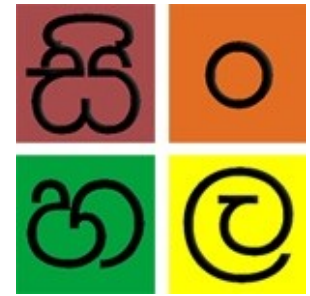
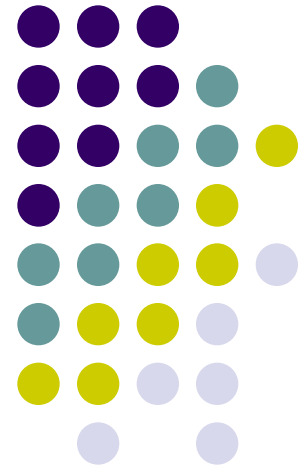


# Recent Developments in Local Language Computing in Sri Lanka



**Gihan Dias**  
University of Moratuwa

**Aruni Goonetilleke**  
ICT Agency of Sri Lanka



# A New Era



- We are currently at the beginning of a new era of computing in Sri Lanka
  - where we do not have to use a foreign language
- Computers, phones and the Internet now work in our own languages
- Very soon we may forget the era where they were only in English



# The Journey



- Started in the mid-1980s
  - pioneering efforts by DMS and Metropolitan
- Sinhala Alphabet
  - සිංහල හෝඩියේ අකුරු කීයද?
- Sinhala in Unicode
  - Documented several issues and published as Sri Lanka Standard 1134:2004
- Unicode has **full** support for both contemporary and classical Sinhala



# Sinhala Unicode



- Unicode uses one code for each vowel and each consonant
  - අ, ආ, ඓ, ග, ම, භ, ක්
- One code for each vowel modifier
  - ක් is represented by two codes, ක + ො
  - කො is represented by two codes, ක + ෝ
- Follows the rules of Sinhala Grammar
- We can write කුරුළු, පකෂ, මිත්‍යා, ශ්‍රී, වගී



# The Journey (Cont.)



- Standard Sinhala Computer Keyboard
  - based on Wijesekera layout
  - easy for existing Sinhala typists
- Fonts
  - ඉස්කෝල පොත
  - මලිකි
  - සරසවි
  - දිනමිණ
  - භාෂිත

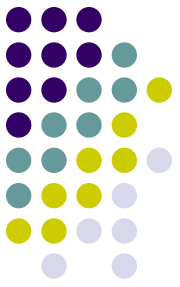




# Tamil Initiatives



# Tamil Keyboard



- Initially proposed use of Tamil99 keyboard
- Not liked by users (who preferred to use what they are familiar with)
- Standardised a variant of Renganathan layout
  - in SLS 1326
  - typewriter-style: type-as-you-write
  - similar in concept (not layout) to Sinhala keyboard



# Tamil Character Code



- Did not make any changes to the existing Unicode encoding
- the Aytham ஃஃ is used to produce the “f” sound
  - e.g. ஃஃபஃ ஃஃ க்ஃஸ் (fax)
- Supports the ஸ்ரீ (shri) and க்ஷ (ksha) letters
- Also defines Tamil numerals, Tamil symbols and the Tamil OOM sign





# Tamil Collation



- Collation is the order in which a list of words or phrases (e.g. names) are sorted
- Several different ways of collating Tamil
- SLS 1326 Part 1:2008 specifies:
  - Vowels first
  - Tamil consonants
  - Grantha consonants
  - the ஃழ (fa) sequence
  - symbols and Tamil numerals come last





# Sinhala Initiatives



# Sinhala Collation Sequence



- SLS1134 Part 1:2007
  - Based on Sinhala alphabetical order
- The letter ළ was a problem
- Dictionary collation sequence
  - treats ළ as the conjunct ඵ+ළ
  - should be used in compiling dictionaries and other scholarly works
- Simple collation sequence
  - treats ළ as a single letter
  - for data processing and lists of names.



# Fonts



- Identified requirements of Sinhala fonts
  - Level 1 – for phones and mobile devices
  - Level 2 – for general use
  - Level 3 – for Pali and classical books
- Developed a new font Family – Bhashitha and a set of font rules
  - භාෂිත, භාෂිත තද
  - භාෂිත සැත්ප, භාෂිත සැත්ප තද
- Font rules may be used by anyone to develop their own fonts



# Sinhala and Tamil Computers



- Vision:
  - you can walk into the local computer shop and buy a computer running in Sinhala and Tamil
- Several versions of Linux (Ubuntu) have Sinhala and Tamil support
- Sinhala and Tamil Windows Vista
- Sinhala and Tamil MS-Office 2007 – July 2009
- Promotion of computers pre-installed with Sinhala
  - collaboration with Intel



# Terminology



- What do you call a **folder** in Sinhala?
- and **undo** ?
- People use various terms when translating software - UoM, ICTA, RedHat, etc.
- Official Languages Dept. glossary on-line  
– <http://www.languagesdept.gov.lk/glossary/>
- ICTA has set up a shared repository for Sinhala and Tamil Terms



# Tools and Utilities



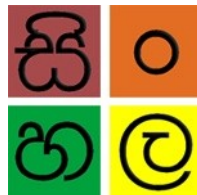
- 10-million word Sinhala Corpus
- Sinhala lexicon
- OCR, Text to Speech - UCSC
- Transliteration between Sinhala, Tamil, English
  - UoM, UCSC/ScienceLand
- Four Colour සිංහල logo
  - Winnie Hettigoda / UCSC
  - we need a proper Tamil logo



# Getting people together

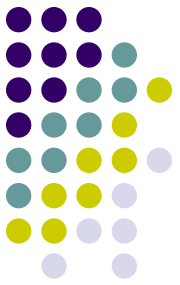


- Lanka Localisation Group  
<http://groups.google.com/group/lanka-localization-group>
- Sinhala Unicode Group  
<http://groups.google.com/group/Sinhala-Unicode/>
- Symposium on Localised Systems and Applications 2009
  - held at UoM on 2<sup>nd</sup> Sept
  - over 60 participants and over 20 presentations
- 2010 Symposium will be held in March 2010





# Acknowledgements



- ICTA Local Language Working Group (LLWG) and CINTEC Sinhala Fonts Committee
- Official Languages Dept.
- Sri Lanka Standards Institution
- UCSC, UoM, ACCIMT, Lake House, Lanka Linux Group, ...
- Microsoft, Microimage, ScienceLand, e-Fusion and others

## Especially:

– Harsha Wijayawardhana

– G. Balachandran



# Conclusion



- After over 20 years of work, we have -
  - extensive use of Sinhala/Tamil on the web, as well as in documents and applications
  - Basic technologies in place to use computers and information systems in Sinhala and Tamil
- We need -
  - more features
  - other platforms (phones, PDAs, etc.)
  - increase awareness and usage
  - more localised applications and content

