

How
ICT was
enabled in

Local Languages

How ICT was enabled in Local Languages



How ICT was enabled in Local Languages

Aruni Goonetilleke

How ICT was enabled in Local Languages

© Aruni Goonetilleke

May 2020

ISBN 978-624-5374-00-7

Cover Design

Pushpananda Ekanayake

LK Domain Registry all rights reserved. No part of this document may be reproduced or transmitted in any form or by any means without prior written permission from the LK Domain Registry.

This book is based primarily on interviews with the experts who implemented ICT in Sinhala and Tamil.

Published

LK Domain Registry

Bernard's Business Park, 106, Dutugemunu Street, Dehiwala.

TP: (011) 421-6061

Email: hostmaster@domains.lk

Web: www.domains.lk

PREFACE

For Sri Lanka to benefit from Information and Communication Technology (ICT), it had to be made available and accessible to the populace throughout the country. The previous apex body on ICT, the Computer and Information Technology Council of Sri Lanka (CINTEC), later named the Council for Information Technology, was established in the mid-1980s. To achieve its mandate, the benefits of ICT had to be taken to the country at large. Therefore, soon after CINTEC was set up, it commenced work on enabling the use of ICT in Sinhala.



Later, in 2003, when CINTEC was closed and its successor, the Information and Communication Technology Agency (ICTA) of Sri Lanka was set up, the work on Local Languages that CINTEC had been working on was taken over to ICTA, due to its necessity for ICTA to achieve what it planned to do.

I was a Board member of ICTA from 2005 and later its Chairman, and throughout my tenure at ICTA, its Board and the Leadership of ICTA ensured that enabling the use of ICT in Sinhala and Tamil was given due recognition, because it was one of the bases or a foundation on which the work of ICTA was built. Everything was being computerized throughout the world and it would have been difficult for people of Sri Lanka to master the linguistic and communicative aspects of English, so as to be able to use the English content that was profusely available. It was imperative that ICT in Sinhala and Tamil was enabled. It was during this period that I specifically became aware of the work that was being carried out and realized its importance. ICTA also set up the Local Languages Working Group, which was known by its acronym, LLWG. This Group comprised people with expertise in ICT and languages, linguists and also those who represented relevant organizations. This Group

and been functional under another name at CINTEC. The work carried out over many years by all, including the Universities, Government organizations, the private sector and individuals has now resulted in a situation where people who wish to use ICT in Sinhala and Tamil can do so, on computers and mobile devices. This is the result of dedicated and meticulous work carried out over many years and for which the country is thankful. This is well documented in this book.

Prof. P.W. Epasinghe

JAYANTHA FERNANDO'S MESSAGE

The LK Domain Registry decided to record the evolution of local language development in the ICT Sector in Sri Lanka. The Information and Communication Technology Agency (ICTA) of Sri Lanka led efforts to standardize (both internationally and domestically) and facilitate the use of Sinhala and Tamil in the digital media which was a pivotal policy achievement in Sri Lanka's digital transformation journey. As a result of this enabler, citizens are able to use Sinhala and Tamil in a variety of digital applications, including Facebook, Twitter and other social media, resulting in this work contributing to bridging the digital divide. Aruni has taken the trouble to meticulously record the important milestones through several interviews from those involved. Therefore, this book becomes the most valuable historical record and compilation for those who wish to know what happened in the area of ICT and local languages.



As ICTA's Program Manager responsible for the ground-breaking initiatives in this area for more than 12 years, Aruni Goonetilleke can testify to the enduring efforts carried out, including the work of collaborators, stakeholders and contributors, whose efforts are succinctly and meticulously recorded in this book. Her efforts to bring out this master-piece should be commended. On a personal note it was a pleasure for me to have had the opportunity to be associated in this area of work as the Program Director for the "ICT Policy, Leadership and Human Resources Development" component at ICTA, under which the Local Languages Initiative was implemented during the multi-donor funded "e-Sri Lanka Development Program". Under the e-Sri Lanka Program it was imperative for ICTA to implement strategies that would proliferate access to ICT and use of ICT applications in rural areas for rural

socio-economic development. In this context the Local Languages Initiative was the central pillar on which our progress was measured and I remain grateful for the efforts of all. Independent evaluations of the e-Sri Lanka Development Program show that its project development objectives were achieved and one of the key reasons was the fact that using ICT in Sinhala and Tamil was enabled by ICTA, its partners and stakeholders.

Jayantha Fernando

CONTENTS

PREFACE.....	i
JAYANTHA FERNANDO'S MESSAGE.....	iii
INTRODUCTION	ix
The need for Local Language Computing.....	1
PROLOG.....	2
THE WAY IT USED TO BE.....	4
Private Sector Initiatives	7
SINHALA IN WANG COMPUTERS.....	8
THE MAHANAMA SINHALA KEYBOARD.....	11
THIBUS	14
KAPUTA FONT AND KAPUTA.COM.....	22
MICROIMAGE, HELAWADANA, SHAN AND LOCALIZATION	26
ICT IN LOCAL LANGUAGES AT ANCL.....	27
SETT BROWSER, HELAKURU, AKURU HURUWA.....	32

Universities and CINTEC.....	37
LANGUAGE TECHNOLOGY RESEARCH LABORATORY – UCSC.....	41
LAKapps CENTER.....	52
THE COMPUTER AND INFORMATION TECHNOLOGY COUNCIL OF SRI LANKA (CINTEC).....	55
Trying to spread the Internet in Sinhala & Impediments encountered	65
The e-Sri Lanka development program, setting up ICTA & the need for local languages.....	71
We have come a Long Way.....	115
EPILOG.....	119
Acknowledgments	120
References.....	121

INTRODUCTION

Many of us are now used to the idea of using the web, computers, phones and other electronic devices in our own language – Sinhala or Tamil. However, some older people still remember when computers worked only in English, and if you didn't know that language, you were completely shut out of the information age. A small, diverse, band of people took on – and successfully overcame – the challenges of using computers in our languages.



The first question we faced was: “How do we represent our language – the letters and symbols – in the computer? We then had to figure out how to show them on the screen or printer, how to get text into the computer using a keyboard or other device, and how to build tools to process Sinhala and Tamil on computers. We also had to face challenges of people who thought that what we were doing was wrong, and we should work on making Lankans learn English – not computers work in Sinhala and Tamil.

How we in Sri Lanka worked to make our languages – and ourselves – first-class citizens of the cyberworld is a fascinating story with many chapters and actors. Aruni Goonetilleke has been a key part of this work and she is uniquely qualified to tell this story.

The narrative is presented in Aruni's inimitable way – as a story. Starting with the difficulties people faced in the 1990s, she moves back to the 1980s, and follows a number of personalities in their work, up to the present. Throughout, she presents her material in a very readable manner – there are no dry, text-book passages. When presenting a story like this, it is not possible to describe every detail, and some matters are covered only at a high level. Some stories

have multiple facets and actors, but not all of them are covered – sometimes because the people involved are no longer with us. Some may dispute details and point out that others too have contributed. However, I hope readers will excuse Aruni for such errors and omissions.

The book concludes with a second instalment of the initial story. Although “Kumara”’s problem is solved, we have not yet come to the end of our quest. Though by the end of 2019 we can now type and read text, many more things remain to be done – including speech recognition, text understanding, etc. The abilities of computing systems grow day-by-day, and we have to work hard just to keep up. Today’s computers can understand what we want, and respond to fulfil our needs – but not yet in our languages. We will continue to be busy for the next decade or more, making sure that our citizens can make full use of IT in our own language.

I hope that by reading this book, you will gain an appreciation of what has been done, and be motivated to contribute to what remains to be done.

Gihan Dias

Moratuwa, December 2019

This book describes the efforts that were made and initiatives carried out by Government organizations, the private sector and many individuals to enable the use of ICT in Sinhala and Tamil, from the beginning upto 2014. When English was the only language that could be used in computers, it left out many people in Sri Lanka, since many were not conversant in English. Enabling the use of ICT in Sinhala and Tamil was a necessity so that all could reap the benefits of ICT and this is the story of those who strove to ensure that this could be done.

Chapter 1

The need for Local Language Computing

We have striven to ensure that the benefits of ICT should be taken to people in Sri Lanka, most of whom would, if given a choice, prefer to use ICT in Sinhala or Tamil.

PROLOG

In most countries in the Asian region, such as in Korea and Thailand, people were using local languages for computing in the mid-1990s. In South Korea, even the user interfaces on the word processors were in Korean. But Kumara sat at his desk in the office complex at the BMICH¹ in Sri Lanka and struggled to type a letter in Sinhala. To do this, he had to use all the function keys, which was tedious. He could use the QWERTY keyboard and the output was displayed as Sinhala letters. But this system was difficult for him to figure out. He could not figure out how to type “Sanyaka” letters. He could not type in correct Sinhala, because he did not know how to type the letters න, ණ, ල, ඵ. It was a tedious process. Another problem that he encountered was that all staff kept coming to him to get their Sinhala letters typed.

Thereafter, the organization where he was working decided to organize an event. It was necessary to print a small booklet for the event written in English, but this booklet included Sinhala characters too. It was to be distributed to the participants at the event. The author insisted that the Sinhala characters had to be in the aesthetically pleasing Sinhala font “Bindumathie”. This font was designed and developed by Mr. Pushpananda Ekanayake. It took a while for Kumara to find the font. The author could not locate it in his computer, neither did he know how to send it to Kumara. This was in the early 1990s and email was not ubiquitous in Sri Lanka then. However, Kumara painstakingly printed the booklet with the Sinhala characters in it. The booklet was distributed to the participants and it well received.



In the meantime, all staff in the organization, including Kumara, were assigned new computers. Then there was another event to be organized and more copies of the booklet were needed for distribution once more. Kumara had

¹Bandaranaike Memorial International Conference Hall – a convention center in Colombo

installed the font Bindumathie in his old computer. Kumara did not understand that Sinhala was not yet included in the universal character encoding standard Unicode. In fact, he did not know anything about the Unicode standard. He did not even know that there was such a standard. He did not know that therefore he would need the font Bindumathie installed in his new computer. He happily opened the computer file which contained the booklet, which he had saved in a floppy disk. The English characters were displayed well. But what was supposed to be Sinhala was displayed as undecipherable symbols. So, once more he started searching for the elusive Bindumathie.

THE WAY IT USED TO BE

What Kumara encountered was a prevalent problem during this period, when trying to produce documents in Sinhala. No standard was used in the industry. A major problem was encountered when a person tried to use a document which had been created by another and produced using a different font or application. The font had to be sent to the recipient together with the Sinhala document, unless one knew that the recipient already had the font. Applications used their own fonts. Documents produced using one application could be accessed and used only through that application. This made the use of Sinhala email impractical, and slowed the use of Sinhala on the web. Functions such as sorting in Sinhala and Tamil were not standardized among applications. The reach of the Internet was said to be global. But most of the information contained in the Internet during this period was in the Latin character set, or in English. Trying to use the Internet in a language with a non-Latin character set such as Sinhala was an acutely frustrating experience. A person conversant only in Sinhala or Tamil could use the Internet only if local language support was enabled, with relevant local language content to keep the user interested.

It was therefore necessary to adapt Sri Lanka's computing environment to suit local needs. Standardization was necessary. Standardization was needed to enable the multitudes in Sri Lanka to use ICT in local languages and reap the benefits of the Internet. Standardization was needed to include the large swathes of people in the country who were left out and were unable to use the benefits of information and communication technology because they were not proficient in English.

Time was in Sri Lanka (when the country was "Ceylon") when schoolchildren in some schools were fined for speaking in Sinhala. This has changed. Sri Lankans may speak in any language in which they choose to speak. But there

had to be a choice too in selecting the language in which one wishes to use ICT. In Sri Lanka, all are not conversant in English. Therefore, users should have a choice – they should be able to select – Sinhala or Tamil for using IT, if they wish to do so. Learning English is a necessity, but people must have a choice in using one's own language to communicate, which includes using it in computing.

A case in point is when Ms. Prasadi Jasinghe, in early 2004, was trying to teach teachers who were mostly from rural areas, to use computers at the Faculty of Education at the University of Colombo. The teachers had been fearful of computers because of their perception that it was necessary to be conversant in English. Ms. Jasinghe overcame this issue at that time by assuring the teachers that they may use English letters to type Sinhala words. The fear subsided to a certain extent. But this was a cumbersome method. Users should be able to use Sinhala and Tamil naturally, without having to figure out what combination of English characters should be used to write a Sinhala word.

The organization in which Kumara was working purchased the package “Thibus”. The software was in a 3 ½ inch diskette which Kumara installed in his computer. The keyboard layout was aligned to the Wijesekera layout. It was easier now to work in Sinhala. Kumara could deftly use the Sinhala keyboard now. It was not tedious as before. Mr. Silva who was in the Administrative section wanted to type in Sinhala too. But only one person in the organization could use the software at a given time. If Mr. Silva were to use the software, then he had to insert the security dongle² into his computer, and during that time, Kumara could not use it. Mr. Silva therefore decided that it was easier for him to hand over all his work which required typing in Sinhala to Kumara. Kumara didn’t mind, but confusedly sought the font Bindumathie in the new software. But it was to no avail -only the Thibus fonts were available.

² Security dongle: denotes the product license and prevents illegal copying of the software

Chapter 2

Private Sector Initiatives

Many people and companies tried to get computers to work in Sinhala. Initiatives were carried out mostly from the early 1980s to introduce Sinhala language computing in Sri Lanka by the private sector. Innovative individuals too, tried through various methods to enable the use of Sinhala and Tamil in computers even though no standards were in place.

SINHALA IN WANG COMPUTERS

In 1977, the then Government adopted the open economy policy. Making use of this opportunity, Data Management Systems Ltd (DMS) was set up in association with Wang Laboratories USA. DMS was one of the pioneers in the sale of turnkey solutions in the late 1970s and in transferring new computing and IT technologies to Sri Lanka with the mission “Making Sri Lanka More Productive”.

Computers at this stage were being used only in English, and were used mainly in Colombo and its environs. But it was felt that the use of Sinhala and Tamil scripts was mandatory for these new technologies to reach the population in Sri Lanka.

This market need drove the Engineering team at DMS to commence localization projects with a view to developing appropriate software/ hardware products. The team specifically started trying to get Sinhala into Wang computers. There were at this time the Wang VS computer introduced in 1977, using virtual memory and the Wang 2200 which was introduced in 1973. These were two multi user computers in the mini-range.

At this time IBM PCs were available and Wang too came up with a PC. It was easier to get Sinhala into a PC computer because the PC had a font memory. Therefore, the first target was to access the font memory. Mr. Chandana

Weerasinghe from DMS and the team researched and found the addresses of the font memory and wrote a routine.

Thereafter, DMS discovered that there was a language initiative at Wang Global through which they were trying to have multiple languages, including Asian languages on Wang OIS (Office Information System) Word Processing systems. The challenges encountered in enabling Asian languages were due to the multi-level nature of the script; firstly, the team had to decide on internal storage, and then there was the issue of rendering Sinhala on the screen. There was also the issue of printing documents in Sinhala.

Wang Global had already developed a “Polyglot editor” which could handle multi-level languages. This had already been carried out for the Thai language. Therefore, the team at DMS contacted those who were involved in localization at Wang International.

Mr. Peter Gordon from Wang Labs Inc. led the localization efforts at that time. He started working on getting Sinhala into Wang computers, based on the initiatives carried out with regard to the Thai language. As stated earlier, the Wijesekera keyboard was widely used and popular at that time, especially in Government organizations. Therefore, the Wijesekera keyboard layout was used for input. The terminals at that time were character based. The Sinhala characters were using three lines. The DMS team continued working with Mr. Peter Gordon’s team. Mr. Chandana Weerasinghe at a later stage went to the US to work with Mr. Gordon’s team.

There were no Sri Lanka Standards (SLS) at that time with regard to ICT and local languages. The ICT language standards were developed later - first, by CINTEC and thereafter by CINTEC’s successor, the ICT Agency (ICTA) of Sri Lanka and these were standardized through the Sri Lanka Standards Institution. The DMS team developed their own standards for Sinhala fonts.

These needed 2-byte storage; the first byte was for the main character and the second was for the vowel modifiers³. Storage had been done in the same way for the Thai language as well. Finally, the product was available. The next step was to be a public launch of this product. The launch was to be on Monday, 9th December 1985. However, on the preceding Friday, DMS was informed via Courts that Mr. A.K. Kumarasena (who was an employee of Metropolitan) was claiming that the product to be launched was an intellectual property infringement of a product he had patented. Despite this, the product launch took place at the Oberoi Ballroom as planned, with DMS counter-suing Mr. A.K. Kumarasena. DMS, in the meantime, also filed for a patent, which was granted later. Unfortunately, the product could not be made available commercially. This was the first case that was contested under the new Code of Intellectual Property Act no. 52 of 1979 as D.C. Colombo case no. 2513/SPL. This case dragged on for a year or more and subsequently there was an out of Courts settlement and both parties mutually withdrew.

However, during this time, Mr. Lal Chandranath, Mr. Ravi Wickremasinghe and Mr. Krishantha Jayaratne of DMS enabled the use of Sinhala for the Commodore PC which was IBM compatible. This was done through a modified version of a VGA card which allowed for raster “interlacing” to allow for enough resolution to support the multi-level nature of scripts. A hardware key (for intellectual property protection) which had to be fixed on a parallel port and two sets of stickers



DMS launching the OIS local language word processing system

³ Vowel modifiers: glyphs on either side or on top or the bottom of a base character which modify the relevant base character.

with the Sinhala and Tamil scripts for users to stick on the keyboard were provided. This product was made available in 1986. The products were named “Mau Basa” (මව් බස) for the Sinhala version and “Thai Mozhi” (தாய் மொழி) for the Tamil version. This package was marketed commercially, mainly to law firms. This could be accessed from other applications, such as databases.

THE MAHANAMA SINHALA KEYBOARD

The Weatherby Training Center, in Sri Lanka, brought used Sinclair machines in 1980 from the UK to Sri Lanka. These were sold to the Department of Education, to be given to schools. But a key impediment in Sri Lanka was the fact that most teachers and the students were not conversant in English. So, Mr. Jayasinghe, who was the owner of the Weatherby Training Center, developed a hardware dongle to change the character map into Sinhala. But the expense with regard to this hardware system was far too high.



Mr. Diliup
Gabadamudalige
explaining a point
to Mr. Saputantri
Mahanama

In 1984, Mr. Saputantri Mahanama joined the Weatherby Training Center as a Programmer. Mr. Jayasinghe then requested Mr. Mahanama to develop a less expensive solution. In response to this request Mr. Mahanama developed two ASCII based keyboard input methods; one was based on using the QWERTY keyboard through which, for example typing the letter “A” would yield the Sinhala letter Ayanna (අ) and typing the letter “B” would yield the Sinhala letter Bayanna (බ).

The other keyboard input system was based on the widely used “Wijesekera” keyboard layout. Sinhala typists were extensively using the Government approved Wijesekera keyboard. A key feature of the keyboard input method which Mr. Mahanama developed was that only one key was necessary to input all the “ispillas”, whatever shape the ispillla may take, such as ේ and ේ and similarly only a single key was necessary to input all the “papillas” whatever shape the papilla may take such as ඩ and ඩ. This was in 1984. Mr. Mahanama obtained a patent for this keyboard input system.

Later, in 1986, Mr. Mahanama was introduced to Mr. Jagath Ranawaka of Jagath Robotics Pvt Ltd (JRL). Mr. Ranawaka had commenced marketing IBM compatible computers. Mr. Jagath Ranawaka asked Mr. Mahanama whether he could develop a Sinhala word processor for these IBM compatible computers. Mr. Mahanama managed to develop a Sinhala word processor for IBM compatible computers in 1 ½ months.

In 1986, JRL commenced marketing the Sinhala language word processor developed by Mr. Saputantri Mahanama who had the patent for the “Mahanama” Sinhala keyboard input method. This Sinhala word processor was very popular among many printing organizations in Sri Lanka. This was used by the former President Mahinda Rajapaksa too, when he was the Leader of the Opposition in 2002. When Mr. Rajapaksa was the Prime Minister, JRL offered him a free copy of this Sinhala word processor. Subsequently, JRL managed to extensively market computers with this Sinhala word processing product.

In the meantime, Bhikkhu Mettavihari had started computerizing the Tripitaka (ත්‍රිපිටකය) in the Pali language using the Sinhala script, sponsored by the Sasana Sevaka Society of Maharagama. In 1992, Bhikkhu Mettavihari met Mr. Jagath Ranawaka who introduced him to Mr. Mahanama. Consequently, from 1992 onwards, the Bhikkhu used the Mahanama Sinhala word processor. With this he could type Sinhala conjunct letters (බැඳි අකුරු).

Thereafter Mr. Mahanama developed a program for the Tripitaka through which it was possible to transliterate Pali written in the Sinhala script to Pali in the Burmese script, and to Pali in the Thai script. This development occurred from 1992 to 1995. The output was made available on www.metta.lk. Mr. Mahanama also developed a program through which it was possible to search and find any word in the entire Tripitaka in the Sinhala script.

Patent for the
Mahanama Sinhala
keyboard

Home	Logout
Edit	
Workflow	
New patent	
Changes log	
Validate	
View POA	
View Assignments	
Similar Owner	
View patent	
Record 1/1	
Application type	National Patent
(219) Application Nbr	P/09845
Reception date	14/12/1987
(220) Filing date	14/12/1987
Novelty date	14/12/1987
Status	(P) Patent Expired
(111) Registration Nbr	P/0001/009845
Registration date	11/03/1988
Entitlement date	14/12/1987
(141) Expiration date	14/12/2007
(71) Applicant	. Wether by Training Services (Lanka) Limited, 8, Balahenmulla Road,Colombo 6, [LK];

Another project was developed on a request made by the late President Premadasa, for the Buddha Sasana Ministry. This was a database system in Sinhala of all Buddhist monks who sat for the Pracheena exam. These monks were given Rs. 100/- per month and through this database it was possible to find out the amounts given.

Mr. Mahanama also developed a searchable Sinhala database for the CD library of the Sri Lanka Broadcasting Corporation (SLBC). The fields in the database included the name of the song, the singer's name etc. SLBC could instantly retrieve the songs that were requested under their "Request Programs" through this system. This was developed using the Assembler language.

Yet another Sinhala program that Mr. Mahanama developed was for the Sri Lanka Navy on the request of the Navy Commander. The Navy Commander wanted to communicate securely with fourteen Naval bases. Mr. Mahanama developed a scrambling password system in Sinhala so that the Navy passwords were secure. The Sinhala passwords were scrambled through this system. Mr. Saputantri Mahanama, over the years, developed many diverse products in Sinhala. He passed away in July 2019.

THIBUS

More than thirty years ago, in 1986, Mr. Rohan Manamudali expected to enter the University the following year, after sitting for his Advanced Level exams. But during this period the country was in turmoil. Universities were closed with no opening date in sight. However, four young men - Mr. Sampath Godamunne, Mr. Rohan Manamudali, Mr. Dhammika Wijeratne and Mr. Vipula Jayampathi - past pupils of Harischandra Vidyalaya in Negombo made good use of their time. They achieved milestones in the local language arena, while other young people who were selected to Universities were languishing, uncertain of the future.

This is how it started. Mr. Rohan Manamudali first saw a computer, a Sinclair computer, as a schoolboy in 1984. But the students at his school were not allowed to even touch the computer. Computers in most schools at this time were sacrosanct and used to be kept locked up carefully. And there were no teachers who were able to teach computing. Students had no access to the computer.

But this did not deter Mr. Manamudali. He was curious about computers. His former classmate Mr. Dhammika Wijeratne had an Apple computer. They had no teachers, but there were books from which they could study the hardware aspects. Mr. Dhammika Wijeratne was conversant in electronics. These two collected discarded parts of old computers and they managed to assemble a computer which actually worked.

The computer which they assembled was IBM compatible. Then they met Mr. Sampath Godamunne who had an Atari computer. Mr. Godamunne wanted to write letters in Sinhala. Mr. Godamunne tried to write his letters in Sinhala with the Atari computer, since he thought his handwriting was not good enough for writing letters. He wanted his letters in Sinhala to be clearly decipherable. When Mr. Godamunne came over and met Mr. Manamudali and Mr. Wijeratne, they decided to see if they could help Mr. Godamunne write his letters using the IBM computer.

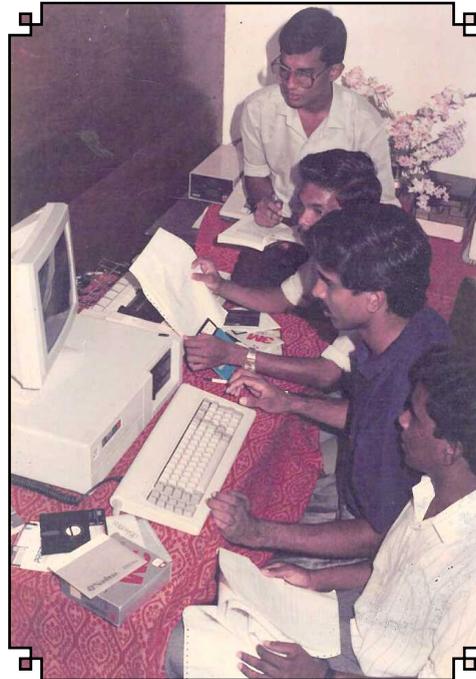
They did not know then that this was the beginning of a long journey in localization. This was the commencement of their localization initiative which went a long way, much further than they ever imagined it would.

The team also got to know of the localization initiatives being carried out by the Companies DMS and Metropolitan. The popular English word processor during this period was WordStar. This team developed a Sinhala and English word processor, “Super 77”, using graphic cards. The next hurdle was computer

input. The team first developed an input method which they could use easily, which they believe may have been one of the first phonetic Sinhala input systems in Sri Lanka. The Wijesekera keyboard layout was popular for typewriting mostly among Government employees.

The team realized that if they were to replace typewriters with computers and word processors, then an input method based on the popular Wijesekera keyboard layout would have to be developed. They used physical keyboards with the English QWERTY layout for their initiative.

The team presented the Sinhala word processor which they developed to Prof. Cyril Ponnampereuma, who was then the Director of the Institute of Fundamental Studies, in Kandy. Prof. Ponnampereuma requested them to develop the package further, with the inclusion of Tamil. This, they achieved within two months. They used the popular keyboard layout Renganathan for Tamil. Key features of this package were that the language could be changed from one to another easily, with just one key. The QWERTY physical keyboard could be used for all three languages. By running a program that was provided, it was possible to switch to the popular Sinhala Wijesekera keyboard. The package

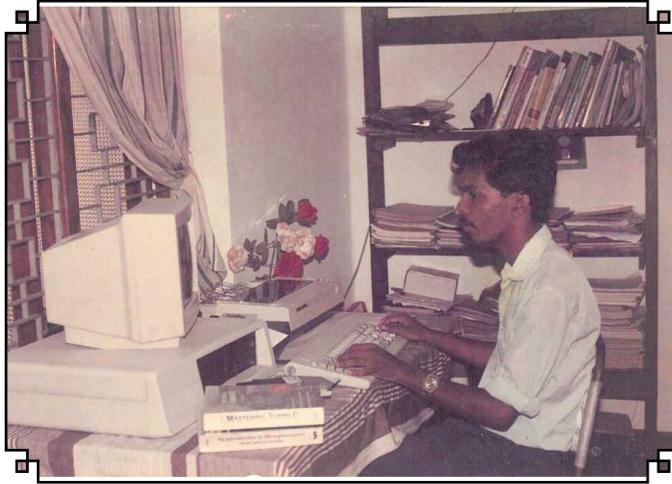


1989, Development of Super 77 -Thibus

(From Left)
Mr. Dhammika Wijeratne
Mr. Rohan Manamudali,
Mr. Sampath Godamunne
& Mr. Vipula Jayampathy

could be used on IBM AT/XT computers and compatibles with operating systems MS-DOS or PC-DOS and at least 256 K RAM and a graphics card.

1989, Super 77,
Preliminary
version of Thibus



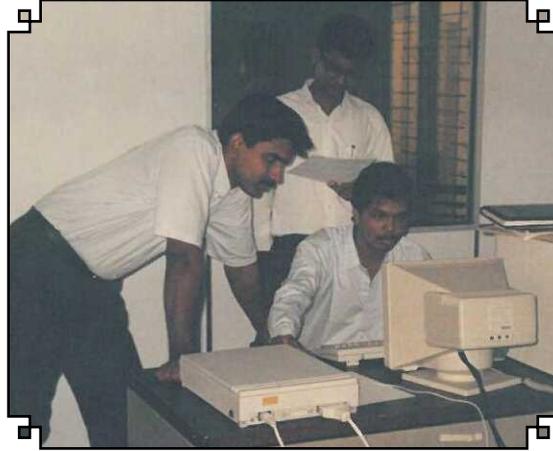
The team's primary objective at this point in time was to deploy it among users, and to make people use their product. They were very young and so they dreamt of all typewriters being replaced by computers

with their product. Therefore, in order to popularize their product, they distributed the package to Government organizations free of charge and then carried out comprehensive awareness programs and demonstrations.

They distributed the package throughout the country, even bearing the cost of travel. If some day someone was to give them a copy of their package, they decided that this would mean that the product's use had truly spread.

This happened one day at the University. Mr. Rohan Manamudali was now an undergraduate at the University of Moratuwa. A friend came over and recommended the word processor that Mr. Manamudali and his team had developed and encouraged Mr. Manamudali to try to use it. When this happened, the team was really pleased and decided that deployment and use of Super 77 had truly spread.

Later, they gave a copy to Dr. Romesh Dias Bandaranaike, who was the Chairman of the Sri Lanka Business Development Center (SLBDC). Dr. Bandaranaike's advice was that it should not be given free but should be sold to SLBDC. Their product was valued at LKR 17,500/- which was then a large sum of money. Dr. Bandaranaike contacted the company Dynamic Ram (Private) Ltd, and requested the team to develop a business model with them. Dynamic Ram (Private) Ltd became the sole authorized distributor for the product which was commercialized and named "Thibus". This was how "Thibus" commenced. Thibus became a household word and became hugely popular.



Working on Thibus

This team had a partnership "Science Land", and they were joined by another member, Mr. Vipula Jayampathy. After the team members graduated from the



1st September 1994, Start of Science Land (From left) Mr. Dhammika Wijeratne, Mr. Rizvy (first employee who marketed Thibus), Mr. Vipula Jayampathy, Mr. Rohan Manamudali & Mr. Sampath Godamunne

Universities, the partnership Science Land was converted into a company, Science Land (Pvt) Ltd. This, they believe, is the very first company

established solely to address the area of ICT and Local Languages.

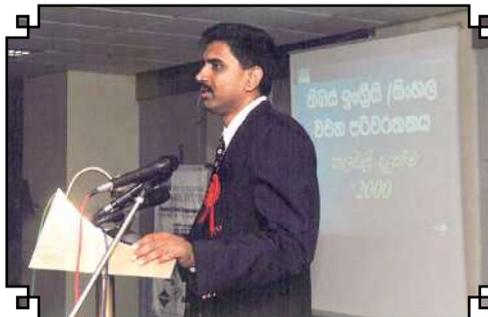
With the introduction of Windows 3.1, a keyboard driver and a font named “Thibus” were also developed. In 1997 new features were added – these were a Sinhala spell checker and a system for sorting in Sinhala according to the Sinhala alphabetical order.

Mr. Rohan Manamudali relates another interesting development. For the Provincial Council Elections of 1999, candidates were to be given numbers according to the alphabetical order of their names. Consequently, candidates purchased the Thibus package, used the sorting feature in Thibus, and found out what their number would be. The candidates then published their numbers before the numbers were issued by the Department of Elections. This meant increased business for Science Land.



September 2000, Nawuyugaya Magazine

Mr. Sampath Godamunne at the launch of Thibus word translator, 2000



By this time, many Government organizations had computers, but in many organizations, staff were not aware of what they could do with the computers. The former Government apex body on ICT, the Computer and Information Technology Agency of Sri Lanka (CINTEC) was also carrying out

awareness programs for staff of Government organizations. The CINTEC mobile computer unit, which was a bus in which the seats were removed and computers installed, visited rural areas to create awareness on computers. Thus, awareness on what computers could do was being built up throughout the country. Meanwhile the Science Land team also visited rural areas to create awareness and teach people how to use the Thibus package. It is estimated that the typewriter was replaced by computers with the Thibus package in about 80% of Government organizations.



Thibus staff at the word translator launch, 2000

In the year 2000, another milestone was reached when Thibus created a Sinhala dictionary. At this time other companies too were entering the Local Languages arena. Specifically,

from 2004 onwards, the Information and Communication Technology Agency (ICTA) of Sri Lanka, CINTEC's successor, promoted adherence to the universal encoding standard Unicode. In line with this, the Thibus software was made Unicode compatible.

A milestone reached in 2006 was the machine translator which was launched by the then President of Sri Lanka. Sinhala text messages for mobiles were enabled in 2006. In 2009, Science Land developed a six-way (for Sinhala, Tamil and English) transliteration system for ICTA. ICTA needed this for

implementation in tri-lingual Government databases being developed under its Re-engineering Government program.

May 2006,
launch of the
Sinhala - Tamil
translator



In 2009, Science Land, in a joint venture with the University of Colombo School of Computing (UCSC) contracted with ICTA in implementing its Language Interface Pack (LIP) project. This LIP project was carried out with Microsoft Corporation. The objective was to develop Sinhala user interfaces for

Windows Vista and Office 2007, and later for Windows 7 and Office 2010. A comprehensive glossary and a Style-guide (which denoted the language style inclusive of locale) were developed under this project. A “Polichcek” which included a comprehensive list of words and phrases in Sinhala and in English which should not be included in Windows and MS Office was reviewed by the LIP team in Sri Lanka and feedback given to Microsoft.

May 2006,
launch of the
Sinhala - Tamil
translator



KAPUTA FONT AND KAPUTA.COM

Mr. Niranjan Meegamma started his career at the Survey Department. He worked initially on the Upper Kotmale project and then joined the Samanala Wewa Project. He had a leaning towards technology and learnt programming from a handheld programmable calculator, an HP 41CX. Then he noticed an advertisement for a vacancy at the project NARA GTZ (National Aquatic Resources Research and Development Agency – German Technical Cooperation) to which he applied. He was recruited as a hydrographic surveyor. The project lead was Mr. Michael Gruber. Mr. Meegamma refers to him as his “Guru”. Mr. Niranjan Meegamma was now working at the National Hydrographic Office (NHO) at NARA, in which he set up a LAN. There Mr. Meegamma wrote a program in QuickBASIC with 30,000 lines to transition navigational maps to AutoCAD.

Then came the issue on language; Several Sri Lankan fishermen were taken into custody near the coast of the Maldives. Sri Lanka’s Fisheries Minister had been informed that the fishermen had not been able to read the navigational maps because these were in English. Around this time, Sri Lanka was heavily dependent on the British Admiralty charts of the Sri Lankan waters. The core issue in this case was that the fishermen were not conversant in English. The Minister had emphasized to the Head of NHO that the maps had to be in Sinhala.

Mr. Niranjan Meegamma was tasked with the conversion of the nautical maps to Sinhala. Therefore, he bought transparent sheets of Sinhala letters, developed by Mr. Colombage, scanned these, digitized and placed the letters in AutoCAD. These developments happened around 1992. Windows version 3.0 was available. Then one morning Mr. Michael Gruber called over Mr. Meegamma. Mr. Gruber was typing the letter අ (Ayanna) – which was the only Sinhala letter that Mr. Gruber knew.

This was a breakthrough. He had used the vector letter they had created and made a True-type font with Corel Draw. Resulting from this, they developed the Sinhala font “Kandy”. For the input system, the letter ඞ was mapped to A, the letter ඞ was mapped to B, etc. Using this font, Mr. Meegammana then wrote and published a book on hydrography in Sinhala. NARA GTZ then purchased the font editor Fontographer. Using Fontographer, Mr. Meegammana developed the font Kandy so that it included commonly used Sinhala letters and he developed the font Kandy-supplement which included letters not used frequently. Now it was possible to type direct on maps in Sinhala. It was also possible to use the fonts on WordPerfect. Mr. Meegammana also developed other fonts such as the Sinhala fonts Anuradhapura, Matara etc., and the Tamil font Jaffna.

Mr. Niranjan Meegammana then left NARA GTZ and formed a partnership, Digital Research, with his brother. Now his focus was exclusively on the development of ICT in local languages. A Sinhala input method editor was developed in 1995. He also developed a Sinhala encyclopedia with Visual Basic. It was possible when using this encyclopedia to easily switch from Sinhala to English and vice versa. Thereafter, Mr. Meegammana joined Ceylon Global Communications as Webmaster. Here, he published Sinhala content on the Internet using ASCII fonts. He developed an interactive CD comprising information on Sri Lanka in Sinhala and English and developed a website “Taru mal yaya” (තරු මල් යාය) and a Sinhala music database consisting of Sinhala songs, in 1997. He also developed a tri-lingual word processor for Windows. Consequently, using this product, the Sinhala newspaper Lankadeepa and the Tamil newspaper Veerakesari were published on the Internet.

Kaputa.com

Kaputa.com was developed in 2000. This portal became instantly well-known. The Sinhala ASCII font Kaputa also became well-known. This was initiated when a group of friends got together and developed an application for Sinhala emails. Mr. Meegammana, in the meantime had set up the start-up Company e-Fusion (Pvt) Ltd. The community portal Kaputa.com became hugely popular. Links were given from this portal to content on various subjects. Earning was through Google advertisements. The Sinhala font Kaputa was used in the online newspapers Divaina and Ravaya, and also in the website of the Ministry of Public Administration.



The advent of the standard Unicode

The ICT Agency (ICTA) of Sri Lanka commenced operations in July 2003. From the start, ICTA's Local Language Initiative (LLI) commenced working on Unicode compliant Sinhala fonts. The Local Language Initiative at ICTA was headed by Prof. Gihan Dias who invited Mr. Niranjan Meegammana to attend a workshop on the Unicode standard. This



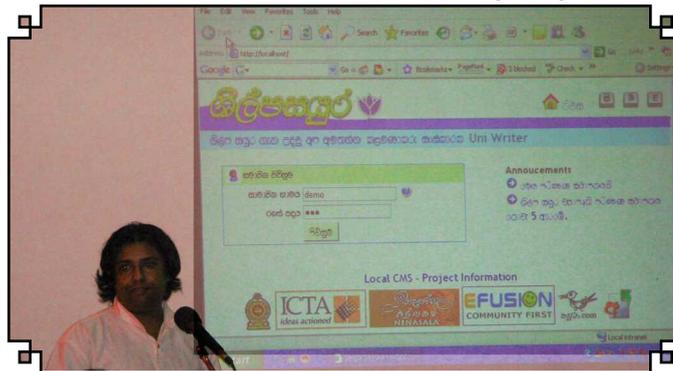
Workshop on the Unicode Standard

workshop was conducted by Mr. Muthu Nedumaran of Malaysia, with expertise in designing Unicode fonts and developing input methods. From this workshop, Mr. Meegammana garnered knowledge on the standard Unicode. He started working on Unicode compliant Sinhala fonts, and developed a Unicode compliant version of the font Kaputa.

Mr. Meegammana then did something which greatly assisted the work being carried out by ICTA. He started in 2005 a Sinhala Unicode Google Group and invited others to join this. This was a community initiative and brought together people who were involved in developing Unicode Sinhala. The Group communicated only in Unicode Sinhala. Technical support was given to those who needed support. Mr. Anuradha Ratnaweera addressed issues on Unix, Linux etc. This Group eventually became a very active forum, communicating entirely in Unicode Sinhala, thus refuting the allegations made against the standard by several detractors, who faded away, one by one with time.

Shilpa Sayura

This initiative commenced in 2006. Mr. Niranjan Meegammana was instrumental in developing Shilpa Sayura, the e-learning system, aligned to the National education curriculum, which was developed through a grant from ICTA's e-Society program. The digital content in Sinhala, Tamil and English was developed by school teachers. This was deployed through the Nenasala centers



Mr. Niranjan Meegammana making a presentation on local languages

which ICTA set up throughout the island, and through rural schools. Shilpa Sayura was of great assistance to children in rural schools where there was a dearth of educational resources. Shilpa Salyura had a direct impact on alleviating school dropouts and improving examination results. This project received awards such as the Stockholm Challenge 2008, e-India 2009, and the Best Corporate Citizen CSR awards in 2008 and 2009 and many others. A mobile version of Shilpa Sayura is presently being developed.

MICROIMAGE, HELAWADANA, SHAN AND LOCALIZATION

When Mr. Sinnathambi Shanmugarajah (Shan) joined Microimage in 1998, his first assignment was to develop a keyboard input system. Microimage had been incorporated as Microimage (Private) Limited in the mid-1990s. It had started as a student club led by Mr. Harsha Purasinghe. By the time Mr. Shanmugarajah joined Microimage, the Microimage Sinhala package Helawadana was already well known and popular among users.

There were extensive discussions on Sinhala keyboard input. Mr. Shanmugarajah remembers struggling and trying hard to get the result “ඒ” when pressing the keys with the English letters “J” and “S”. Later Mr. Shanmugarajah found a low-level keyboard tool. Consequently, Sinhala keyboard input systems in accordance with the Wijesekera layout and another based on a phonetic system were developed. The phonetic system was sound based. For example, typing the letter “B” yielded the pure consonant “බ”. In the late 1990s, Microimage launched the package Helawadana, which now included these two Sinhala keyboard input systems and six Sinhala fonts. The Sinhala fonts had been developed by the founder of Microimage, Mr. Harsha Purasinghe and Mr. Dasun Gunaratne etc. The fonts included MI Harsha, MI Dasuni, and the font Ridma.

The Helawadana package could be used for word processing, accounting, and desktop publishing and included a comprehensive spell checker. It also included a tutorial through which users could be guided on learning how to use the Sinhala keyboard. Later “Aricchawadu” was added to the Helawadana package. This included a Tamil keyboard input system based on the Tamil keyboard layout aligned to the “Renganathan” layout and two Tamil fonts. These were ASCII based systems, prior to Unicode Sinhala implementation. Helawadana was an instant success with tremendous sales. In 2003 Helawadana was re-launched with added features which comprised Sinhala games and dynamic font converters.

ICT IN LOCAL LANGUAGES AT ANCL

Computer technology was brought to Associated Newspapers of Ceylon Ltd (ANCL) - known as Lakehouse – in the mid-1980s. During this period only proprietary systems were used. The printing industry was using typesetting machines to print the output onto photo-sensitive paper (known as bromide paper) and developed in a chemical bath. At the final stages, pages of the newspaper were made up on a desk by pasting the bromide papers onto a dummy page. Thereafter, a photo image was taken and from the photo image, the plate image which would be mounted on the offset press, was produced.

ANCL then sought a system which was capable of typesetting in all three languages, Sinhala, Tamil and English. The selected system was Amicus, the typesetting and classified advertisement system from the UK. Convincing people who were mostly familiar with the Latin alphabet to facilitate printing with the full character sets of Sinhala and Tamil, was difficult. Sinhala characters have modifiers which are displayed before, after, on top and the bottom of the actual base character. In addition, combined characters were used frequently. In this case, when two characters are combined the resultant character is of a different shape, especially with regard to the

consonants ජ (ra) and ය (ya). With old Sinhala writing styles, conjunct and touching letters, and the rakaransya, rephaya and the yansaya were frequently used.

The team at ANCL believed that language should not be changed to adjust to technology. Therefore, the team leader, Mr. Anura Tissera, who was the Senior Electronic Engineer at ANCL, (and later the Chief Innovation Officer) contracted with Amicus.

This resulted in changes carried out to the Amicus system to suit ANCL's requirements. Aesthetic features of letters were also considered. The typesetter output was a bitmap of composite characters. This may have been the first time a newspaper in a local language was published using a computer system.

The story of how the system was implemented is also interesting; there had been a strike by letter block making operators who were concerned that automation may be a risk to their jobs. To overcome this issue, Mr. R. Bodinagoda, who was then Chairman of ANCL asked Mr. Anura Tissera whether the newspapers could be printed through the typesetting system. This was carried out and ANCL consequently managed to publish a newspaper through a typesetting system for the first time in Sri Lanka.

ANCL had two Amicus systems with several workstations. One system was for the Editorial section and the other for the Advertising section. Through this it was possible to sort the classified advertisements in the alphabetical order in Sinhala for the first time in Sri Lanka. The input methods were based on the Wijesekera keyboard layout for Sinhala and the Renganathan keyboard layout for Tamil.

In late 1990s printing was converted to open systems. The typesetters were to be replaced by imagesetters. These could directly print the negative image of a page and with that the plate could be produced. The systems were to be changed smoothly. First, open system support for local languages was made available at ANCL. The Company Microimage Pvt Ltd developed the keyboard driver and renowned font developer Mr. Pushpananda Ekanayake designed a set of fonts. The keyboard driver and the fonts were specific to ANCL as there were no standards at that time.

The next task was to migrate to open system software. QuarkXPress was the prevailing desktop publishing software. It used ASCII character codes. Firstly, the editorial system was set with open systems. The formatting tags necessary for QuarkXPress were introduced through software developed in-house.

Newspaper columns are narrow and lengthy. Therefore, for these, text has to be fully justified. This meant that it was necessary to break words at the end of the line. This feature was not available for local languages in word processors. Ms. Dineesha Ediriweera of ANCL, analyzing Sinhala words under the guidance of Mr. Anura Tissera, developed a system to break the words at the most suitable position. This was probably the first-time hyphens were introduced in a system for a local language. Then the pages were made on QuarkXPress and sent to the imagesetter via an imagesetter software.

While the editorial pages were moved to open systems and imagesetter output, advertising was still being carried out by using the Amicus system. Advertising has two categories; classified advertisements and display (artwork) advertisements. The display advertisements were moved so that

map exclusively designed for ANCL was used. The remaining work to be done, was to develop a classified system for any publication in any of the three languages.

In early 2000s, an in-house system was developed to work with a SQL server. SQL server did not support local languages or the collation order but 2-byte characters could be used for Sinhala and Tamil text. The sorting order is important in a classified system. Hence, the sort string was built up from the first two words of the text in the advertisement and used in listing the classified advertisements.

Later, there were discussions and meetings held on the need for Sinhala and Tamil to be compliant to the Unicode standard, first at CINTEC and later at CINTEC's successor, ICTA. There were extensive discussions on the Sinhala keyboard layout, which had to be standardized. Mr. Anura Tissera represented ANCL initially at these discussions. Later, ANCL was represented by Ms. Dineesha Ediriweera at meetings of ICTA's Local Language Working Group.

The Sinhala Unicode font DinaminaUniweb was also developed by ANCL. With this font it was possible to prove that all Sinhala characters could be rendered properly in compliance to the Sinhala Unicode standard. ANCL gave the font to ICTA to be offered free to users.

At present, every aspect of work at ANCL is Unicode compliant, including publishing. But an impediment encountered is that Adobe software does not fully support Unicode in Sinhala (viz. version CS5 did not support at the time of purchase) therefore a third-party software company made a template to be used with InDesign for ANCL.

SETT BROWSER, HELAKURU, AKURU HURUWA

Mr. Dhanika Perera was an Advanced Level student at Holy Cross College, Kalutara when he first developed software in Sinhala. The software he developed was an encyclopedia in Sinhala, Manavaya (මනවයා), on the human physiological system. In 2007, he received the Young Computer Scientist Silver Award from the Federation of IT Industry (FITIS) and the Sri Lanka Association for the Software Industry (SLASI), for this project. This is what triggered Mr. Dhanika Perera's interest in and passion for software development, in particular for software development in the arena of local languages.

Mr. Dhanika Perera was selected to the University of Moratuwa. In 2011, during his third year at the University, there was a requirement for students to implement an individual project. Due to his deep interest in the area of local languages, he naturally wanted to do a project involving Sinhala software development. During this period, Sinhala character rendering support was not ubiquitously available on mobile devices. This was a problem for Mr. Dhanika Perera, as he was writing a blog in Sinhala, but it was not possible to read the blog on mobile devices due to this issue.

Therefore, he wanted to develop a solution for this issue as his University project. Consequently, he developed SETT browser, a web browser for mobile devices with application level support to render Sinhala characters for Unicode compliant Sinhala websites. This product worked on any Android phone available in the market at that time.

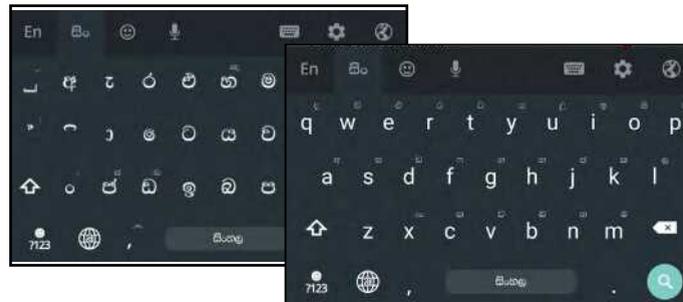
Mr. Dhanika Perera did not give this up after his University project evaluation - he applied to the mBillionth Award South Asia. The SETT browser for Android was a winner under the category m-inclusion. At this competition, Etisalat, the mobile telecommunication provider in Sri Lanka was also a winner. Etisalat was impressed with Mr. Dhanika Perera's product and realized its potential.

Etisalat consequently made a business proposition to Mr. Perera; Mr. Dhanika Perera's product would be made available to Etisalat customers. Mr. Perera agreed and consequently, each time someone downloaded the product, Mr. Perera received revenue. This inspired him to start a company of his own.

Mr. Dhanika Perera was yet a third-year student at the University of Moratuwa. He was innovative. With two of his University batch-mates he established the company Bhasha Lanka (Pvt) Ltd. There was no physical office. But they were profitable from the first day onwards, because of the venture with Etisalat. The team re-invested the profits and developed more products. One such product was Helakuru, which ultimately became well-known and widely used.

The product Helakuru consisted of a Sinhala input method for Android and iOS mobile devices. It enabled typing in Sinhala on mobile devices. Helakuru gave the capability to input Sinhala according to the standard Wijesekera layout or through a phonetic layout. The user could select the preferred layout.

Helakuru Mobile Keyboard
- Smart Wijesekera Layout

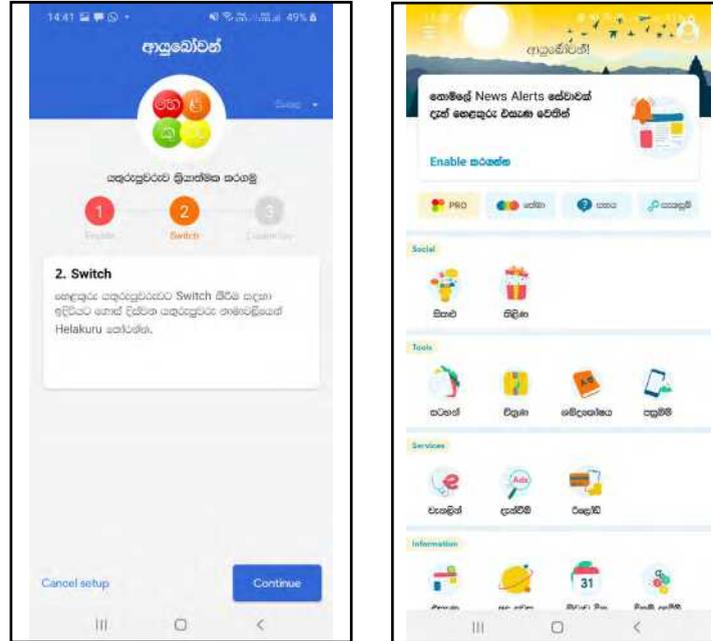


Helakuru Mobile
Keyboard - Phonetic Layout

Mr. Dhanika Perera then applied to Spirallation, the Tech Startup Program, implemented by the ICT Investment and Private Sector Development Program of the ICT Agency (ICTA) of Sri Lanka. Mr. Perera's company did not receive the subsidy under the Spirallation Program, but received mentoring, capacity development, and relevant training through workshops. Mr. Perera also found

the opportunity to network with others in the industry, which he found greatly beneficial.

Helakuru Super App -
Sinhala Digital Services



Helakuru had, by now, over five million downloads and about two million active users, and was one of the most downloaded mobile apps in Sri Lanka. Today it empowers millions of Sri Lankans to write in Sinhala digitally. The Bhasha team then introduced Helakuru Pro, which included features for personalization and custom dictionaries, for a monthly subscription. The team scaled up from three employees to twenty employees within a few years and consequently, there were more innovative software products. A desktop version of Helakuru, which worked on Windows, Linux and Mac operating systems was also developed. This enabled Sinhala to be typed phonetically. But

the company's core competency and focus were in the arena of mobile devices. The team added Tamil input capability to Helakuru in 2019. At present, the team is developing Helakuru into a Super App intending to make it similar to WeChat in China, offering more digital services in local languages to the two million users of Helakuru.

Bhasha has also ventured into teaching small children to read and write in Sinhala through mobile devices, through their brand Hapan (හපන්). Under this brand, the product Akuru Huruwa (අකුරු හුරුව) consists of gamified educational content targeted to small children to assist them in learning to write Sinhala letters. The product Hodi Potha (හෝඬි පොත) consists of similar content for learning to read Sinhala letters and words. The team is developing the brand Hapan as a digital education platform with local language educational content.

Kumara was now the de facto Sinhala typist although his designation was Accounts Assistant. People kept plying him with more and more work which included Sinhala keyboard input. But Kumara was unfazed. He stayed on after the office closed and kept working until nightfall. But then his Chairman handed him a large printed document which he wanted computerized. This resulted in an unpredictably sudden reaction in Kumara who felt he just could not do this. He had read about optical character recognition (OCR) systems. He knew that research on an Optical Character Recognition system for Sinhala was being carried out at the University of Colombo. He went over to the University but there was no OCR product which was working at that time.

Chapter 3

Universities and CINTEC

State sector initiatives in local language computing commenced in the mid-1980s when the Computer and Information Technology Council of Sri Lanka (CINTEC) was established. Institutions under State Universities such as the Institute of Computer Technology which was later restructured as the University of Colombo School of Computing (UCSC) commenced work on Sinhala language computing in the mid-eighties.

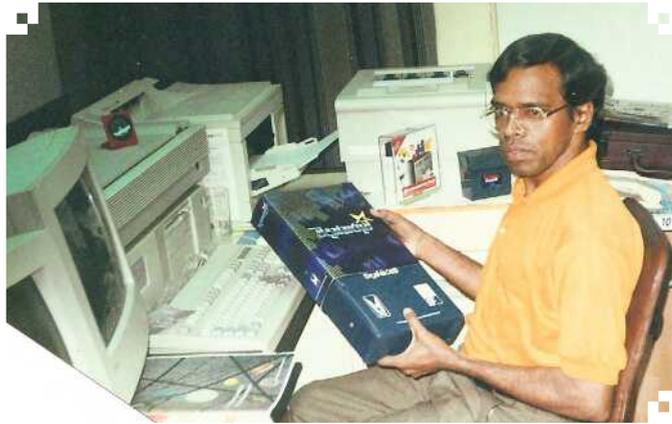
FONT SARASAVI AND WADAN THARUWA

Dr. Kevin Seneviratne was a Senior Lecturer at the Department of Statistics and Computer Science at the University of Colombo. He thought of developing Sinhala letters using a BBC computer in 1983. Staff members therefore created a set of Sinhala characters and the program parade of ITN TV was displayed in Sinhala.

Around this time, IBM released the personal computers IBM PC XT and IBM PC AT. Mr. S.T. Nandasara, of the Institute of Computer Technology of the University of Colombo, on the instructions of Prof. V.K. Samaranayake, commenced the development of ICT in Sinhala. Specifically, he reprogrammed the IBM PC's Basic Input / Output system (BIOS). His intention was to convert the display from English to Sinhala. Dr. Thaweesak Koanantakool from the University of Thammasat in Thailand was involved in developing the use of ICT in the Thai language. Mr. Nandasara therefore studied the development of ICT and local languages at the University of Thammasat with the intention of carrying out similar development with regard to Sinhala. He learnt how the Thai language was used on a desktop IBM PC XT, returned to Sri Lanka and commenced development. One issue was that the IBM PC XT did not have sufficient resolution for Sinhala. IBM released the video graphic array card, VGA, with very high resolution. Thereafter it was possible to get Sinhala characters displayed perfectly. He worked together with the University of

Thammasat (although he was based in Sri Lanka) and managed to reprogram the entire VGA card to handle both Sinhala and Tamil. This happened in 1989.

He was able to get the first basic input / output system for Sinhala and Tamil where it was possible to type in Sinhala on the command prompt and change the language in the computer using the combination of the keys control + shift. There was no Alt key on the keyboard at that time. (At present, the keys Alt+shift are used to change the language in the computer). The operating system was reprogrammed for using Sinhala and Tamil and it was named S-BIOS for Sinhala and T-BIOS for Tamil. Thereafter Mr. Nandasara developed the tri-lingual word processor “Wadan Tharuwa” (වදන් තරුව) using the C programming language. The name was given by Prof. Samaranayake. This product was marketed by companies such as John Keells Holdings, CBA Pvt. Ltd and Fentons Computers etc.



Mr. S.T. Nandasara

In 1994, Mr. Nandasara developed the typesetting software “Athwela” (අත්වැල) with trilingual support (Sinhala, Tamil, and English) for Ventura Publisher, an early desktop publishing package. This product was marketed by the Computing Services Center of the Institute of Computer Technology, University of Colombo. With this product, it was possible to create complex documents consisting of all three languages.

Later, in 1995, Mr. Nandasara developed the “Sarasavi” (සරසවි) packages for Windows 95/98. Mr. Nandasara was also instrumental in the development of the tri-lingual national web window www.lk which was launched on 15th September 1996.



COLOMBO DISTRICT
DEHIWALA MOUNT LAVINIA MUNICIPAL COUNCIL

ප්‍රතිචාරකයා	ඡන්ද	ප්‍රතිශතය
එ.ජ.ප. පී.එ.ආ. PA	37 481	46.37%
එ.ජ.ප. පී.එ.ආ. UWP	30 791	38.09%
එ.ජ.ප. පී.එ.ආ. JVP	4 622	5.72%
එ.ජ.ප. පී.එ.ආ. NDI 1	4 156	5.14%
එ.ජ.ප. පී.එ.ආ. NDI 2	3 125	3.87%
එ.ජ.ප. පී.එ.ආ. NDI 3	656	0.81%

එ.ජ.ප. පී.එ.ආ. VAO 80,831 95.19%
එ.ජ.ප. පී.එ.ආ. REJECTED 4,081 4.81%
එ.ජ.ප. පී.එ.ආ. POLLED 84,912 99.41%

Election results in Sinhala

MIS FOR KOTMALE MILK PRODUCERS' COOPERATIVE

The Kotmale Milk Producers Cooperative wanted their accounts and the milk production statistics computerized. The project was contracted to the University of Colombo School of Computing and handed over to Mr. Harsha Wijayawardhana and to Ms. Shiromi Arunatilake. The Kotmale Milk Producers Cooperative required this work to be carried out in Sinhala and this was when they encountered an impediment. There was an immense problem at this time when handling ICT in Sinhala. It was not possible to use both Sinhala and English in the same document. It was necessary to incorporate the font with the application. Mr. Harsha Wijayawardhana worked with Visual Basic and tried to create an Accounting Management Information System for the milk producers of the Kotmale Cooperative. He worked with the font Sarasavi, which had been developed by Mr. Nandasara. This was an 8-bit ASCII true type serif font.

The website kirana.lk which Mr. Harsha Wijayawardhana developed had information in Sinhala. The user had to download the font. This site had useful information but the site was not effectively used.

LANGUAGE TECHNOLOGY RESEARCH LABORATORY – UCSC

The Language Technology Research Laboratory (LTRL) was established at UCSC in 2004. It was headed by Dr. Ruwan Weerasinghe and the initial team consisted of Mr. Vincent Halahakone, Mr. Dulip Herath, Mr. Viraj Welgama, Mr. Asanka Wasala, Mr. Nishantha Medagoda, Mr. Rajathurai Premkumar and Ms. Kumudu Gamage.



The LTRL Team

The objective of LTRL was to address the ever-increasing need for local language computing through localization and through research into language processing and development. LTRL has developed numerous useful applications, most of which have been made available through its website. Some of these are described below.

TEXT CORPUS COLLECTION

Sri Lanka, through the UCSC joined the PAN Localization Project in 2003. Sri Lanka joined seven other countries in the region on developing the localization initiative. These localization initiatives were supported through the PAN Localization grant from the International Development Research Center (IDRC) and administered by the Center for Research in Urdu Language Processing, National University of Computer and Emerging Sciences, Pakistan.

Initially, it was necessary to build resources before applications in local languages could be developed. Accordingly, a corpus of 10 million words of contemporary Sinhala was developed by LTRL of UCSC. This was one of the key deliverables under Phase 1 of the PAN Localization Project. The corpus was developed during the period March 2004 to July 2006. It was developed with the intention of building the language processing infrastructure for Sinhala which was needed for subsequent research. Sources such as novels and short stories were used. Mr. Vincent Halahakone had also been collecting words for a corpus, and he was later connected to LTRL's initiatives.

LTRL also built a news corpus separately, consisting of articles from the Sinhala newspapers Divaina and Dinamina. The articles were classified into several categories such as news articles, editorials, feature articles and sports news etc.

After completing this task, LTRL embarked on annotating⁴ the corpus with parts of speech (POS) labels⁵. This task was carried out during the period

⁴ Adding linguistic information

⁵ a set of labels to identify words by its functions

August 2006 to March 2007. As the first step of the tagging process, the LTRL defined a ‘Parts of Speech Tagset⁶’ for Sinhala in consultation with senior linguists and also with reference to other Indic language Tagsets.

THE SINHALA COLLATION SEQUENCE

Users who were maintaining databases and lists of information in three languages, especially people from the Government sector, had been first asking CINTEC and thereafter ICTA for a standard Sinhala collation sequence. The Sinhala collation sequence was needed when setting up databases and other lists of information in Sinhala. There were many issues with regard to the collation sequence of Sinhala that needed to be clarified. Work on the Sinhala collation sequence was carried out by UCSC. The team was headed by Dr. Ruvan Weerasinghe. The UCSC team studied all major Sinhala dictionaries to determine the main points of divergence with respect to the collation sequence. They also consulted a group of eminent Sinhala scholars drawn from the different linguistic persuasions, individually, to ascertain their stand on each of these issues together with their justification.

Collation sequences that were being adopted by key State institutions were researched and documented. Consequently, UCSC recommended a suitable collation sequence to ICTA’s Local Languages Working Group (LLWG). It was presented and discussed thoroughly at LLWG meetings. This collation sequence consisted of a dictionary collation sequence and a simple collation sequence.

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

⁶a set of labels to identify words by its functions

The simple collation sequence is to be used for preparing lists of names, lists of places etc., and produces results identical to the dictionary collation sequence when collating personal names, place names and other common data. This is easier to implement, thus encouraging vendors to support Sinhala in their products, and produces a result which will not confuse a user who is not aware of the subtleties of the language. The two collations will produce different results only between words with the letter ජ or ඛ and the letter ඥ in a given position.

After much discussion the recommended collation sequence was included in the draft Sinhala ICT standard and was approved by the Sri Lanka Standards Institution as a Sri Lanka Standard. This Sinhala collation sequence was standardized as Part 1 of the Sri Lanka Standard, SLS 1134 : 2004.

SINHALA SPEECH RECOGNITION

Later, in 2008, LTRL set about designing and developing a speech corpus for Sinhala, since this would facilitate future research on Sinhala speech recognition.

Most modern automatic speech recognition systems use statistical models which are trained using speech corpora. A speech corpus is a main part of speech recognition research and it greatly impacts the performance of a speech recognizer. A speech corpus is a database of speech audio files and text transcriptions, in a format that can be used to create acoustical models using speech recognition engines. Such systems are highly reliant on the comprehensiveness of the relevant speech corpora.

Therefore, the LTRL team realized that efforts should be made to build a good speech corpus for a low-resourced language such as Sinhala. The Sinhala language lacks proper speech corpora for speech recognition research.

The team, consequently, designed and developed a speech corpus to be used in Sinhala automatic speech recognition. The corpus consisted of 78,667 utterances recorded from mobile phone calls and the length was around 65 hours of speech. Research on Sinhala speech recognition continued under a project initiated later by the ICT Agency (ICTA) of Sri Lanka, described in a subsequent chapter.

MACHINE TRANSLATION

Through machine translation text from one source language can be translated to another target language using an automated procedure. Machine translation systems can be based on rules or can be data driven. Statistical Machine Translation (SMT) is a well-known and well-established data-driven approach used for language translation. SMT is based upon statistical models built on bilingual corpora.

LTRL commenced a project on machine translation in 2008, on translating from Sinhala to English and English to Sinhala. This project was also a component of the PAN Localization project. The purpose was to find out how translation from English to Sinhala could be accomplished using SMT techniques.

After completion of the project, the LTRL team intended to use larger corpora for the Sinhala language with regard to future work. This research was a promising start towards better information access for rural communities via English to Sinhala machine translation.

Later, in 2013, LTRL commenced a project on machine translation to translate from Sinhala to Tamil and vice versa. The purpose was to find out how the SMT systems perform for Sinhala to Tamil and Tamil to Sinhala translation.

The focus of this work was to develop a statistical machine translation system for Sinhala and Tamil language sentence pairs. LTRL ascertained how Sinhala-Tamil SMT performance varied with the amount of parallel training data being used. This was done in order to find out the minimum amount of data needed to develop a machine translation system with acceptable performance.

Tamil to Sinhala and Sinhala to Tamil translation was unable to produce intelligible output with a parallel corpus of only 25,000 sentence pairs of relatively short length. But it was expected that performance would approach usable levels by collecting a larger parallel corpus.

However, the error analysis showed that the sentence length limitations of the Sinhala-Tamil parallel corpus could not be the only reason for the comparatively lower BLEU scores⁷. Morphological richness could have been the reason for getting lower results since misspelled words and proper names are common to other languages too. Furthermore, a preliminary study shows that better perplexity values can be taken for the same dataset used for this research by stemming suffixes of the Sinhala and Tamil parallel sentences. This research is still in progress using the current state of the art machine translation technique; Neural Machine Translation (NMT). In the future, LTRL is planning to implement a system capable of producing acceptable translations between Sinhala and Tamil for use by the wider community.

OPEN-SOURCE LANGUAGE TEACHING FRAMEWORK

LTRL developed Shikshaka, a computer-based teaching framework for teaching languages using a dialogue-based andragogy. Effective technological methodologies were used to develop the system as an interactive tool. The framework was used to implement scenarios: to teach Tamil using Sinhala and to teach Sinhala using English.

⁷ Bilingual Evaluation Understudy Score; a measurement for comparing a particular translation of text to one or more reference translations.

The content management framework and the user interfaces of the system were implemented using Adobe Flash technology. ActionScript was used for communication between the user interface and XML files in developing the on-demand loaded content for the flash application. The choice of Flash enabled the delivery of an efficient, lightweight, platform independent, multimedia-rich and web-enabled system. One important gain in using this combination of technologies was that it was able to overcome the issue of Flash not supporting Sinhala Unicode rendering. Two language learning courses were developed for teaching Tamil through Sinhala and Sinhala through English as a proof of concept for the framework. The courses can be downloaded through LTRL website (ltrl.ucsc.lk). The Shikshaka framework can also be extended to support levels of study.

TEXT TO SPEECH

A Text to Speech (TTS) system is a computer-based system capable of converting text to its spoken form. This application is useful in many industries and is especially useful for those who are visually impaired and for senior citizens. TTS systems are capable of reading text from different sources such as newspaper articles, websites or e-books, and converting the analyzed text into spoken language through a computational mechanism.

Since UCSC was the partner organization in Sri Lanka for the PAN Localization project, the Sinhala TTS system too, was originally developed under this initiative. At the very early stages in developing a TTS system for Sinhala, research was carried out by LTRL.

The LTRL team emphasized that finding correct pronunciation for a given word is one of the first and most significant tasks in the linguistic analysis process. This problem was addressed by formulating a set of rules. The rules were tested using 30,000 distinct words obtained from a corpus and compared with the same words which were manually transcribed to phonemes by an expert.

This Sinhala TTS system was developed based on Festival, an open source speech synthesis framework. The system was developed up to a state where it could be used by those who were blind. This TTS system had a male voice. But the voice sounded robotic. It was tested by the Ability Center of the University of Colombo. The members of the Ability Center were visually impaired. It was also used by Prof. Weerakkody of the University of Peradeniya. He too, was blind, and was extensively using TTS systems. This product won the award “Best Innovative Product” at INFOTEL 2008.

The dataset of this product was later, around 2009/2010, used by the University of Moratuwa to develop a TTS system. This TTS system too, had a robotic voice – it was a female voice – but this voice sounded more natural than that of the product developed earlier by UCSC.

Thereafter, in 2014 the World Intellectual Property Organization (WIPO) provided funding for a Sinhala TTS Project. With the experience gained through a student’s MSc project on TTS, which had been supervised by Dr Ruvan Weerasinghe at UCSC, the LTRL team embarked on developing a human voice TTS system. LTRL commenced developing this TTS system mainly based on a request made by the Ability Center of the University of Colombo. The people at the Ability Center could sense gaps in the spoken voice and errors in the system which most people with sight could not discern. Therefore, the Ability Center specifically wanted the existing TTS system to be further improved.

For this project, the open source Mary TTS framework was used. A set of rules for mapping text to sound was identified. The datasets used for this study were gathered from newspaper articles. The corresponding sentences were recorded by a professional speaker. The pronunciation dictionary which was used for the study consisted of 5000 words, which the team agreed, was not sufficient. This was enlarged later. The voice in this product sounded far more natural than in the earlier versions.

User evaluation was conducted with 20 persons from the Ability Center. Evaluation was carried out on intelligibility, speech quality and naturalness. Intelligibility and the naturalness of the Sinhala TTS system received an approximate score of 70%. The overall speech quality was approximately 60%.

CHECKING AND CORRECTING SPELLING ERRORS IN SINHALA

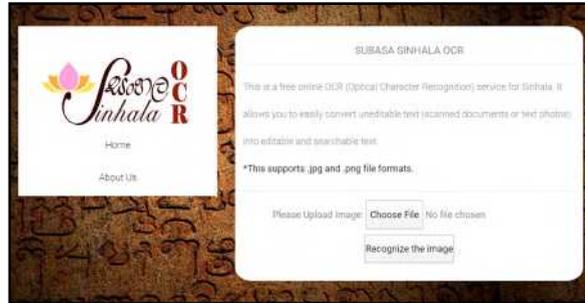
LTRL was involved in developing a spell checker for Sinhala, as an independent desktop application with a graphical user interface and also as a web-based version.

Results showed a promising performance, achieving an average accuracy of 82%. The LTRL team was of the view that it was possible to apply this technique to develop spell checkers for other phonetic languages where linguistic resources are scarce or non-existent. This product is hosted on the web portal www.subasa.lk.

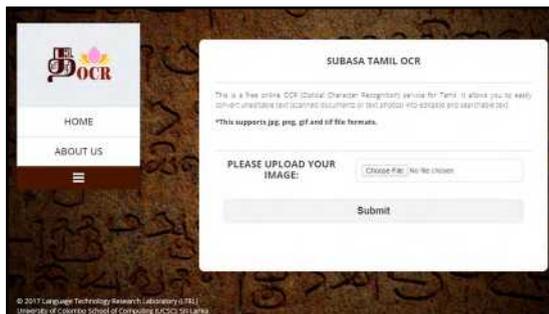


OPTICAL CHARACTER RECOGNITION (OCR) SYSTEM FOR SINHALA

OCR is the electronic or mechanical conversion of handwritten or printed letters within images or pdf documents into text. Commercial systems are available for English and other languages which have Latin-based character sets. LTRL first tried to build a font specific OCR system for Sinhala as a component of the PAN Localization project in 2004. After Google's Tesseract OCR was released, LTRL tried developing an OCR for Sinhala using deep leaning – a part of machine learning - techniques. The latest system which LTRL developed has been made available on their website.



TAMIL DIGITIZATION



LTRL implemented the project “Tamil Digitization” under ICTA’s e-Society program. The objective was to develop the tools needed to automatically recognize the most common printed Tamil fonts from scanned

images of books and documents for digitizing the content.

There were no widely available optical character recognition solutions for printed Tamil scripts globally. Therefore, scanned or hardcopy printed Tamil documents to be digitized were being retyped manually. Under this project, the LTRL team developed an OCR solution for Tamil through which a satisfactory level of accuracy could be obtained.

SOME OTHER LTRL INITIATIVES

LTRL has also developed a computational grammar for Sinhala, and a morphological analyzer for Sinhala. Other work done under the PAN Localization Project include; developing a Sinhala Wordnet, developing domain names for Sinhala and developing a framework to create language learning tools. Two language learning tools to learn Tamil through Sinhala and Sinhala through English were also developed using this framework.



The LTRL team with the e-Swabhimani award

LAKapps CENTER

People who were more proficient in Sinhala or Tamil than in English prefer to use localized computer applications. The University of Moratuwa established the Center of Excellence on Localized Applications – the

LAKapps Center – to provide expert knowledge and to carry out research related to localized applications. Prof. Gihan Dias and his team aimed to “extend the benefits of ICT to those who find the English language a barrier in connecting to the modern world and its competence”.



Dr. Sanath Jayasena



The Center was launched with a Symposium on Localized Systems and Applications, held on 2nd September 2009 at the University of Moratuwa. Dr. Sanath Jayasena, a Senior Lecturer at the Department of Computer Science and

Engineering, pointed out that the symposium provided a forum whereby developers could present their original work, meet peers and widen their horizons.

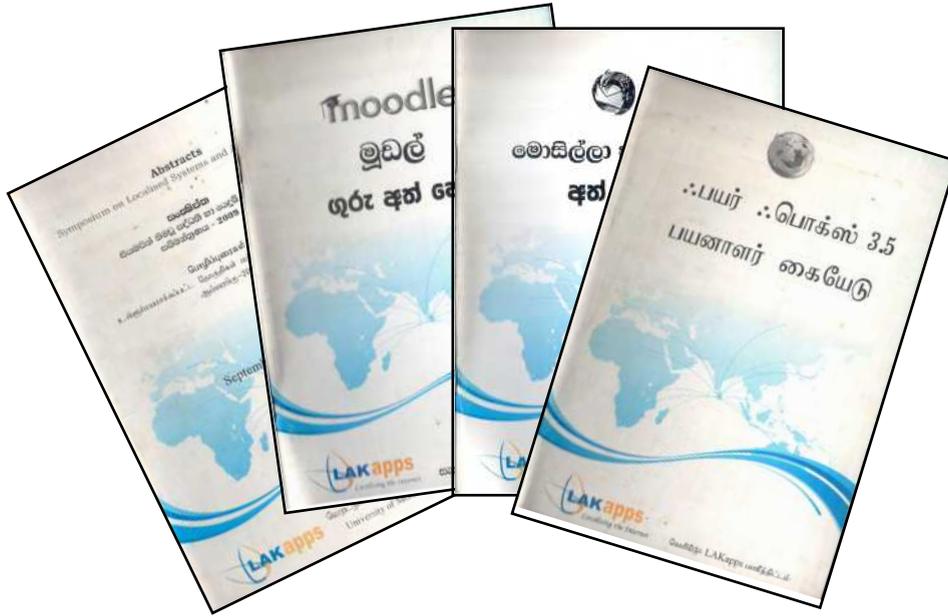
The work on the pilot project implemented by the LAKapps Center covered several Districts and locations. The Sinhala and Tamil “pack” comprising the keyboard driver and several Unicode fonts were installed at many locations, including in several Nenasala Centers, the access centers set up by the ICT Agency of Sri Lanka.

The locations at which the language pack with the keyboard driver was installed included the following;

- The Nenasala at Sevanagala
- Embilipitiya Central College
- The Nenasala at the Kotmale Community Radio
- The Computer Resource Center at Bandaragama Maha Vidyalaya
- Kottawa Dharmapala Maha Vidyalaya
- St. Sylvester’s College, Kandy
- Ramanathan Hindu Ladies’ College, Colombo

In addition to installing the Sinhala/Tamil language pack the Center also installed other localized software; viz. Sinhala Firefox web browser; Tamil Firefox web browser; Sinhala Thunderbird email client; Tamil Thunderbird email client; Sinhala Joomla content manager (which was not localized by the LAKapps Center); Tamil Joomla content manager; Sinhala Moodle learning management tool (not localized by the Center); and the Tamil Moodle learning management tool.

Training:



The LAKapps Center developed user guides and training material in Sinhala and Tamil for the software. This is noteworthy, as many find it difficult to follow user guides written in English. Multimedia learning material for OpenOffice (word processor) and Calc (spreadsheet) were also developed.

Training on local language computing and use of localized applications was conducted at each site. Nearly 150 adults and 300 students were trained in the use of IT applications in their own language. The coordinators at each site were also given specific training so that they could support end-users, monitor the use of the software and provide feedback for evaluation to the LAKapps Center. This initiative was partly funded by the e-Society program of the ICT Agency of Sri Lanka.

The National Language Processing (NLP) Center was set up in 2017. One of its key projects is intended to improve six-way translations between Sinhala, Tamil and English. In order to meet this objective, a tool for professional translators was developed. Documents are provided to the NLP Center by Government organizations, and these are translated and returned.

Building the language processing infrastructure is essential for subsequent research. Resources are needed to develop tools. Consequently, the University of Moratuwa and the Language Technology Research Lab (LTR) of UCSC jointly implemented a project on developing a parts-of-speech (POS) tagset for Sinhala. This is the process of marking up a word in a text to a part of speech. As stated before, UCSC had developed a POS tagset earlier. Full grammatical tagging for Sinhala and Tamil will be developed next and thereafter full computational grammar will be developed.

THE COMPUTER AND INFORMATION TECHNOLOGY COUNCIL OF SRI LANKA (CINTEC)

The Computer and Information Technology Council of Sri Lanka (CINTEC) was established as a body corporate under the Computer and Information Technology Act no. 10 of 1984, as the apex Government body on Computers and IT. One of the functions of the Council as stated in the Act was to take all necessary measures to promote, facilitate and assist the use and application of computers and IT in Sri Lanka with a view to improving the quality of life of its people and enabling Sri Lanka to acquire the necessary capability to meet the challenges of technological change. Facilitation and assisting the use and application of IT could not be done in English. The implication here is that the use of IT in Sinhala and Tamil had to be enabled if the functions stated in the Act were to be carried out properly. CINTEC carried out its work through setting up Working Committees with a few people recruited as staff. Several

Working Committees were set up at the inception. Some of these were the Committee on Telecommunications and Data Transfer, the Committee on Computer Applications in the Public Sector and the Committee on Computer Education.



The first CINTEC Board

Another such Working Committee was the Committee on Recommending Standards for the use of Sinhala and Tamil in Computer Technology. The first Chairman of

CINTEC was Prof. Mohan Munasinghe. CINTEC was later chaired by Prof. V.K. Samaranyake, who later became known as the “Father of IT in Sri Lanka”. Prof. Samaranyake took an enthusiastic interest in trying to enable the use of ICT in Sinhala and worked extensively on the area of ICT and Local Languages.

CINTEC worked on the following areas:

- Defining the Sinhala character set and its order.
- Developing the Sinhala Standard Code for Information Interchange (SLASCII).
- Inclusion of Sinhala in the Unicode Standard and addressing the issues encountered.
- Working on the Sinhala keyboard layout which was to be standardized in a consultative manner.
- Addressing terminology issues and developing a Sinhala glossary of technical terms.

CINTEC was later renamed as the Council for Information Technology. The Working Committee on Recommending Standards for the use of Sinhala and Tamil in Computer Technology was set up in 1985. This Committee took steps to meet the immediate need to agree on an acceptable Sinhala character set and order. This Committee joined a Committee appointed by the Natural Resources, Energy and Science Authority of Sri Lanka (NARESA) to form the Committee on Adaptation of National Languages in IT (CANLIT), which agreed on a unique Sinhala character set and order. No action was taken on Tamil in the 1980s because work was being undertaken in India. CANLIT consisted of experts in the Sinhala language as well as in IT.

CANLIT arrived at defining the Sinhala character set as having 16 vowels, 2 semi consonants and 41 consonants as shown in the CINTEC publication of 1990. Thirteen (13) consonant modifiers were also identified. A new character to denote “fa” (ඞ) was introduced. CANLIT also agreed on the order of the Sinhala character set, with a slight modification. This exercise took a representative group of language and technology experts several months to arrive at a consensus.

IT GLOSSARIES - SINHALA GLOSSARY

In the early nineties Dr. Gamini Wickremasinghe of Informatics Ltd. discussed with Mr. S.M. Banduseela, who was then the Director, IT at the Urban Development Authority, the possibility of compiling and publishing a dictionary of computer terms in Sinhala. Mr. Banduseela had extensive experience in translating scientific books from English to Sinhala. Dr. Wickremasinghe was of the opinion that unless sufficient literature on computing was made available in local languages, it would not be possible to attract a sufficient number of young students to the computer field.

As a first step Dr. Wickremasinghe wanted to compile a dictionary of computer terms in which the meanings of the technical terms would be given in Sinhala. He therefore selected the 'Penguin Dictionary of Computers' and obtained copyright to translate and publish this book in Sinhala. He requested Mr. Banduseela to carry out the translation.

When Mr. Banduseela commenced working on the translations, he was faced with the problem of finding suitable Sinhala terms. He therefore met Prof V.K. Samaranayake who was then the Chairman of CINTEC to seek his advice on this issue. Prof. Samaranayake informed Mr. Banduseela that CINTEC had established a Committee to develop suitable technical terms in Sinhala and requested Mr. Banduseela to join the CINTEC Sinhala Glossary Subcommittee.

The Sinhala Glossary Subcommittee was established under the main Working Committee on Recommending Standards for the use of Sinhala and Tamil in Computer Technology, to work on a Sinhala Glossary of Technical Terms in Computer Science. It was this Subcommittee that Mr. Banduseela joined.

The CINTEC Glossary Subcommittee comprised experts from a wide range of areas such as language experts and IT professionals. Prof. J.B. Disanayaka was a key member. This Subcommittee took a long time to compile the dictionary because the Sinhala language experts could not agree on the terms. This was because they represented divergent schools. In this Subcommittee there were scholars in oriental languages (Prof Mahinda Palihawadana), those representing the Hela school (Mr. Aelian de Silva) and moderates such as Prof J.B. Disanayaka. Those representing the Hela school were opposed to coining technical terms using Sanskrit words. But they were not totally opposed to using the English term with appropriate modifications. The Hela school also opined that it is the verb that should be translated and not the nouns.

The Subcommittee meetings at which the words were coined were held at CINTEC, usually once a week. The meetings were coordinated first by Ms. Mayura Wijesinghe and at a later stage by the late Ms. Prasadi Jasinghe. The Subcommittee took a while to finalize the glossary, but ultimately, there were results.

This Committee completed the first Glossary of Technical Terms in Computer Science in 1991. This was deemed to be the first Sinhala glossary on Computer terms. It was published by the Department of Educational Publications. Since this publication had only 1000 terms the Glossary Subcommittee continued to work, taking into consideration the needs of this rapidly developing field. The technical terms developed by this Subcommittee over several years were handed over to the Official Languages Commission. After it was reviewed by another committee, the Official Language Commission published these terms as “Glossary of Technical Terms – Information Technology” in 2000.

IT GLOSSARIES - TAMIL GLOSSARY

In the early 1990s CINTEC established a Subcommittee to work on a Tamil Glossary of Technical Terms in Computer Science. It was accepted that the Tamil terms used in Sri Lanka are not the same as those used in India and in other countries. CINTEC purchased two copies of the Penguin Dictionary of Computers and handed over one to the Sinhala Glossary Subcommittee and the other to the Tamil Glossary Subcommittee. The Tamil Glossary Subcommittee was chaired by Professor Uvais and comprised Tamil language experts, linguists and IT professionals. The members worked rigorously and effectively. In a very short time they had completed translations up to the letter “T” in the English Dictionary. The members used to gleefully check on the work being carried out in parallel by the Sinhala Glossary Subcommittee. The Tamil Glossary Subcommittee was always ahead. But later CINTEC introduced a new member who was an ICT professional, to the Tamil Glossary Subcommittee.

This new member deemed that this was not the way in which the work should be carried out. This new member stated that the translations should be carried out under three areas, namely hardware, software and communication, much to the chagrin of the other Subcommittee members. After this intervention, unfortunately there was no progress and no Glossary was published. All the translations that had been done were lost. CINTEC could not retrieve the work that had been done.

THE SINHALA STANDARD CODE FOR INFORMATION INTERCHANGE SLASCII

The requirement for a standard code for Sinhala characters was identified in the mid-eighties. In developing the Sinhala Character set for use in IT, the work already done in Thailand for the Thai language, which was somewhat similar to Sinhala, was studied with the Thammasat University, Bangkok. At this stage the aim was to develop a 7-bit code to fill the positions A0 to FF in the single byte ASCII code table (ISO 646). The draft standard code was approved by the Council of CINTEC on the advice of its Working Committee for Recommending Standards for the use of Sinhala and Tamil in Computer Technology.

A standard Sinhala encoding known as SLASCII was developed by CINTEC and was approved by the Sri Lanka Standards Institution (SLSI) as Sri Lanka Standard 1134 in 1996. However, no implementation of SLASCII was done. The SLASCII standard differed in many aspects with Unicode for Sinhala.

Sinhala was first included in the Unicode standard and the Universal Multiple Octet Coded Character Set (known as the Universal Character Set, UCS) at a meeting of the Working Group2 (the Group was officially termed ISO/IEC JTC1 /SC2 /WG2) at its meeting held in Crete, Greece from 30th June to 4th July 1997. Prof. J.B. Disanayaka and Mr. S.T. Nandasara from Sri Lanka attended this meeting. This was a significant milestone.

ISO/IEC JTC1/SC2/WG2 is the international standardization working group of the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC) for the Universal Character Set (i.e. the Universal Multiple-Octet Coded Character Set - ISO/IEC 10646)

The scope of the Working Group 2 was to develop the Universal coded character set that encompasses all the world's language scripts, all the symbols, characters and emoji used in ICT. Each character in the Universal Character set is identified by a unique code. ISO works with the Unicode Consortium to develop the Universal Character set.

SINHALA KEYBOARD LAYOUT

Typists in Government Departments had been using Olympia and Olivetti typewriters. In these typewriters, the Sinhala characters had been on various different keys; specific Sinhala characters were not on the same key on these typewriters.

Therefore, typists faced difficulties. One Sinhala keyboard layout was necessary for all Government Departments. Consequently, the



Ministry of Public Administration set up a Committee to decide on the keyboard layout. Members of the Typists Union were requested to be present in Colombo so that their views could be sought. Online surveys were not possible then. The Committee that was set up was headed by Dr. Nandadeva Wijesekera. Hence, the perennial Wijesekera keyboard layout.

Thereafter, Sinhala typists used the Government approved Wijesekera keyboard. CINTEC carried out extensive consultation on the Sinhala keyboard layout to be used with computers. The final agreement was on the Wijesekera layout. CINTEC first developed and obtained Government approval for the “Extended Wijesekera Keyboard for Electronic Typewriters”, the intention being the introduction of Daisywheel and Golf-ball electronic typewriters then used as an interface for microcomputer output. It included the new character fa (ආ) and three other additional key positions. This layout was once again modified for use with a standard 101-Key computer keyboard.

In the late 1990s, some users from whom CINTEC sought views recommended keyboard layouts with parts of letters on keys – for example, a part of the letter ක and a part of the letter ඞ to create කඞ - to create conjunct letters. This was not necessary and these issues were smoothly sorted out. CINTEC held the final consultation with stakeholders for taking a decision on the Sinhala keyboard layout in June 2003, a few weeks before CINTEC closed its operations and ICTA was set up. It was unanimously agreed on at this consultation that the standard keyboard layout should be based on the Wijesekera layout. The keying in sequences are based on the “type-as you-write” method. Each letter is typed in the order in which it is written. Later, in 2004, this layout and the keying-in sequences were included in the second revision of the Sri Lanka standard, SLS 1134.

Kumara browsed the Internet during his lunch-hour. The Lakehouse newspapers were now online. He read the Sinhala website Budusarana, which was a newspaper published by Lakehouse and he was impressed. The Internet was not ubiquitous then. Kumara lived in Weligama, in Southern Sri Lanka. During this period, there was no widespread availability and affordability of telecommunication services and Internet access in areas away from the Colombo district. Kumara therefore decided to create a small booklet comprising the articles in Budusarana, without the pictures. He wanted to take the booklet home so that his family could read Budusarana. The Lakehouse newspapers Daily News and the Sunday Observer had been published online by Lanka Internet for the first time on 4th September 1995. But Unicode Sinhala newspapers were not online. So Kumara tried to copy the article and paste it on to Word. The result was unreadable symbols not Sinhala words. The articles had been published as images, not Sinhala text. Unicode Sinhala fonts were not yet available. Lakehouse had not developed its Unicode Sinhala font Dinamina. He gave up the effort.

Chapter 4

Trying to spread the Internet in Sinhala & Impediments encountered

THE INTERNET COMMITTEE TRIES TO SPREAD THE INTERNET IN SINHALA

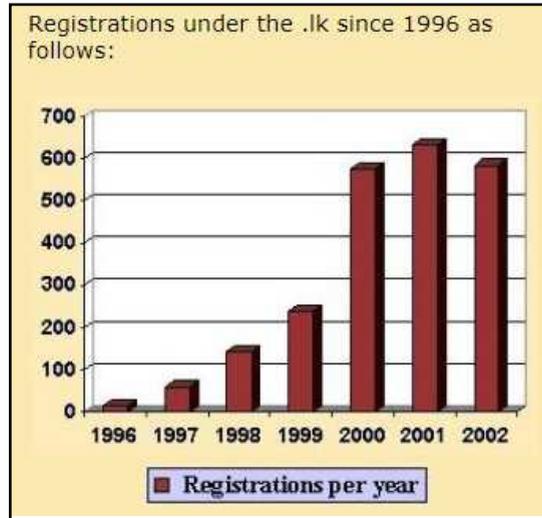
There were several initiatives taken to spread the use of ICT and the Internet in Sinhala. But a key impediment was that there were no standards being used.



The CINTEC Internet
Committee

The CINTEC Internet Committee was set up to develop the Internet in Sri Lanka, to address and alleviate issues in using the Internet in Sri Lanka and to spread its use in the country. Its members comprised the LK Domain Name Registrar, representatives from CINTEC, UCSC, the Telecommunications Regulatory Commission, the Licensed Internet Service Providers Association (LISPA), the legal profession, the Computer Society of Sri Lanka and the private sector. In 2002 the Committee realized that the Internet was being used mainly in Colombo and its environs. Therefore, the Committee agreed that the Internet should be popularized throughout the island so that people in rural areas are also able to reap the benefits of the Internet.

Consequently, the Committee assigned one of its members Mr. Navin Peiris to draft a prototype for a Sinhala website. The carefully developed prototype was presented at one of the Internet Committee meetings and it was approved. The next step was to create relevant content in Sinhala, develop websites and make information available in Sinhala to the country at large. This was when the Committee members came across a huge impediment. They realized that it was not possible to use ICT and the Internet in “actual” Sinhala.



The fundamental issue was the lack of standards. A multitude of legacy fonts and keyboards fragmented the market and hindered development. The fonts that were being developed and used then, mapped Sinhala letter symbols to the codes that were used by the English alphabet in ASCII. The two most common mapping schemes were based on the Wijesekera keyboard layout and the “phonetic” layout. Consequently, the codes allocated to Sinhala letters were based on the codes of the corresponding English letters.

The Internet Committee studied the existing fonts and realized that the only international standard for a Sinhala character set was the Unicode (Universal Encoding) standard. Unicode included provision for all languages, including Sri Lanka’s local languages, Sinhala and Tamil. It provides a unique encoding for every character, every symbol and even for many emojis. This was 2002 and

all major computer systems supported the Unicode standard in 2002, therefore they could easily be adapted to support Sinhala. The Unicode standard was accepted as the answer to multilingualism on digital devices without changing the way in which we wrote Sinhala.

Therefore, the Internet Committee realized that there was a lot of work to be carried out before the Committee could promote the Internet in Sinhala. The Committee could not proceed at this point in popularizing the Internet in Sinhala. The end was not in sight.

Therefore, the Internet Committee decided to establish another Committee through CINTEC. It was necessary to look into the issues of supporting Sinhala under the Unicode standard, to clarify the use of Unicode, and make recommendations on resolving ambiguities, but not make any changes to the approved Unicode standard, to make recommendations on a keyboard layout, character representation, etc. and to promote the implementation of Unicode compatible Sinhala fonts.

In December 2002, CINTEC set up the Committee on Unicode Compatible Sinhala Fonts. This name was later abbreviated and it became known simply as the “Fonts Committee”. The Committee was chaired by Dr. Gihan Dias, the LK Domain Registrar, and comprised Ms. Aruni Goonetilleke from CINTEC, Mr. Harsha Wijayawardhana and Dr. Ruvan Weerasinghe from the UCSC, representatives from newspapers, and font developers. This Committee was to define the basic minimum requirements for Unicode compatible Sinhala fonts; define the essential features which should be present in a Sinhala character set, character combinations and their input; address the requirements for a standard Sinhala keyboard, keyboard stroke sequences, and issues relating to the glyphs and keyboard drivers.

The Fonts Committee decided that work should be carried out in four major areas in order to enable computing in Sinhala. These were: a standard encoding of Sinhala characters; development of Sinhala fonts; standardize the Sinhala keyboard; and availability of standards-based applications and utilities (such as spell-checkers).

January 2003 was a time of uncertainty at CINTEC. CINTEC was to be closed and a new organization was to be established in its stead. Consequently, most of the CINTEC Working Committees ceased to function. Despite the uncertainty in the physical environment, the Fonts Committee worked smoothly and effectively and continued their work in striving to enable the use of ICT in Sinhala until the closure of CINTEC.

The functions of this Committee were incorporated into the Language Requirements Working Group – which was later named “the Local Languages Working Group - of the Information and Communication Technology Agency (ICTA) of Sri Lanka, CINTEC’s successor.

Kumara had now temporarily stopped his pursuit of the font Bindumathie and instead made lists of the furniture and lists of the computers which would be taken from CINTEC to the new ICTA premises at Kirimandala Mawatha, in Colombo. He also had to order stationery and other office requisites for this new office. ICTA was to inherit CINTEC's assets and liabilities. The CINTEC work which ICTA should continue was also selected; ICT in local languages was a key area to be continued at ICTA. So Kumara – the person ceaselessly seeking the font Bindumathie - ordered a lorry to take the furniture. Stakeholders in the ICT arena were invited to Kirimandala Mawatha. ICTA commenced operations on 1st July 2003 at Kirimandala Mawatha, in Narahenpita. It was a heady time with drive and energy; it was necessary to define projects and achieve targets, so Kumara had to keep his quest for the font Bindumathie in abeyance.

Chapter 5

*The e-Sri Lanka development program, setting up
ICTA & the need for local languages*

In November 2002, the Government of Sri Lanka launched the e-Sri Lanka Development Program with the objective of using ICT in all its aspects for the benefit of the people of Sri Lanka and to further socio-economic development throughout the nation. Through the implementation of the e-Sri Lanka Development Program an enabling environment was to be created, where Government would work in partnership with stakeholders to create the infrastructure and establish e-Government services. The e-Sri Lanka roadmap was transformed into the multi-donor funded e-Sri Lanka Development Program, which resulted in the enactment of the Information and Communication Technology Act No. 27 of 2003, under which the Information and Communication Technology Agency (ICTA) of Sri Lanka was established as the legal successor to CINTEC.

ICTA's strategy was to provide affordable access and opportunities by improving ICT infrastructure; improving the secure delivery of public services and governance using ICT; enhancing the quality of education, promoting learning and research through the use of ICT; building a strong ICT industry; proliferating access to ICT and use of ICT applications in rural areas for rural socio-economic development; and creating the institutional leadership to carry forward e-Development in Sri Lanka.

To achieve the above, a six-program strategy was defined. It encompassed the programs Information Infrastructure; ICT Policy, Leadership and Institutional Development; Re-engineering Government; ICT Human Resource Development; ICT Investment and Private Sector Development and the program e-Society.

The e-Sri Lanka development program, setting up ICTA & the need for local languages

TAKING THE DIVIDENDS OF ICT TO EVERY VILLAGE AND EVERY CITIZEN – THE NEED FOR SINHALA AND TAMIL

This was 2003. ICTA was newly set up. Everyone, led by the CEO, Mr. Manju Haththotuwa, participated in deciding on a very appropriate vision after much discussion; “to harness ICT as a lever for economic and social advancement by taking the dividends of ICT to every village, every citizen, to every business and also re-engineer the way Government thinks and works”.

But then everyone came across a colossal roadblock; the dividends of ICT had to be taken to every village and every citizen in Sinhala and Tamil. Many people in Sri Lanka, especially in the rural areas, were not conversant in English. Acquiring sufficient conversance in English to be able to make use of the content that was profusely available in English on the Internet, takes time. The ability to read a drop-down menu on a computer or fill up a form in English would not enable a person to make real use of the rich content that was available in English on the Internet. Therefore, a fundamental and crucial requirement for all the work that ICTA was to do was to enable the use of ICT in Sinhala and Tamil. Enabling the use of ICT in Sinhala and Tamil was a foundation. Thereafter, relevant content in Sinhala and Tamil had to be developed and other entities had to be triggered to develop local language content.

RE-ENGINEERING GOVERNMENT DEMANDS SINHALA

The Re-engineering Government program needed to work in Sinhala and Tamil to develop the Government websites, they needed the use of IT in Sinhala and Tamil enabled for developing trilingual databases with lists of names. This team was planning to set up the Lanka Government Network (LGN), through which central and provincial Government organizations throughout the island were to be connected. The team implementing this project LGN wanted to

know what the standard keyboard layouts were for both Sinhala and Tamil. Agreement had been reached on the Sinhala keyboard layout. CINTEC had worked hard on the Sinhala keyboard layout, carried out public consultations and an agreement had been reached on what the Sinhala keyboard layout would be; the Wijesekera layout had been agreed on and keying-in was based on the “type-as-you-write” method. This was not yet standardized, but work had to commence immediately on the Tamil keyboard layout. The LGN team needed to know whether keyboard drivers were developed for input, whether keying-in was enabled and what the standards were. This meant that the Sri Lanka Standards (SLS) had to be in place. In fact, most of the work the Re-engineering Government team intended to do had to be in Sinhala and Tamil.

NENASALA CENTERS NEED SINHALA

Under the “Infrastructure Program”, Nenasala centers, or access centers were to be set up throughout Sri Lanka. These centers were to provide public access to information and communication services through different means, and extend the benefits of ICT to communities which found it difficult to access such services. These were to be set up in easily accessible and convenient locations. The team implementing the Nenasala project came across impediments which they had not fully envisaged. ICTA placed advertisements for recruiting Nenasala owners who would also be the operators of the Nenasala centers. But people were hesitant in applying as Nenasala owners and operators



The e-Sri Lanka development program, setting up ICTA & the need for local languages

because of their perception that English would be needed for such a role. They were hesitant in applying due to the fact that user interfaces on computers were all in English. They were not conversant with the language that was used. They did not think they could play the role of a Nenasala operator effectively due to the language barrier.

E-SOCIETY NEEDS LOCAL LANGUAGES

One of the strategies of ICTA's e-Society program was to facilitate the development of innovative locally relevant ICT solutions and develop local content to make ICT more meaningful to rural communities. The e-Society program needed to use ICT in Sinhala and Tamil in order to develop content in Sinhala and Tamil, relevant to rural communities and the "grass-roots" as they termed their target groups. Before they could develop Sinhala and Tamil content, the computer operating systems had to support Sinhala and Tamil, Sinhala and Tamil fonts had to be available and keyboard drivers should be developed.

ALL NEED LOCAL LANGUAGES

All these teams set to work and started planning and developing their respective strategies. They had realized that to implement all of these and "to take the dividends of ICT to every village and every citizen", they all needed one thing, and that was the ability to use ICT in Sinhala and Tamil. This was the base.

Everyone realized that the ability to use ICT in Sinhala and Tamil was one of the foundations on which the rest of ICTA's work had to be built. Therefore, all at ICTA realized the importance of the Local Language Initiative and agreed that this area should be given priority and promoted.

The objective of the Local Language Initiative was to ensure that the benefits of ICT should be taken to the population in Sri Lanka, the majority of whom prefer to use ICT in Sinhala or Tamil. The Local Languages Initiative Team at ICTA comprised mainly three people at the outset, in 2003 – Prof. Gihan Dias, the Director, Ms. Aruni Goonetilleke the Program Manager and Ms. Phyllis De Alwis and later Ms. Mahiya Rafeek who provided Secretarial Support.

The Local Language Initiative (LLI) team immediately, at ICTA's commencement, reactivated the Local Languages Working Group. This was a Working Group which had functioned throughout CINTEC's lifetime under different names. This was a Group which could transition smoothly and work effectively under whatever host organization to which it was shifted and in which they sometimes found themselves. The Working Group comprised stakeholders and experts in languages and IT, from the University of Colombo School of Computing (UCSC), the University of Moratuwa, the Sri Lanka Standards Institution, the private sector and individual experts. The Working Group recommended policy and direction for ICTA's Local Languages Initiative, reviewed the work being carried out and contributed immensely towards the work that ICTA was carrying out.

The very first LLWG members, when work commenced at ICTA were Prof. V.K. Samaranayake, Ms. Amara Nanayakkara from the



Prof. V.K. Samaranayake

The e-Sri Lanka development program, setting up ICTA & the need for local languages

National Library, Prof. Gihan Dias, Dr. Ruwan Weerasinghe, Mr. Harsha Wijayawardhana, Prof. J.B. Disanayaka, Mr. Anura Tissera from ANCL, Mr. J. Kulasingham from Spectrum, Mr. G. Pratheepan from the University of Moratuwa, Mr. J. Yogarajah, Prof. Suseendrarajah, Mr. Lalith De Silva, Deputy Government Printer, and from ICTA there were Mr. Wasantha Deshapriya and Ms. Aruni Goonetilleke. Later, there were other members. Ms. Dineesha Ediriweera from ANCL, Prof. Sandagomi Coperahewa and Prof. Rohini Paranavitana, both from the University of Colombo joined the Group. Mr. Reshan Dewapura, who was initially a Program Director and later CEO of ICTA was also a member. LLWG was chaired by Prof. Gihan Dias, Prof. V.K. Samaranyake, and thereafter by Dr. R.B. Ekanayake. Dr. Ekanayake was a Board Director at ICTA. Prof. Epasinghe who chaired ICTA after Prof. Samaranyake passed away, was also very supportive of the Local Languages Initiative. Dr. Shahani Markus also oversaw the LLI arena during her two-year period as Chief Technology Officer of ICTA. Ms. Jayampathy Dewasurendra from the Sri Lanka Standards Institution (SLSI) contributed immensely towards the work that was carried out.

The Local Language Initiative consisted of several strata, each of which was to be built on the previous ones. These were; development of standards; development of technology and infrastructure; development and localization of software with overall and ongoing awareness, promotion and support.

The first LLWG meeting was held at ICTA on 12th January 2004. At this meeting the members decided that they should prioritize the development of Sinhala fonts and have a plan for disseminating the fonts to get people to use these. It was necessary to standardize the keyboard layouts and get local language keyboards developed. Keyboard drivers were necessary for input. It was necessary to hold awareness sessions for people developing local language websites. It was necessary to dissuade them from using non-Unicode legacy fonts; it was also necessary to decide on the Sinhala collation sequence with a view

to getting it standardized. As described earlier, in this regard it was agreed that Dr. Ruwan Weerasinghe and his team from the UCSC would submit a proposal. It was also necessary to develop converters to convert the existing fonts to Unicode fonts.

While the rest of ICTA was agitating about the need to do their work in Sinhala and Tamil, the Local Language team had been working all the while in ensuring implementation of everything that was necessary for using ICT in Sinhala and Tamil. It was not possible to get fonts to render perfectly overnight. To decide on a standard keyboard layout for Tamil, comprehensive and extensive consultation was required. (with regard to the Sinhala keyboard layout, consultation had been carried out by CINTEC). Therefore, in spite of the demand for using ICT in local languages to be enabled, time was needed to finetune and get everything to work well.

Mr. Wasantha
Deshapriya



One factor which created impetus and drove the Local Language team to speed up work was the constant and relentless demand made from the Re-engineering Government team, particularly from Mr. Wasantha Deshapriya who was the Director, Re-engineering Government. One trigger was an email from a member of ICTA's Communications Team

which he received around this time, pointing out the fact that the Sinhala section of the BBC website was in English. This was unacceptable.

Consequently, Mr. Deshapriya pointed out to Mr. Manju Haththotuwa, who was the CEO and to everyone else, that ICTA being the apex Government body on ICT, it was imperative that ICTA should without any delay enable the use of ICT in local languages. This produced what was a “chain-reaction”. Mr. Haththotuwa came up to Prof. Gihan Dias and Aruni Goonetilleke and kept asking when it would be possible to use ICT in Sinhala seamlessly without any problem. Telling Mr. Manju Haththotuwa that time was needed to get fonts to render, for keyboard drivers to be developed was of no avail. He was in a hurry. On retrospect, maybe this helped in speeding up the efforts to find solutions.



Mr. Manju Haththotuwa

Mr. Harsha Wijayawardhana, who was a member of LLWG, in the meantime, was an enthusiastic advocate of Unicode Sinhala. He had been working throughout in developing Unicode Sinhala fonts in his lab starting in 2001. His main contribution had been in trying to understand the rendering of Unicode Sinhala. Mr. Wijayawardhana worked hard during this period in trying to find out if Sinhala characters could be rendered using the Microsoft Windows rendering engine Uniscribe. Linux used several rendering engines such as Pango and QT. During this period, Mr. Wijayawardhana was involved in developing a Unicode Sinhala font using the glyphs of the Sinhala font Sarasavi. He experimented with Unicode font rules. Between 2002 and 2003, Mr. Wijayawardhana managed to render some of the font rules for Sinhala

using Uniscribe, except for the kombuwa. To render the kombuwa Microsoft's intervention was required; their rendering engine had to be patched so that the kombuwa appears before a consonant, although it would be stored in the sequence after a consonant, in the memory. Subsequently, Mr. Wijayawardhana wrote a paper theorizing on how to use the zero-width joiner, ZWJ and the non-zero width joiner, ZWNJ for rendering the rakaransaya, yansaya and touching letters. Mr. Wijayawardhana had been a strong advocate of encoding the rakaransaya, yansaya and rephaya. If the rakaransaya and yansa forms were not encoded, then Mr. Wijayawardhana suggested that bendi akuru (බැඳි අකුරු) or conjoined letters could be the default form. This idea had some following although Mr. Wijayawardhana subsequently agreed with Prof. Gihan Dias' proposal of using only ZWJ.

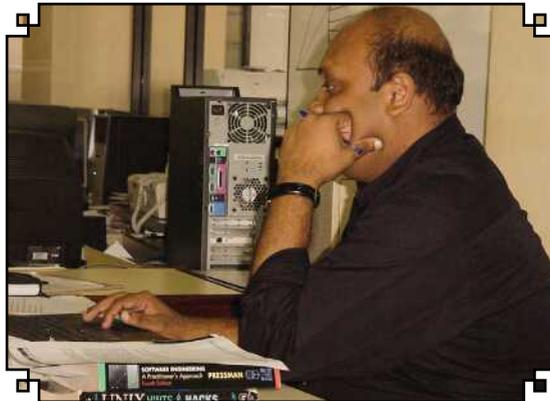
Mr. Wijayawardhana used Fontographer, a font editor for Windows, for this purpose. This was given to him by Mr. S.T. Nandasara, who had obtained a license for using it. The Sinhala script at this period was not in the list of scripts which were supported by the Windows rendering engine Uniscribe. Therefore, Sanskrit encoding was used. Discussions were held with teams from Microsoft. This was made easier by having a Sri Lankan, Mr. Prasanna Samarawickrama heading the Asian Desk at Microsoft, Redmond. Mr. Prasanna Samarawickrama managed to coordinate the Sri Lankan teams who were working on Sinhala fonts and the Microsoft team involved in developing Indic fonts for the Indian Subcontinent.

Then in December 2004, there was the Tsunami which devastated parts of Sri Lanka with an extensive loss of life and property. A team comprising many individuals and companies came together and contributed towards developing the Sahana FOSS Disaster Management System. This system included the development of a database consisting of the names of people who were missing, which would help in searching. Ideally this database had to be in local languages.

AWARENESS PROGRAMS

When ICTA commenced, there were many projects which had to be implemented and many things which had to be done. The standards had to be in place. The draft Sinhala ICT standard had to be approved and standardized by the Sri Lanka Standards Institution. Unicode compliant Sinhala fonts had to be developed. The Sinhala fonts had to render properly. Then there was the issue pertaining to the yansaya, rakaransaya, and rephaya which represent combinations of two or three letters, where the initial letter or letters are pure consonants. It had to be ensured that these do not break up.

At an early stage, Mr. Harsha Wijayawardhana recommended separate code points for each of these three, but this did not happen. He continued to advocate this at most meetings. This issue, later, became an unending problem, with ICTA receiving phone calls and emails from fractious users all over the island, about the yansaya, rakaransaya and the rephaya breaking up.



Mr. Harsha
Wijayawardhana

The typing sequences had to be defined, the delete sequences had to be defined, a keyboard driver had to be developed. It was also necessary to have a comprehensive dialog with Companies and groups developing operating systems.

Users had got used to particular typefaces. In several organizations there were specific fonts which were being used, such as “Abhaya”, “Amila” or

“Bindumathie”. The aesthetically pleasing Sinhala font Abhaya developed by Mr. Pushpananda Ekanayake was hugely popular. But at this stage, these fonts were not Unicode compliant. There was great resistance to switching away from the particular typeface which was being used. Therefore, an initial requirement was to make users and developers aware of the benefits of adhering to the international standard Unicode. It was necessary to introduce this encoding standard to those who wanted to use ICT in Sinhala and Tamil, specifically heavy users, such as the Parliament and the Legal Draftsman’s Department.



In parallel, font developers had to be dissuaded from developing legacy fonts. There were many individuals and companies at this time who were developing legacy fonts. Font developers had to be convinced of the overall long-term benefits of switching to the Unicode standard. This was not an easy task. The Local Language team had to carefully explain to users and developers that they would no longer have to attach fonts to emails if the recipient did not have the particular font, nor download fonts to view websites, if they were compliant

to this international standard. The intention was to ensure that using ICT in Sinhala and Tamil would be on par with ICT use in English.

Therefore, ICTA's Local Language team organized a series of workshops and awareness sessions on the necessity and benefits of using ICT in Unicode compliant Sinhala and Tamil. Most of these were held in the ICTA Auditorium. One such workshop was targeted to Font Developers and the resource person was Mr. Muthu Nedumaran, from Malaysia. In the early 2000s, getting resource persons from overseas helped font developers in Sri Lanka to augment their knowledge.

Some attendees who later became proficient font developers, stated that this particular workshop helped them greatly. Mr. Shanmugarajah, who extensively developed the keyboard drivers later for ICTA, attended this workshop. Members of the LLWG, particularly Dr. Ruvan Weerasinghe, Prof.



Dr. Ruvan Weerasinghe carrying out a presentation

Prof. Gihan Dias and Mr. Harsha Wijayawardhana conducted workshops for specific audiences. Font developers and users from Government organizations attended these sessions. Participants were from all over the island. A font developer from Anuradhapura, Mr. Nimal K. Wijesinghe, who developed beautiful legacy fonts attended this workshop. At first there was resistance, trepidation and even perhaps some resentment. Some developers even alleged that their livelihoods were at stake. But most were finally convinced. Font developers thereafter started being compliant to the Unicode standard and developing Unicode compliant Sinhala fonts.

Meanwhile, Mr. Wasantha Deshapriya, with greatly difficulty managed to dissuade Mr. Lalith Weeratunga, who was then the Secretary to H.E. the President, from using a legacy font with a particularly font-face. Mr. Weeratunga wanted no other font but this. This was the Thibus font which was initially not Unicode compliant. Therefore, Mr. Wasantha Deshapriya requested Mr. Sampath Godamunne from Science Land to convert this font to being Unicode compliant and explained the benefits of being Unicode compliant to Mr. Weeratunga. Mr. Weeratunga accepted this. He thereafter sent a Circular to all Government organizations stating that they should comply to the Unicode standard when using local Languages.

Awareness on using Unicode compliant Sinhala was also built when Mr. Nirajan Meegamma, who created kaputa.com, set up a Google Group in the mid-2000s, to read, write and communicate in Unicode compliant Sinhala. This Group communicated only in Unicode compliant Sinhala. The fact that they acted as a testbed for new implementations was greatly helpful to the work on local languages that ICTA was carrying out. The Group members used different operating systems and diverse technologies, but participated in very interesting discussions on diverse topics, writing to each other in Sinhala smoothly with hardly any perceptible issue. This Group also helped and acted as a buffer when there were detractors and disparagers. Detractors were many. There was the issue of one person requesting a code point each, for all Sinhala characters, combinations and variants. And many years later, another proposed a system for writing and displaying Sinhala using Latin characters as an alternate method for Unicode Sinhala on digital devices. According to this latter system the Sinhala Script was to be replaced with the Latin script. This technology, if it had been widespread would have led to the extinction of the Sinhala Script which has kept the uniqueness of the Sinhala language. In the long run, this technology would have led to Sinhala data to be stored in the Latin Unicode space on the Internet and in Database Management Systems. According to this person's technology, the display looked like Sinhala

on the surface, but it would have brought in key problems when handling digital Sinhala data. The above illustrates the fact that intricate and complex communication mechanisms were needed.

ANNUAL REPORT 2004

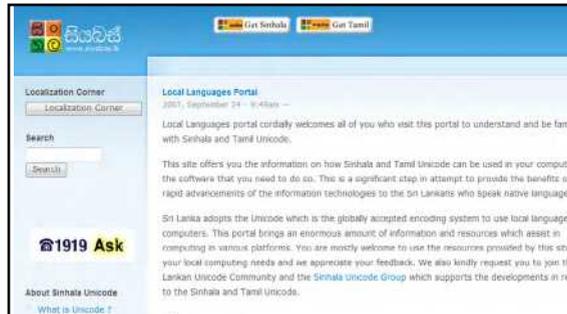
It is noteworthy that the Local Language Initiative was the area at ICTA which had significant project output and achieved outcomes during the first year of ICTA's existence and therefore had considerable progress to report to ICTA's very first Annual Report in 2004:

- “SLS 1134, revision 2: The second revision of the Sri Lanka standard Sinhala character code for information interchange was approved by the Sri Lanka Standards Institution, as a Sri Lanka standard.
- Sinhala kit for Microsoft Office: Work has been carried out to ensure that the β 2 version works well with Word XP and 2003 and PowerPoint XP.
- Unicode Compatible Sinhala Fonts: Unicode compatible fonts were available by the end of the year.
- IT words and phrases in Sinhala: A meeting held on planning out Sinhala translations of IT words and phrases.
- Sinhala upgrade for Internet Explorer (IE)6: The upgrade allows Windows users to view Unicode web pages.
- Sinhala Collation algorithms: A report has been prepared on Sinhala Collation Algorithms by the University of Colombo School of Computing.
- Sinhala Glossaries: Preliminary translation of the Microsoft list was carried out.
- Standardizing of the name of the language in ISO 639: The Standard is “Sinhala” with “Sinhalese” as a variant.”

LOGO USED IN WEBSITES

Mr. Harsha Wijayawardhana continued to work on Sinhala font rules, and specifically on the issue of the yansaya, rakaransaya and the rephaya breaking up. He had received formal training in Lahore, Pakistan, in mid-2000, under the PAN Localization project under the tutelage of Prof. Sarmad Hussain.

Around this time, he created the popular four-color logo “සිංහල ගන්න” using a font designed by Mr. Winnie Hettigoda. The four letters reading සිංහල were intended to identify Unicode Sinhala. The four colors of the Sri Lankan flag, gold, red, yellow and green are depicted in the logo. The logo was used in the local language portals siyabas.lk, emathumozhihal.lk and locallanguages.lk which ICTA developed in Sinhala, Tamil and English respectively.



The logo was subsequently used on most tri-lingual websites. The logo was used on the website on local language resources www.fonts.lk which was developed at this stage. ICTA used this logo on all the Government websites which were developed under its Re-engineering Government program. On these websites, the logo was linked to the keyboard driver which enabled keyboard input in accordance with the keyboard layout and keying-in sequences in SLS 1134.

FONTS

Kumara for some reason found it difficult to install the 30 MB Sinhala “kit” which included the keyboard driver and a Unicode Sinhala font, into his computer. He was using a Windows XP computer at this stage. And it took some time for Kumara to realize that he no longer had to use the function keys to type in Sinhala. He could now change the language in his computer with the keys Alt+ Shift from English to Sinhala and vice-versa. Then he no longer had to select a particular font to type in Sinhala. But he continued to seek Bindumathie.

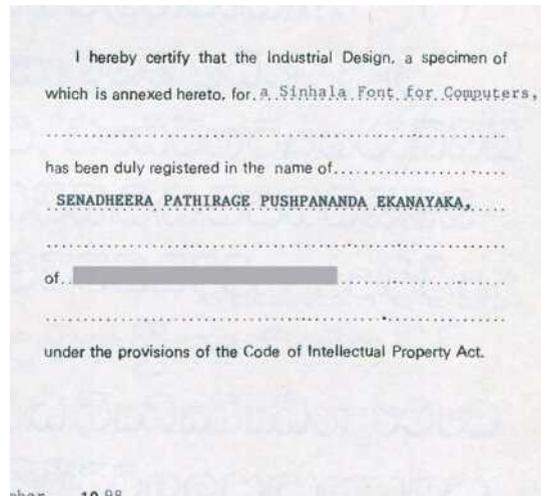
FontMaster (FM)

Mr. Pushpananda Ekanayake’s Sinhala font Abhaya was extremely well-known. It was the font that was used in Government organizations and by printers and publishers. He created the font Malithi in 1996. He launched ten Sinhala fonts on 26th September 1996 at the Mahaweli Center. Mr. Pushpananda Ekanayake says that there were no text fonts among these. Developing a text font is difficult. A reader should be able to read at least 500 pages without fatigue, when reading a document printed with a text font designed properly. Mr. Pushpanada Ekanayake registered his fonts as Industrial Designs at the Intellectual Property Office. Protecting a font from copyright infringement is difficult.

There were also cases of intellectual property infringements in relation to Mr. Ekanayake’s fonts. He found it extremely difficult at that time to explain to law enforcement personnel that Sinhala fonts or “letters” had been subjected to IPR infringement. Law enforcement officers could not, at that time comprehend how he could claim IPR to a set of Sinhala “letters”. These issues were addressed under the Code of Intellectual Property Act no. 52 of 1979,

The e-Sri Lanka development program, setting up ICTA & the need for local languages

and were sorted out satisfactorily. This case was discussed later in lectures for law students. Mr. Pushpananda Ekanayake registered the Company “The FontMaster” in 1999, for the purpose of designing and developing fonts. The large number of Sinhala FM fonts which were available and used widely were fonts developed by Mr. Pushpananda Ekanayake through this Company; 64 fonts were later developed.



Certificate of registration Bindumathie font

And it was Mr. Pushpananda Ekanayake who designed the beautiful and elusive Sinhala font Bindumathie.

Lihil

Another little-known initiative was carried out in the late 1980s; Mr. Jayantha de Silva developed an HP PCL Sinhala font named Lihil, with a phonetic keyboard driver. This was done when Bhikkhu Gangodawila Soma wanted to produce a newsletter “බොදු පුවත්” in Sinhala.

More stylized fonts

The font developer, Mr. Nimal K. Wijesinghe developed a portfolio of stylized ASCII fonts which included the aesthetically pleasing fonts AG-Diththala, AG-Panhida and AG-Sesatha etc.



Several Unicode fonts

When ICTA started operations, users were still using legacy fonts. ICTA wished to dissuade people from using non-Unicode legacy fonts. To do this, there had to be a large number of Unicode compliant Sinhala fonts. ICTA, therefore, at the inception, held extensive awareness sessions on the Unicode standard targeted to Font Developers. Several Unicode fonts were developed thereafter. With the release of Microsoft's Uniscribe shaping engine in 2003-5, Microsoft initially introduced the Unicode Sinhala font Potha which was later upgraded as Iskoola Potha. The font Iskoola Potha later became immensely popular. Some users even regarded it as a synonym for Unicode. During this period, ANCL (Lake House) created the Unicode Sinhala font Dinamina. This font was made available to ICTA with the understanding that it could be given free to users. With regard to the font MalithiWeb, the glyph was designed by Mr. Pushpananda Ekanayake and the rules were set by Prof. Gihan Dias. This was a Unicode compliant font. Mr. Pushpanada Ekanayake presented this font to ICTA to be given free to users. UCSC developed Sarasavi (Unicode), the Linux font LKLUG was available. Other Unicode fonts which were

developed were Science Land's Thibus (with keyboard input), Microimage's Helawadana kit with several Unicode fonts and Niranjana Meegammana's font Kaputa. All these fonts were the result of research carried out into creation of Sinhala and Tamil Unicode fonts.

ඉස්කෝලපොත

භාෂිතසැත්ප්

දිනමිණ

සරසවි

මලිකිවෙඩි

හෝඩිපොත

එන්-එයි-එච්

පුස්තකාල හොත

වර්ණ

විනි

විනි1

ව්‍ය.ව්‍ය.පුලකණ

Since many people had commenced developing Unicode Sinhala fonts, ICTA decided to develop a font which Font Developers could use as a guide. ICTA therefore commenced the development of a Sinhala font and make the font rules freely available to font developers. The Sinhala font was to be compliant with the standard SLS 1134 : 2004. It would enable the generation of accurate Sinhala documents using the full complement of characters facilitated in the standard SLS 1134 : 2004.

Developing such a font would propagate the use of aesthetically correct Sinhala character glyphs in Government documents and websites and would discourage the continuation of legacy fonts that were not compliant with the standard SLS 1134 : 2004. It would also enable old Sinhala and Pali books, written in the Sinhala script to be published on the Internet for worldwide audiences.

In order to meet the above targets, ICTA contracted with Ms. Dineesha Ediriweera who was working at ANCL, to develop the font. The font glyphs were designed by Mr. E.D. Pemasiri who had also been working at ANCL earlier. Mr. Pemasiri had developed fonts for the ANCL newspaper Silumina. The ICTA font was developed with the guidance of a Review Team and overseen by LLWG. The Review Team scrutinized the shape of each letter carefully. It was necessary to think of a suitable name for the font. Ms. Dineesha Ediriweera came up with the name “Bhashitha” for the font. This name ultimately became well-known.

The Sinhala font Bhashitha conformed to the Sinhala Standard SLS 1134 : 2004. It comprised two Unicode compatible Sinhala Font families, one serif and the other sans serif, with Regular and Bold fonts in each family. It was aesthetically correct. This font was launched in 2009.

Ms. Dineesha Ediriweera and Mr. Pemasiri together developed another Sinhala font, Hodipotha, for ICTA. This font was developed as a result of a request made by the National Institute of Education. It was a font meant for young schoolchildren in Grades 1 and 2 who are learning to read and write. One feature of this font was that the “eye” of letters such as ඌ had to be clearly discernible. In most fonts targeted to older people, the “eye” is just a smudge. Usually in fonts of sizes about 11 pt it is not possible for these details to be clearly visible. But small children have to learn to write such letters properly, in the correct way. And books meant for small children have large letters in which all the minutiae of a letter are visible. With this necessity in mind, the font Hodipotha was designed taking these aspects into account. ICTA carried out a demonstration for NIE and handed over the font.

ICTA thereafter ensured the development of the Tamil font SriTamil. ICTA contracted with Mr. S. Shanmugarajah (Shan) to develop the font

The e-Sri Lanka development program, setting up ICTA & the need for local languages

rules. The font rules of these two fonts, Bhashitha and SriTamil were given free for use, (with the condition that the glyphs and the rules are not to be changed).

Training on font development

After getting several fonts developed, it was decided that font developers should be trained to design and develop Unicode compatible Sinhala and Tamil fonts. The target groups were students, self-employed developers and the private sector. The training was held at UCSC in 2009 with an eminent team of resource persons. The team included Ms. Dineesha Ediriweera, Prof. Gihan Dias, Dr. Ruvan Weerasinghe, Mr. G. Balachandran, Mr. Harsha Wijayawardhana and Prof. Rohini Paranavitana.

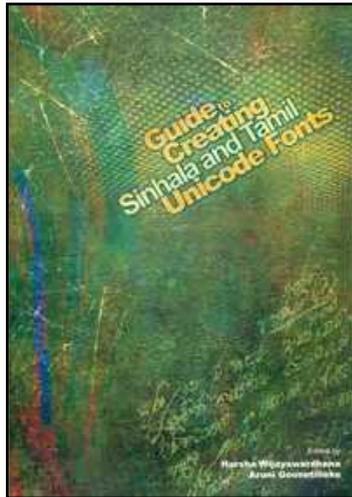
Areas of coverage included the Sinhala Script and Tamil scripts, essential features of the scripts, personality of fonts and their recommended use, typography of fonts, introduction to Open Type fonts, Font Families, Glyphs, features of Open Type Fonts, and selecting the glyph set based on use; Screen Fonts, Office use Fonts, SMS Fonts, using a Font development program – glyph size selection, naming conventions, laying out, composite glyphs and a hands on session on font developer software, cleaning up glyphs, widths, spacing (side bearings), fine tuning the accents and generating fonts.

The trainees designed and developed several stylized Sinhala fonts; Ms. Nirmali Allahendra developed the font Puskola Potha, Mr. Warna Somaratne developed the font Warna, Mr. Sushikshita De Silva developed the font SS Sulakna, Mr. Nirmala Handapangoda developed the font NYH and Mr. Winnie Hettigoda developed the fonts Winnie and Winnie1. The rules of the font Bhashitha were used in developing these fonts. The trainees agreed that ICTA could offer the fonts free to users. Therefore,

these fonts were available on the ICTA site www.icta.lk. These are now available on the LK Domain Registry's local language portal www.language.lk, which was designed, developed and is maintained by Ms. Chamali Perera of the LK Domain Registry.



www.language.lk



Output of the training, Guide to Creating Sinhala and Tamil Unicode Fonts

One output of the training was a guide on how to develop Unicode local language fonts which was published as an ICTA book. The book consists of the presentations that the trainers made at the training session. A hard copy was published and it was also made available free of charge on the website language.lk. Compiling the presentations and making it available as a book was Mr. Harsha Wijayawardhana's idea.

ICTA later contracted with Mr. Shanmugarajah and completed a Sinhala font and a Tamil font, AdBhashitha and AdSriTamil respectively, which work on Adobe publishing applications. After this development, it was not necessary for the book designer to re-type ICTA's Annual Report 2014. He used Adobe InDesign into which he copied the text from the original Annual Report in Sinhala, Tamil and English, which was written on Word.

characters ෂ and ් which are not commonly used and six sets of touching letters which are described in the standard. A level 3 font should also support all combinations of strokes with conjuncts.

Font Converters



The Language Technology Resource Laboratory (LTR) developed a font converter to convert non-Unicode text to Unicode text. This was made available on the web in 2006.

It facilitated several Sinhala fonts and one Tamil font. Using this converter, it was also possible to convert Unicode compliant Sinhala text to non-Unicode text; sometimes printers and publishers needed conversion in this direction. LTR also developed an offline converter for converting non-Unicode text to Unicode text, for Sinhala and Tamil fonts.

SRI LANKA STANDARD SINHALA CHARACTER CODE FOR INFORMATION INTERCHANGE, SLS 1134 : 2004

The standards were developed in a participatory manner. ICTA, in early 2004, partnered with other organizations, specifically the Sri Lanka Standards Institution, the University of Colombo School of Computing, the University of Moratuwa and language experts in developing the second

The e-Sri Lanka development program, setting up ICTA & the need for local languages

revision of SLS 1134, following on the work carried out on this area by CINTEC. Prof. Gihan Dias who was heading the Local Language Initiative



Prof. Gihan Dias Program
Director ICTA, carrying
out a presentation on local
language keyboard layouts

at ICTA worked with the LLWG on developing this second revision of the SLS 1134. LLWG discussed the draft extensively and thereafter forwarded it to SLSI. Ms. Jayampathy Dewasurendra, from SLSI followed up on the standardization process. SLSI carried out public consultation and notices were published in

newspapers in all three languages. Comments from the public were taken into account. Thereafter the draft was standardized as the Sri Lanka Standard Sinhala Character Code for Information Interchange, SLS 1134 : 2004.

The standard consisted of the encoding for Sinhala compliant with the Unicode standard and with the Universal Character Set, ISO/IEC 10646; the Sinhala keyboard layout based on the Wijesekera layout and the keying in sequences based on the “type-as you-write” method, and the code sequences. The Sinhala Collation Sequence was Part 1 of the Standard.

SRI LANKA SINHALA CHARACTER CODE FOR INFORMATION INTERCHANGE, SLS 1134 : 2011.

It was Mr. Michael Everson, a contributing editor to the Unicode Standard, and to the standard ISO/IEC 10646, (the Universal character set) who first

kindled interest in Sinhala numerals among the ICT community in Sri Lanka. He first submitted a proposal on the inclusion of Sinhala numerals in the Sinhala character set in the Unicode standard in the mid-1990s. CINTEC was made aware of this. Consequently, Professor J.B. Disanayaka and Mr. S.T. Nandasara of the UCSC attended the relevant meeting through CINTEC and managed to persuade Michael Everson to postpone the inclusion of Sinhala numerals in the Unicode standard.

About a decade later, after CINTEC was closed, ICTA was made aware that Mr. Michael Everson had re-submitted a paper to propose the inclusion of old Sinhala numerals in the Sinhala page of the Unicode standard. The Local Languages Working Group discussed this new development and was of the view that there was no agreement on what Sinhala numbers are and how they should be represented. ICTA needed time to carry out research on this. This was an initiative which would impact users of the Sinhala language throughout the country and elsewhere. Therefore, ICTA requested Prof. J.B. Disanayaka, who was then the Ambassador of Sri Lanka to Thailand, to again attend the relevant meeting of the Unicode Technical Committee, which was held in Cupertino, California and hold the proposal in abeyance, so that LLWG members could gain time to study the subject. LLWG discussed this issue extensively. The Local Language Initiative was by now within the area “ICT Policy, Leadership and Institutional Development”, and Mr. Jayantha Fernando was the Director. He was greatly supportive of this venture.



Mr. Jayantha Fernando

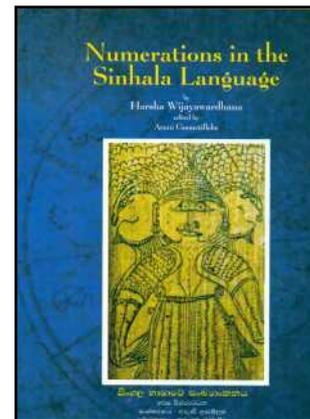
The e-Sri Lanka development program, setting up ICTA & the need for local languages

ICTA contracted with Mr. Harsha Wijayawardhana to conduct research on the subject of Sinhala numerals. Mr. Wijayawardhana studied papers and publications on old Sinhala numerals and original documents which had some form of numerals or numerations. He visited the Vatadageya in Polonnaruwa, visited the Colombo Museum and the Museum at Sigiriya, studied rock inscriptions and studied Ola leaf page numbering.

This extensive research yielded the fact that five different types of numerations were used in the Sinhala language:

1. Sinhala Numerals or Sinhala Illakam
2. Lith Illakam or ephemeral numerals
3. Katapayadia
4. Page numbering of Ola leaves using Sinhala “Swara” (ස්වර)
5. Bhootha Anka or Butha Sankya

Of these, two types of numerations had been widely used; one set (Sinhala Illakkam) was in use up to the early part of the nineteenth century, and the other set (Lith Illakkam) was in use well into the twentieth century. Sinhala Illakkam had been used to express simple numbers. They had separate symbols for 10, 40, 50, 100, and 1000 etc. Lith Illakkam had been generally used for casting horoscopes and for carrying out astrological calculations. This set had also been used for numbering pages of Ola palm leaf books. Numbers in Lith Illakkam are depicted by Sinhala letters and vowel modifiers. Lith Illakkam includes a zero. The zero is depicted by the Halantha or Hal lakuna (kodiya) of the Sinhala language. In Lith Illakam, numbers greater than zero were



written in the same way as Indo-Arabic numbers, where the value of the number on the left was increased by ten with the zero. Concisely, Lith Illakkam has a zero and a zero place-holder concept.

The findings were presented to members of the Royal Asiatic Society. Mr. Wijayawardhana participated over the phone in the Unicode Technical Committee meeting held in August 2010. He also participated at ISO's Working Group 2 meeting held in South Korea. Mr. Wijayawardhana, thereafter worked with LLWG on developing the draft third revision of SLS 1134 which includes encoding for Sinhala numerals. The draft was forwarded to the Sri Lanka Standards Institution. SLSI put it through the usual approval process, sought public opinion, published advertisements in newspapers and considered public comments. Thereafter, the third revision of SLS 1134 was approved as the Sri Lanka Sinhala Character Code For Information Interchange, SLS 1134 : 2011, in July 2011. Sinhala numerals were also included in the Unicode standard and in ISO/IEC 10646; Lith Illakkam, (which includes the zero concept) were included in the main Sinhala code page, and Sinhala Illakkam were included in the supplementary area of the Unicode standard and in ISO/IEC 10646.

THE SINHALA COLLATION SEQUENCE

The Sinhala Collation Sequence, developed by Dr. Ruvan Weerasinghe and his team from UCSC, was approved by ICTA's Local Language Working Group, as described earlier. This was forwarded to the Sri Lanka Standards Institution. Subsequently, it was approved by the Sri Lanka Standards Institution as a Sri Lanka Standard. This Sinhala collation sequence was standardized as Part 1 of SLS 1134 : 2004.

The e-Sri Lanka development program, setting up ICTA & the need for local languages

TESTING PRODUCTS FOR COMPLIANCE WITH SLS 1134 : 2004

ICTA, through LLWG, and in partnership with SLSI, developed a standard Test Method for Sinhala ICT products through which organizations could get their Sinhala ICT products, such as keyboard drivers, Unicode Sinhala fonts, and keyboards tested. Thereafter these products could be certified as being compliant with the standard SLS 1134 : 2004 and an SLS mark could be obtained. This standard, “Part 2 of the standard SLS 1134 : 2004 – Requirements and Methods of Test” was approved by the SLSI as a Sri Lanka Standard. The scope was for computers only. But unfortunately, due to numerous reasons, this standard was not used. In retrospect, ICTA should have perhaps tried harder on the implementation aspects, rather than focusing mainly on getting the standard developed.

SRI LANKA STANDARD TAMIL CHARACTER CODE FOR INFORMATION INTERCHANGE, SLS 1326 : 2008

Mr. Gnanasekaraiyer Balachandran, under his contract with ICTA, had worked on a Tamil keyboard layout and on the Tamil collation sequence. These were accepted by users and formally accepted by LLWG. It was now necessary to get these included in a Tamil ICT standard. Therefore Mr. Balachandran with LLWG and stakeholder participation developed a draft document comprising encoding for Tamil characters, the keyboard layout and a collation sequence for Tamil. The Sri Lanka Standards Institution standardized this in September 2008, as the Sri Lanka Tamil Character Code for Information Interchange, SLS 1326 : 2008. The standard was launched at an event which ICTA organized for key stakeholders in the local language arena, on 24th November 2008. The standard comprised; encoding for Tamil; the Tamil keyboard layout and keying-in sequences; and the Tamil Collation Sequence.



Handing over
SLS 1326 to the Hon. Minister

Mr. Reshan Dewapura CEO, ICTA
at the launch of SLS 1326



ENCODING FOR TAMIL

The encoding for Tamil in the standard is based on the Unicode standard Version 5.1 and ISO/IEC 10646 : 2003. Character code encoding / sequences are provided for the following.

- Vowels
- Pure-Consonants
- Vowel-Consonant syllable
- āytam ஃ,
- Grantham Conjunct syllable - ழுழு
- Tamil numerals
- Tamil symbols
- Tamil OOM, ழு

THE TAMIL KEYBOARD LAYOUT AND KEYING IN SEQUENCES

ICTA, at the very outset, with the recommendation of LLWG, set up a Subcommittee of the LLWG to specifically address issues relating to using ICT in Tamil. This Subcommittee's ultimate objective was to develop a Tamil ICT standard equivalent to the Sinhala ICT standard SLS 1134 : 2004. The code chart for Tamil compliant to the Unicode standard was agreed on.

Thereafter the Subcommittee had to recommend a suitable keyboard layout. After several meetings and discussions, the Subcommittee agreed that the Tamil99 keyboard layout was suitable and efficient. The Subcommittee mainly consisted of ICT professionals who deemed that it was most logical and consequently, that it was suitable. From a logical standpoint it was a most efficient layout. It was said to be a "phonetic" layout. The keyboard layout was presented to users. A team from ICTA with members of the Subcommittee even travelled to Jaffna in May 2004, to present the keyboard to users. This was not easy during those difficult times in 2004, when the country was

torn with terrorism. The workshop was held at the Jaffna Hindu College. At this workshop, most participants agreed that the layouts Renganathan or Remington are widely used, but in the long run “Romanized” keyboard layouts may be suitable. Some participants expressed views that the layout Tamil99 is good but there are some places in the layout where “people can go wrong”.

The keyboard layout was also presented to Government employees and heavy users. This layout was rejected outright. The “phonetic” keying in was not popular. Users did not wish to take time to figure out which combination of keys would yield the desired result. Thereafter, working on a Tamil keyboard layout and a Tamil ICT standard stalled for several years.

ICTA, at this time was setting up the Lanka Government Network (LGN). The LGN team implementing this project kept asking what the standard Tamil keyboard layout was, and also requested the keying-in sequences to be recommended to Government organizations which were to be connected to the LGN. They also wanted a keyboard driver to be made available.

Mr. Gnanasekaraiyer
Balachandran presenting
on the Tamil Keyboard
Layout



These issues were sorted out in 2006 and 2007. ICTA set up a team consisting of people who were proficient in the Tamil language, linguistics, and users to work on the keyboard layout. And ICTA contracted with Mr. Gnanasekaraiyer Balachandran to work with this team to decide on a Tamil keyboard layout, carry out consultation and to develop a draft Tamil ICT standard. From this point onwards the work on ICT and Tamil got sorted

The e-Sri Lanka development program, setting up ICTA & the need for local languages

out. Mr. Balachandran extensively checked on the available Tamil keyboard layouts and presented his work to LLWG and LLWG's recommendations were taken into account.

Mr. Balachandran and the team first held a workshop in October 2006 for stakeholders at which representatives from the Government sector were present. The consensus was that they wish to use the Renganathan keyboard layout. The team thereafter conducted extensive research, had meetings with other key stakeholders and it transpired that the Renganathan layout has about 10 variations or more. After further analysis and stakeholder consultation, a keyboard layout based on Renganathan, on which some modifications were carried out, was selected. All possible Tamil letters and symbols were included in this layout. The layout did not deviate significantly from the Renganathan and Bamini layouts. Keying-in sequences were defined on the "type as you write" method. The proposed layout was also presented to a wider audience in January 2007, with a keyboard driver for testing. The audience consisted of users from ICTA, the University of Moratuwa, the University of Kelaniya, the Legal Draftsman's Department, the Parliament, the Sri Lanka Standards Institution, the Ministry of Public Administration, users from Nenasala centers in various parts of the island, and from the Associated Newspapers of Ceylon Ltd. (ANCL). The participants accepted the proposed layout and there was no dissension.

ICTA thereafter contacted Mr. Mahesh Perera at the Parliament and arranged a demonstration for users at the Parliament. Carrying out a presentation at the Parliament was not easy during those difficult times; every piece of equipment that was taken in and every person who entered the Parliament was carefully checked. But it was necessary to obtain the views of the users at the Parliament before finalizing the layout. The demonstration was done with a demo keyboard driver to users in the Parliament. The proposed keyboard layout, mainly based on the Renganathan keyboard layout, was unanimously accepted

by the users. This layout was consequently included in the Sri Lanka Standard Tamil Character Code for Information Interchange, SLS 1326 : 2008.

THE TAMIL COLLATION SEQUENCE

The Tamil collation sequence was needed for compiling dictionaries, other scholarly works, use in data processing and sorting lists of personal data and other information. The following elements, contained in the contemporary Tamil script were included:

Vowels, e.g. அ, ஓள,

Tamil pure-consonants, e.g. க், ங்,

Grantha pure-consonant, e.g. ஸ், ஷ், ஜ்,

Tamil and Grantha vowel-consonant

syllables, e.g. க், செ, ஸ, ளே, ழ,

ஃ (F) and vowel-consonant syllables of ஃ

Special letter, āytam – ஃ,

Conjunct syllable - (Sri) - ழ்,

Tamil Symbols, e.g. உ, டீ. ள்,

Tamil numerals, e.g. ச, கூ, ய, ற, சூ

TESTING PRODUCTS FOR COMPLIANCE WITH SLS 1326 : 2008

A standard Test Method for Tamil ICT products was made available through which organizations which have developed Tamil ICT products such as keyboard drivers, Unicode compatible fonts, and keyboards etc., could get these certified as being compliant with the standard SLS 1326 : 2008. This standard, “Part 2 of the standard SLS 11326 : 2008 – Requirements and Methods of Test” was approved by SLSI as a Sri Lanka Standard in July 2011. This standard too, like the Sinhala standard was not used.

The e-Sri Lanka development program, setting up ICTA & the need for local languages

SINHALA AND TAMIL “KIT”

A Sinhala and Tamil “kit” was created incorporating standard Sinhala fonts, keyboard driver and installation instructions, which was made available free of charge on the site www.icta.lk (under Local Languages Initiative).

LOCAL LANGUAGE PHYSICAL KEYBOARDS

ICTA promoted the development of physical keyboards based on the standard Sinhala and Tamil keyboard layouts which were made available in the market. Users had been sticking transparent Sinhala letters onto the English letters on the physical keyboards. And there was some concern on whether it would be possible to have tri-lingual keyboards. Some people were of the view that it was not possible to print characters pertaining to three languages on the keys. The keys were supposed to be too small. But this was done. ICTA had extensive discussions with Companies such as SALA Enterprises. The result was that the Sinhala and Tamil characters on the keyboard were in different colors.

LOCALE INFORMATION FOR SINHALA AND TAMIL

To support users of different languages, software has to be adapted to the conventions specific to that language, script, region and country etc. “The Unicode Common Locale Data Repository (CLDR) provides key building blocks for software to support the world’s languages. CLDR is by far the largest and most extensive standard repository of locale data. This data is used by a wide spectrum of companies for their software internationalization and localization”. www.unicode.org

ICTA implemented two projects on locale information – one for Sinhala and the other for Tamil. Locale information includes localized date and time formats, currency formats, calendars, collation order etc. The project entailed collecting and documenting the national preferences and use with regard to Sinhala locale information and Tamil locale information respectively, to maintain locale information pertaining to each language in Sri Lanka, and to define, collect and agree on and make available through the Unicode Consortium’s Common Locale Data Repository the common, necessary locale data for the Sinhala language. The locale for Sinhala Si_LK was available. But ICTA had to send an email to the Unicode Consortium to get a locale developed for Tamil-Sri Lanka. The locale for Tamil-India which was Ta_In was available. A locale Ta_LK was created on condition that whatever information common with Ta_In is “inherited” into Ta_LK.

USER INTERFACES IN SINHALA AND TAMIL

ICTA signed a Contract with Microsoft Corporation in August and September 2006 for implementation of the project “Language Interface Pack (LIP)” on the localization of Windows Vista and Office 2007. ICTA then contracted with the alliance Science Land and UCSC on translating from English to Sinhala. The translated files were made available to Mr. S.M. Banduseela, whom ICTA appointed as moderator for linguistic review

This project comprised the translation of the user interfaces such as the drop-down menus, from English to Sinhala, and creation of a language Style-guide which provided guidelines on the tone of localization. The Style-guide included provision of date formats and time formats, currency formats, eras, months, calendar preferences, week conventions, etc. For example, in early localization efforts before the Style-guide was produced, the time in Sinhala on Windows appeared as “2.30 ප.ව.” when it should have been “ප.ව. 2.30”.

The e-Sri Lanka development program, setting up ICTA & the need for local languages

Microsoft Corporation also sent documents termed “Polichcek” to ICTA; these were lists of terms that should be avoided and not be included in Windows or in MS Office, such as terms which were obscene, offensive, profane or politically incorrect. Microsoft made these lists available in English with the corresponding Sinhala equivalent in the Sinhala script. These were sent over to ICTA and ICTA’s LIP team reviewed and confirmed whether each word or term was to be accepted or rejected and informed Microsoft.

The project also included the development of a glossary of about 2,000 terms. This glossary was printed by Microsoft as a small booklet and there was a big demand for this. The completed product was handed back to Microsoft Sri Lanka and to Microsoft Corporation in May 2007. With the glossary Microsoft made a transparent sheet of the standard keyboard layout, from which users could take out the Sinhala characters and stick these on the keys of the English QWERTY keyboard. People who were not too conversant with the Sinhala keyboard layout during this period found these useful.

Prof. V.K. Samaranyake who was the Chairman of ICTA, together with several members of the LLWG and ICTA staff handed over the files containing the project output to officers from Microsoft Sri Lanka, on 30th May 2007. This was probably the very last activity in the local language arena in which Prof. Samaranyake participated (after working extensively on this area for more

than 20 years). He passed away a few days later in early June 2007.



Handing over the project LIP to Microsoft; L to R Mr. Jayantha Fernando, Mr. Anura Tissera, Ms. Aruni Goonetilleke, Prof. J.B. Disanayake, Prof. V.K. Samaranyake, Mr. Sriyan de Silva Wijeyeratne, Dr. Shahani Markus, Mr. Reshan Dewapura & Mr. S. Shanmugarajah.

INTERNATIONALIZED DOMAIN NAMES (IDNS)

LLWG, in March 2008 approved the setting up of a Task Force to work on Internationalized Domain Names (IDNs). The Task Force was to initially have comprehensive discussions on what should be the equivalent to LK in Sinhala and Tamil and discuss implementation. Consequently, the Task Force on IDNs was established and commenced work in May 2008. The Task Force on IDNs comprised representation from key stakeholders including the CEO of the LK Domain Registry, ICTA, the Telecommunications Regulatory Commission of Sri Lanka (TRCSL), the Department of Official Languages, the University of Colombo School of Computing, ANCL and two key ISPs - Sri Lanka Telecom and Eureka. It was agreed that when Sinhala and Tamil IDN Top Level Domains are implemented and made available, these will have an impact on all local language users. Sri Lanka required two IDN top level domains: one in Sinhala and the other in Tamil. Both languages have official status in the country.



Therefore, it was necessary to agree on the IDN top level domains in Sinhala and in Tamil. The Task Force agreed that Sri Lanka was willing to support the Fast Track procedure proposed by ICANN. The IDN Task Force was to be the forum which would determine the official position with regard to IDN ccTLDs for Sri Lanka. It was necessary for Sri Lanka's ISPs to agree to support

The e-Sri Lanka development program, setting up ICTA & the need for local languages

IDNs. The Task Force further agreed that only Sri Lanka should register domains in the Sinhala script. A public consultation process was followed. Advertisements were placed in national newspapers calling for those interested to attend a workshop held in August 2008. Specific invitations were also sent to stakeholders. The public consultation workshop was held on 6th August 2008, at which the IDN Task Force was able to ascertain the views of a wider representative gathering. Views of stakeholders and those interested were also collected through a questionnaire. The outcome of the public consultative workshop was taken into account. The responses to the questionnaire that had been sent out to stakeholders, were reviewed.

After extensive discussions the IDN Task Force unanimously agreed that the IDN for Sri Lanka in Sinhala would be *.ලක* and the IDN for Sri Lanka in Tamil would be *.இலங்கை*. Mr. Chamara Disanayake of the LK Domain Registry tested IDN ccTLDs in Sinhala and Tamil on browsers and with a few ISPs, as a pilot project. These domains are now being registered by LKNIC.

DEPLOYMENT

The areas that were necessary for local language computing were addressed. It was possible to use ICT in Sinhala and Tamil smoothly, to a great extent. There were a few glitches; there was the issue of the yansaya, rakaransaya, and the rephaya breaking up. Mr. Harsha Wijayawardhana continued to work on this to sort out this issue. One key impediment which a user encountered during this period was the necessity to install a keyboard driver to type in Sinhala and Tamil in accordance to the two SLS standards. The ultimate objective was for a user to be able to sit at a computer and start using it without doing anything additional such as installing software, which was a deterrent. It had to be possible to use a computer in Sinhala and Tamil in the same way in which a computer is used in English.

Therefore, the LLWG communicated with Mr. Indika De Zoysa who was the Country Manager of Intel. A meeting was held with him at which he agreed to facilitate deployment of the Sinhala and Tamil keyboard drivers through the computer vendors. The target was to ensure that when a computer is sold the Sinhala and Tamil capability – ideally out-of-the-box – is available for the purchaser. The overall deployment plan was to target Corporate users, the Government sector, and also home users. The program was to be deployed in partnership with the private sector, to inform the private sector that it was possible to use ICT in Sinhala and Tamil, and to secure their participation in deploying local languages.

Through this program computer vendors were made aware of the availability and benefits of using ICT in standards based local languages. The vendors were to pre-load computer systems with the Sinhala and Tamil keyboard drivers with the fonts. (Sinhala and Tamil “kits”). LLWG also discussed the possibility of making available standards-based engraved keyboards in the market and the inclusion in Government tenders for pre-installed standards-based local language products.

Therefore, the core solution which included the Sinhala and Tamil “kits” for Windows Vista and Windows XP - the Windows operating systems which were available and being used during this period - comprising the relevant keyboard drivers and Unicode compliant fonts, were made available to vendors on CDs. The vendors were expected to install these into the computers being sold, with the buyers’ concurrence.

The e-Sri Lanka development program, setting up ICTA & the need for local languages

The circular no. SP/SB/01/07, dated March 04, 2007, was circulated to all Secretaries of Ministries and Heads of Departments. It was signed by Mr. Lalith Weeratunga, Secretary to the President. The requirement was that all State organizations should adhere to the Sri Lanka Standard, SLS 1134, which in turn meant that State organizations should be Unicode compliant when working in local languages. Three more circulars on this subject, signed by Prof. V.K. Samaranayake, by Prof. P.W. Epasinghe and Mr. Wasantha Deshapriya were circulated.; ICTA was determined that State organizations were compliant to the Unicode standard.

But, if ICTA were to follow up on whether other organizations complied, then ICTA had to first ensure that its staff adhered to the Unicode standard. Therefore, these were circulated to all ICTA staff, including Kumara. Kumara read these carefully. The contents were clear and precise and he knew what he had to do. But at unexpected moments the font Bindumathie entered his mind. He tried to stay in focus, pressed left-hand Alt+shift and used Sinhala, not heeding the font that came up, which probably was Iskoola Potha.

We have come a Long Way

Chapter 6

We have come a Long Way

Now we can type in Sinhala and Tamil, exchange information in Sinhala and Tamil using computers, and browse the web in Sinhala and Tamil without having to download various fonts. Standards for Sinhala and Tamil are in place. Sorting in Sinhala and Tamil has been standardized. There are Sinhala and Tamil fonts available. When we use a word processor, we do not have to painstakingly select a Sinhala font the way we used to do. There is no need to select a font, the content is displayed as Sinhala. When we use a word processor, we do not need to install specific fonts to read the content. If the document had been sent by someone else, then we do not need the specific font that the sender used - if it is in Sinhala, then it will be displayed as Sinhala. The content on websites in local languages is displayed as Sinhala and Tamil content, rather than being seen as undecipherable symbols. We do not have to download fonts to view Sinhala and Tamil content on the web. Domain names have been registered in Sinhala and Tamil. There is no necessity to send emails with Sinhala fonts attached. Mobile phones are being used in Sinhala. Text messages are being sent in Sinhala. People view websites on mobile phones in local languages. Newspapers are available in Sinhala and Tamil on the web. Social media is being used in Sinhala and Tamil. Social media in Sinhala has been used to influence other users and bring about changes. It is possible to search in Sinhala. A search on any topic in Sinhala usually yields many results. And a phenomenon which we did not foresee, but nevertheless demonstrates how the availability of technology triggers innovation, is the multitude of Sinhala and Tamil blogs. There are blogs on many a topic, some are serious blogs, others are more frivolous. Many young people seem to be writing blogs. Now anyone can work with anyone else.

And teams at ICTA, which were vocal at the start about local languages, were now peacefully carrying out their work in Sinhala and Tamil. The Re-engineering Government team communicated in Sinhala with their stakeholders. The first Unicode compliant tri-lingual website www.gov.lk was

developed by e-Fusion (Pvt) Ltd, as a project under ICTA's Re-engineering Government program. Tri-lingual websites were being developed for Departments and Ministries and for Divisional and District Secretariats. Tri-lingual databases were being developed. The e-Society team was profusely developing local language content. There were issues which were not resolved, such as publishing software not fully supporting Sinhala. Due to this problem, it has not been possible for publishers to be Unicode compliant.

When the fundamental issues were being sorted out users could work in local languages and vociferous critiques, to a great extent were getting lulled, then with the respite it was possible for ICTA to address issues concerning people with impairments and disabilities. There was a necessity to ensure that assistive technologies in local languages were developed. A Sinhala and Tamil text to speech system would aid the visually impaired, so that a spoken voice would read to the user. It would be of great use to the elderly population and senior citizens who, for example, often find it difficult to keep a cursor still and click on a link due to shaking hands. It was therefore necessary to develop assistive technologies.

Therefore, in 2016, ICTA decided that a text to speech system for Sinhala and for Sri Lanka Tamil should be developed which would facilitate access to digital content for people with visual impairments, and the elderly. An integrated product would be developed as a screen-reader, a program that helps a blind or low-vision person use a computer.

A speech to text software for Sinhala was also to be developed. The objective was that when this was finalized it would be initially deployed at the Ratmalana School for the Blind. The organization that was selected for development of the software was UCSC. LTRL of UCSC had previous experience carrying out research and in building TTS systems for Sinhala.

UCSC used the Mary TTS, an open source Text-to-Speech Synthesis platform to develop Sinhala TTS voice and integrated it with the NVDA (Non-Visual Desktop Access) screen reader. Screen readers work closely with the computer's Operating System to provide information about icons, menus, dialogue boxes, files and folders. A screen reader uses a TTS engine to translate on-screen information into speech, which can be heard through earphones or speakers. UCSC's initial model under this project showed some promising results, and the quality of the voice was being improved so that it sounded natural and not robotic.

In parallel with the development of text to speech tools, an optical character recognition (OCR) system for Sinhala was also to be developed. Printed material could thus be scanned into a device and converted to editable digital text. Thus, more content could be made available to blind people. ICTA contracted with Theekshana, which was the consultancy body of the UCSC to develop an OCR system for Sinhala, to facilitate the development of digital content. There were old documents in Government organizations which had to be preserved and sometimes confidentiality of the documents had to be ensured. Government organizations needed an OCR system. It was expected that through the OCR system it should be possible to detect Sinhala characters in images and in PDF documents. It was expected to recognize Sinhala characters in conformance to Level 3 Fonts defined in the standard Sri Lanka Sinhala Character Code for Information Interchange, SLS 1134; Part 2 : Requirements and Methods of Test. The ultimate idea was for these products, which ICTA deemed should be made available as a service, to be made freely available country-wide.

EPILOG

Kumara was not struggling with fonts anymore. He was requested to help out with the ICTA Annual Report. This was one continuous word processing document, with the report in Sinhala, Tamil and English. All three languages were there smoothly and well-displayed in this same document. Kumara had now realized that three languages could be displayed correctly in the same document on the same page. He helped out with the Sinhala section. He could edit it easily. There was no need to block the Sinhala text and select a Sinhala font. It was emailed to him for follow-up action, by the person who compiled and wrote the report. He did not have to check whether he had the same Sinhala font which she had used. It was not necessary. He could switch the language in his computer to Sinhala and type in Sinhala. He found it easy to use. He used whatever Sinhala font that was available when he switched his computer to Sinhala. And he was no longer gripped by the Sinhala font Bindumathie. The memory of Bindumathie gradually faded. He had adjusted.



Acknowledgments

All those who participated in enabling the use of Sinhala and Tamil in ICT, whose names are mentioned in this book are gratefully acknowledged.

Those who contributed towards this arena and who are inadvertently not mentioned are also appreciated.

All members of the Local Languages Working Group are gratefully appreciated for their invaluable contribution throughout the years.

The former Chairmen, CEOs, Directors and the Leadership Team of the ICT Agency of Sri Lanka who fully supported ICTA's Local Languages Initiative are greatly appreciated.

The CEO of the LK Domain Registry (LKDR) and the review team set up by the LKDR comprising Ms. Nayeni Fernando, Mr. Wasantha Deshpriya, Mr. Chandana Weerasinghe, Mr. Harsha Wijayawardhana, Ms. Dineesha Ediriweera, Mr. S.M Banduseela, Mr. Rohan Manamudali provided great assistance.

Lastly Ms. Janakie Mampitiya's and Mr. Wasantha Deshpriya's comprehensive and meticulous final reviews were of immense value.

References

1. Development of Standards for Sinhala Computing
Gihan Dias & Aruni Goonetilleke, ICT Agency.
1st Regional Conference on ICT& E-Paradigms, June 2004, Colombo,
Sri Lanka
2. An Introduction to Unicode for Sinhala Characters, UCSC Technical
Report 03/01, UCSC, 2003
V.K. Samaranayake, S.T. Nandasara, J.B. Disanayaka, A.R. Weerasinghe,
H. Wijayawardhana
3. Research Papers on Natural Language Processing, Language Technology
Research Laboratory, UCSC

Aruni Goonetilleke's book takes me back almost twenty-five years, tracing the long and arduous journey that Sinhala and Tamil computerization had taken. Aruni has succinctly managed to document most landmark events in enabling Sinhala and Tamil on computers and subsequently on all digital devices. She also, in her inimitable style narrates the story of a struggle of a group of individuals who were passionate and volunteered to make Sinhala and Tamil computing possible over the years. In this story, she tells us how the development of original 7-bit Sinhala fonts took place, which initiated Sinhala computing many years ago to the latest True Type Unicode Sinhala fonts, while patching most rendering engines which are used today. Owing to the research carried out for the encoding of Sinhala in Unicode, the existence of hitherto forgotten Sinhala Numerals surfaced, showing the proud history of Sinhala mathematics and prowess of Sinhala mathematicians, unraveling them as one the first users of a symbol for zero to the world. So, it is a book that must be read by all who have an interest in finding how Sinhala and Tamil came into being on digital devices and the Internet.

Harsha Wijayawardhana B.Sc. (Miami), FBCS
COO/CTO Theekshana R & D, Director, Bank of Ceylon
Internet and Sinhala Unicode Pioneer in Sri Lanka

LK Domain Registry

ISBN 978-624-5374-00-7



9 786245 374007