form should appear before the composed form.*

Each syllable in the de-composed text is evaluated from left-to-right, and collated in the following order:

අ, ආ, , ඇ, ඉ, ඊ, උ, ඌ, සa, සaa, ඎ, ඎෘ, එ, ඒ, ඓ, ඔ, ඕ, ඖ,

ා්, ාඃ,

ක, කා, කැ, කෑ, කි, කී, කු, කූ, කa, කaa, කෘ, කෲ, කෙ, කේ, කෛ, කො, කෝ, කෞ, ක්,

බ, බා, බැ, බෑ, බි, බී, බු, බූ, බa, බaa, බෘ, බෲ, බෙ, බේ, බෛ, බො, බෝ, බෞ, බ්,

ග, ගා, ගැ, ගෑ, ගි, ගී, ගු, ගූ, ගa, ගaa, ගෘ, ගෲ,ගෙ, ගේ, ගෛ, ගො, ගෝ, ගෞ, ග්,

ස, සා, සැ, සෑ, සි, සී, සු, සූ, සa, සaa, සෘ, සෲ, සෙ, සේ, සෛ, සො, සෝ, සෞ, ස්,

ඩ, ඩා, ඩැ, ඩෑ, ඩී, ඩී, ඩු, ඩූ, ඩa, ඩaa, ඩෘ, ඩෲ, ඩෙ, ඩේ, ඩෛ, ඩො, ඩෝ, ඩෞ, ඩ්,

හ, හා, හැ, හෑ, හී, හී, හු, හූ, හa, හaa, හෘ, හෲ, හෙ, හේ, හෛ, හො, හෝ, හෞ, හ්,

ච, චා, චැ, චෑ, චි, චී, චු, චූ, චa, චaa, චෘ, චෲ, චෙ, චේ, චෛ, චො, චෝ, චෞ, ච්,

ජ, ජා, ජැ, ජෑ, ජී, ජී, ජු, ජූ, ජa, ජaa, ජෘ, ජෲ, ජෙ, ජේ, ජෛ, ජො, ජෝ, ජෞ, ජ්,

ඡ, ඡා, ඡැ, ඡෑ, ඡී, ඡී, ඡු, ඡූ, ඡa, ඡaa, ඡෘ, ඡෲ, ඡෙ, ඡේ, ඡෛ, ඡො, ඡෝ, ඡෞ, ඡ්,

ඛ, ඛා, ඛැ, ඛෑ, ඛී, ඛී, ඛු, ඛූ, ඛa, ඛaa, ඛෘ, ඛෲ, ෙඛ, ඛේ, ෛඛ, ඛො, ඛෝ, ඛෞ, ඛ්,

ඦ, ඦා, ඦැ, ඦෑ, ඦී, ඦී, ඦු, ඦූ, ඦa, ඦaa, ඦෘ, ඦෲ, ඦෙ, ඦේ, ෛඦ, ඦො, ඦෝ, ෙඦෞ, ඦ්,

ඥ, ඥා, ඥැ, ඥෑ, ඥී, ඥී, ඥු, ඥූ, ඥa, ඥaa, ඥෘ, ඥෲ, ඥෙ, ඥේ, ෛඥ, ඥො, ඥෝ, ෙඥෞ, ඥ්,

ඣ, ඣා, ඣැ, ඣෑ, ඣී, ඣී, ඣු, ඣූ, ඣa, ඣaa, ඣෘ, ඣෲ, ඣෙ, ඣේ, ෙඣ, ඣො, ඣෝ, ඣෞ, ඣ්,

ඨ, ඨා, ඨැ, ඨෑ, ඨි, ඨී, ඨු, ඨූ, ඨa, ඨaa, ඨෘ, ඨෲ, ෙඨ, ඨේ, ෙඨෛ, ඨො, ඨෝ, ඨෞ, ඨ්,

ඬ, ඬා, ඬැ, ඬෑ, ඬී, ඬී, ඬු, ඬූ, ඬa, ඬaa, ඬෘ, ඬෲ, ෙඬ, ඬේ, ෙඬෛ, ඬො, ඬෝ, ඬෞ, ඬ්,

ඪ, ඪා, ඪැ, ඪෑ, ඪී, ඪී, ඪු, ඪූ, ඪa, ඪaa, ඪෘ, ඪෲ, ෙඪ, ඪේ, ෙඪෛ, ඪො, ඪෝ, ඪෞ, ඪ්,

ඵ, ඵා, ඵැ, ඵෑ, ඵී, ඵී, ඵු, ඵූ, ඵa, ඵaa, ඵෘ, ඵෲ, ෙඵ, ඵේ, ෙඵෛ, ඵො, ඵෝ, ඵෞ, ඵ්,

ණ, ණා, ණැ, ණෑ, ණී, ණී, ණු, ණූ, ණa, ණaa, ණෘ, ණෲ, ෙණ, ණේ, ෛණ, ණො, ණෝ, ෙණෞ, ණ්,

ඞ, ඞා, ඞැ, ඞෑ, ඞී, ඞී, ඞු, ඞූ, ඞa, ඞaa, ඞෘ, ඞෲ, ඞෙ, ඞේ, ෛඞ, ඞො, ඞෝ, ඞෞ, ඞ්,

ත, තා, තැ, තෑ, ති, තී, තු, තූ, තa, තaa, තෘ, තෲ, තෙ, තේ, ෛත, තො, තෝ, තෞ, ත්,

ථ, ථා, ථැ, ථෑ, ථී, ථී, ථු, ථූ, ථa, ථaa, ථෘ, ථෲ, ෙථ, ථේ, ෙථෛ, ථො, ථෝ, ථෞ, ථ්,

ද, දා, දැ, දෑ, දි, දී, දු, දූ, දa, දaa, දෘ, දෲ, දෙ, දේ, ෛද, දො, දෝ, දෞ, ද්,

ධ, ධා, ධැ, ධෑ, ධී, ධී, ධු, ධූ, ධa, ධaa, ධෘ, ධෲ, ෙධ, ධේ, ෙධෛ, ධො, ධෝ, ධෞ, ධ්,

න, නා, නැ, නෑ, නි, නී, නු, නූ, නa, නaa, නෘ, නෲ, නෙ, නේ, ෛන, නො, නෝ, නෞ, න්,

ඳ, ඳා, ඳැ, ඳෑ, ඳි, ඳී, ඳු, ඳූ, ඳa, ඳaa, ඳෘ, ඳෲ, ෙඳ, ඳේ, ෙඳෛ, ඳො, ඳෝ, ඳෞ, ඳ්,

ප, පා, පැ, පෑ, පි, පී, පු, පූ, පa, පaa, පෘ, පෲ, පෙ, පේ, ෛප, පො, පෝ, පෞ, ප්,

ඵ, ඵා, ඵැ, ඵෑ, ඵී, ඵී, ඵු, ඵූ, ඵa, ඵaa, ඵෘ, ඵෲ, ෙඵ, ඵේ, ෙඵෛ, ඵො, ඵෝ, ඵෞ, ඵ්,

බ, බා, බැ, බෑ, බී, බී, බු, බූ, බa, බaa, බෘ, බෲ, බෙ, බේ, ෛබ, බො, බෝ, බෞ, බ්,

හ, හා, හැ, හෑ, හී, හී, හු, හූ, හa, හaa, හෘ, හෲ, හෙ, හේ, ෛහ, හො, හෝ, හෞ, හ්,

ම, මා, මැ, මෑ, මී, මී, මු, මූ, මa, මaa, මෘ, මෲ, මෙ, මේ, ෛම, මො, මෝ, මෞ, ම්,

ඹ, ඹා, ඹැ, ඹෑ, ඹී, ඹී, ඹු, ඹූ, ඹa, ඹaa, ඹෘ, ඹෲ, ඹෙ, ඹේ, ෛඹ, ඹො, ඹෝ, ඹෞ, ඹ්,

ය, යා, යැ, යෑ, යී, යී, යු, යූ, යa, යaa, යෘ, යෲ, යෙ, යේ, ෛය, යො, යෝ, යෞ, ය්,

ර, රා, රැ, රෑ, රි, රී, රෘ, රෲ, රa, රaa, රෘ, රෲ, රෙ, රේ, ෛර, රො, රෝ, රෞ, ර්,

ල, ලා, ලැ, ලෑ, ලි, ලී, ලෘ, ලෲ, ලa, ලaa, ලෘ, ලෲ, ෙල, ලේ, ෛල, ලො, ලෝ, ලෞ, ල්,

ව, වා, වැ, වෑ, වී, වී, වු, වූ, වa, වaa, වෘ, වෲ, වෙ, වේ, ෛව, වො, වෝ, වෞ, ව්,

ශ, ශා, ශැ, ශෑ, ශී, ශී, ශු, ශූ, ශa, ශaa, ශෘ, ශෲ, ශෙ, ශේ, ෛශ, ශො, ශෝ, ශෞ, ශ්,

ෂ, ෂා, ෂැ, ෂෑ, ෂී, ෂී, ෂු, ෂූ, ෂa, ෂaa, ෂෘ, ෂෲ, ෂෙ, ෂේ, ෛෂ, ෂො, ෂෝ, ෂෞ, ෂ්,

ස, සා, සැ, සෑ, සි, සී, සු, සූ, සa, සaa, සෘ, සෲ, සෙ, සේ, ෛස, සො, සෝ, සෞ, ස්,

හ, හා, හැ, හෑ, හී, හී, හු, හූ, හa, හaa, හෘ, හෲ, හෙ, හේ, ෛහ, හො, හෝ, හෞ, හ්,

ළ, ළා, ළැ, ළෑ, ළී, ළී, ළු, ළෟ, ළa, ළaa, ළෘ, ළෲ, ෙළ, ළේ, ෛළ, ළො, ළෝ, ළෞ, ළ්,

ෆ, ෆා, ෆැ, ෆෑ, ෆි, ෆී, ෆු, ෆූ, ෆa, ෆaa, ෆෘ, ෆෲ, ෙෆ, ෆේ, ෛෆ, ෆො, ෆෝ, ෆෞ, ෆ්.

4.2 The Simple collation algorithm

The algorithm for the Simple collation is the same as for the Dictionary collation sequence, except that the decomposition in step d) of 4.1 is omitted. Therefore, ඦ is not decomposed into ඦ‍ඦ but treated as a single letter.

NOTES:

1. *The vowels ඔ and ඖ (ilu, iluu) and their modifiers ෘ and ෲ do not occur in Sinhala words. However, they are included in the collation sequnce.*

2. *Not all of the above syllables occur in Sinhala but are listed for completeness as words with such combinations may, though incorrect, appear in text to be collated.*

For example, the vowels ැ, ෑ are not found in Sanskrit, and would never modify ඎ. Also, the vowels ො, ෝ would never modify ඓ or ඖ. Inclusion in the above listing does not imply that such combinations are valid.

3. *The decomposition specified in 4.1 a) to e) is only for the purpose of determining the collation order. The actual text will not be decomposed, and will appear in the correct Sinhala representation.*

4. *Combinations of letters with the yansaya, rakaransaya an repaya do not appear in the above list, as they are decomposed in steps c) and d).*

5. *The sound "ru" may be written either as in කෘත or in ඍෂිය depending on the origin of the word. These two constructs are collated differently, based on their composition, as consonant + ෘ and consonant + ʰ+ ර + ු respectively.*

Loan words such as ග්‍රෑන්ඩ් containing the sound ෑ are collated as consonant + ʰ+ ර + ෑ.

The long versions of these vowel modifiers are treated similarly.

6. *The characters ඍ and ඎ are interchanged in the above list when compared with the Unicode code-point order. This is relevant only in the Simple collation, as the letter ඎ is decomposed when performing the Dictionary collation.*

7. *The Kundaliya ෴ is not included in these algorithms, but should be collated before all the letters.*

------------------/

SLS CERTIFICATION MARK

The Sri Lanka Standards Institution is the owner of the registered certification mark shown below. Beneath the mark, the number of the Sri Lanka Standard relevant to the product is indicated. This mark may be used only by those who have obtained permits under the SLS certification marks scheme. The presence of this mark on or in relation to a product conveys the assurance that they have been produced to comply with the requirements of the relevant Sri Lanka Standard under a well designed system of quality control inspection and testing operated by the manufacturer and supervised by the SLSI which includes surveillance inspection of the factory, testing of both factory and market samples.

Further particulars of the terms and conditions of the permit may be obtained from the Sri Lanka Standards Institution, 17, Victoria Place, Elvitigala Mawatha, Colombo 08.

SRI LANKA STANDARDS INSTITUTION

The Sri Lanka Standards Institution (SLSI) is the National Standards Organization of Sri Lanka established under the Sri Lanka Standards Institution Act No. 6 of 1984 which repealed and replaced the Bureau of Ceylon Standards Act No. 38 of 1964. The Institution functions under the Ministry of Science & Technology.

The principal objects of the Institution as set out in the Act are to prepare standards and promote their adoption, to provide facilities for examination and testing of products, to operate a Certification Marks Scheme, to certify the quality of products meant for local consumption or exports and to promote standardization and quality control by educational, consultancy and research activity.

The Institution is financed by Government grants, and by the income from the sale of its publications and other services offered for Industry and Business Sector. Financial and administrative control is vested in a Council appointed in accordance with the provisions of the Act.

The development and formulation of National Standards is carried out by Technical Experts and representatives of other interest groups, assisted by the permanent officers of the Institution. These Technical Committees are appointed under the purview of the Sectoral Committees which in turn are appointed by the Council. The Sectoral Committees give the final Technical approval for the Draft National Standards prior to the approval by the Council of the SLSI.

All members of the Technical and Sectoral Committees render their services in an honorary capacity. In this process the Institution endeavours to ensure adequate representation of all view points.

In the International field the Institution represents Sri Lanka in the International Organization for Standardization (ISO), and participates in such fields of standardization as are of special interest to Sri Lanka.