2. CONQUERING THE DIGITAL DIVIDE AND UNITING PEOPLE THROUGH INDIAN LANGUAGE TECHNOLOGIES

TDIL

Major programme on the 'Technology Development for Indian Languages" (TDIL) was initiated by the Government since 1991 for capacity building in language computing in diverse situations in a number of languages for the first time.

Under the aegis of the TDIL programme, various centres have been established spreading across the country that have developed broad based capabilities, manpower, several niche technologies and enabling software and tools for Indian Language support. (TDIL) has been instrumental in generating wide interest in developing technology and resources relevant to the use of Indian languages in ICT (Information and Communication Technology). Through resource centres and funded projects, TDIL has created rich human resources, linguistic resources, software tools, etc. across the country in many languages.

With these seeding efforts of the government spread across the country and subsequent development of niche technologies and products, multilingual computing has gathered momentum in the country. Slowly private players and institution of DIT like C-DAC through have brought to the market products which are used widely addressing the growing and small market.

Meanwhile government initiatives noted above also helped to accelerate the participation of industries and also developed considerable interest among the multinational companies to tune their products with Indian Language interface as well as start research in the emerging language technology areas.

Genesis

India is a large multi-lingual country with 22 constitutionally recognized languages. Computer technology in India has both a developmental as well as a social role. In its developmental role, it is concerned with the designing and development of newer technologies for various applications. In its social role, it breaks the language barrier and bridges the gap between the various sections of the society through easier access to information using their respective mother tongues or local languages. Language

here has a major role to play and, therefore, language computing becomes central to the exchange of information across speakers of various languages.

Proliferation of Information Technology in the society requires availability of user-friendly tools and technologies in Indian Languages. These demands have necessitated launching of the programmes to accelerate development, awareness (through high visible public events) making available basic tools and technologies for mass usage at no cost to the users, support and encourage for productization and interaction with applications. Availability of such basic Indian language tools & technologies in diverse sectors is expected to act as catalyst to derive maximum advantage from the IT proliferation to the society.

National Roll Out Plan project (NROP)

Under the aegis of Ministry of Communications and Information Technology (MC&IT), TDIL programme under Department of Information Technology (DIT), has initiated National Roll Out Plan project in year 2005 for proliferation of languages through new media.

It is the intension of the Government to accelerate the reach of Indian language computing use through making available free software tools and fonts so that the pace of language computing use in Indian Languages accelerates enormously benefiting large segment of users not fully comfortable and conversant with English (90% of the population). This is expected to pay the dividends to kick off IT revolution in the country, increase PC and internet penetration, increase Indian Language contents and enable applications (such as e-governance, education etc.) to be available to wide range of users in daily life.

The program also aims at consolidating the Indian language tools & technologies developed so far by various Public (academia, R&D labs), Private players and brings out such development-tools for mass usage. The aim was to consolidate even the commercially successful tools & technologies from Private players with a bear minimum or no compensation & make it available to users for unlimited free usage.

This project focuses on development, awareness creation and mass proliferation of tools & technologies, support to end users & developers, support productization efforts through integration of Tools/components and support widespread deployment in practical applications in 22 Indian Languages in phased manner

Language CD

Since the CD is targeted towards common man, it contains tools for common man, productivity enhancement tools & beta tools for getting more feedback from user for future research. The general contents of the CD's released are as follows:

- Basic Information Processing Kit (BIPK) : text processing tools
- Productivity Enhancement Tools like Spell checker, code converter
- Beta tools for Power User for feedback

The CD has an Indian language Graphical User interface to facilitate use and instructions related to installation of the various software. The Graphics of CD and the CD cover have been designed keeping in view the cultural aspects associated with the language or the area where the language is most widely spoken or used.

Basic Information processing Kit (BIPK) for the masses.

It consists of Indian language tools and technologies which are required by the majority of users, including office automation, fonts and data entry and content creation tools. The BIPK consists of highly calligraphic True Type and UNICODE compliant Open Type fonts of various languages, keyboard drivers and layouts for inputting.

Akruti Multi-font keyboard engine basically will be able to support typing and creating documents using any of the fonts supported by the tool. Basically there are two types of fonts which are popular for desktop usage. These are True type and Open Type fonts. True type fonts are basically vendor dependent. This means that the code used for storing the characters is designed by individual organization. To support typing using the said font, we need to have the driver which is developed by that organization only. Tomorrow if someone wants to edit the document already created, he will also need to have the same driver else it will not be possible for him

to edit that document. As opposed to True Type, there are something called Open Type or Unicode enabled fonts. Unicode consortium specifies the codes for each character for all the languages worldwide using a unique code for each character. Therefore it becomes very simple for editing the documents created using Unicode fonts.

It also includes Localized Indian version of Open Office which includes word processor (writer). Spread sheet (calc), presentation tool (impress), drawing and math tool. For more information please see point 4.4

The Firefox browser, thunderbird email client and GAIM multiprotocol messenger have also been localized into Indian languages.

Keyboard and Language Learning tools have been included in the CD for proliferation of use of computers in Indian languages.

Productivity enhancing tools

Apart from the BIPK, the CD also consists of productivity enhancing tools such as Dictionaries, spellcheckers, legacy code converters, etc. The language learning, multimedia and keyboard learning tutor software are also included. We feel that the tools like Dictionaries are very helpful at various places, may it be office work or even for that matter building vocabulary of your child. Officials often create the documents in different languages and sometimes because of various reasons we need help of spell-checkers to see if the created document does not have any wrongly spelled words. Document converters have their own importance.

Today there are various companies involved in Indian language computing. Some of them are Akruti, Modular InfoTech, C-DAC, IMRC, C. K. Technologies, Softview, etc. These companies have their fonts in the market and used by lot of people especially in the areas of Data Entry, desktop publishing, media and others. As seen in the Multi-font keyboard engine explanation, people should have the drivers available with them to use the particular font in their documents. It may happen that the document which is created by one person needs to be edited/ changed by other. In this case there is a dependency that the driver which was used for creating the document is also needed for editing the document. Also if someone wants to use

another font from different vendor (may be because that font is better than the current font) then there should be a utility which can convert into the other font. Otherwise he will have to retype the entire document again. This is not that someone would really want because the amount of time and money and other resources required would be huge.

This tool converts various 8-bit True Type fonts encoding to and from ISCII / UNICODE. This is especially used when the data is created using font encoding.

Beta Tools

It takes lot of research and other efforts to develop or build tools such as optical character recognizer for Indian languages. Such tools should be used and evaluated by as many people as possible so as to make it more versatile. We generally call these products as Beta Products. Beta tools and research material such as Text to speech systems, Optical character recognition systems, morphological analyzers and generators etc. have also been included so that these are used by the experts, common people, Researchers, evaluate them and give the feedback to us so as to carry out further work and research to make these products better and useful.

Language wise Availability of Tools
The following chart shows the software tools and the language CDs in which these are available:

S. No.	CD Tools and technologies	Languages for which the Tools are included	Total language
1.	True Type Fonts with Keyboard driver	Hindi, Assamese, Gujarati, Kannada, Malayalam, Marathi, Oriya, Punjabi, Sanskrit, Tamil, Telugu, Urdu, Nepali, Bangla, Manipuri, Kashmiri, Sindhi, Konkani,	18
2.	Multi Font Keyboard Engine for True Type Fonts	Hindi, Assamese, Gujarati, Kannada, Malayalam, Marathi, Oriya, Punjabi, Telugu, Bangla.	10
3.	Unicode Compliant Open Type Font	Hindi, Assamese, Bodo, Dogri, Gujarati, Kannada, Maithili, Malayalam, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Tamil, Telugu, Urdu, Bangla, Manipuri, Kashmiri, Sindhi, Santali, Konkani.	22
4.	Unicode Compliant Keyboard driver	Hindi, Assamese, Bodo, Dogri, Gujarati, Kannada, Maithili, Malayalam, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Telugu, Tamil, Urdu, Bangla, Manipuri, Kashmiri, Sindhi, Santali, Konkani.	22

5.	Generic Font's code and storage code converter	Hindi, Assamese, Gujarati, Kannada, Malayalam, Marathi, Oriya, Punjabi, Telugu.	9
6.	Bharateeya open office (Open Source)	Hindi, Assamese, Bodo, Dogri, Gujarati, Kannada, Maithili, Malayalam, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Tamil, Telugu, Urdu, Bangla, Manipuri, Kashmiri, Sindhi, Santali, Konkani.	22
7.	Firefox- browser	Hindi, Bodo, Dogri, Maithili, Nepali, Punjabi, Assamese, Kannada, Malayalam, Marathi, Tamil, Telugu. Bangla, Manipuri, Kashmiri, Sanskrit, Sindhi, Santali, Konkani, Gujarati.	20
8.	GAIM- Multi protocol Messenger	Hindi, Telugu, Kannada, Punjabi, Oriya, Marathi, Assamese, Malayalam, Urdu, Tamil	10
9.	Columba – Email Client & Lime wire	Hindi, Telugu.	2
10.	OCR	Hindi, Malayalam, Marathi, Punjabi, Tamil.	5
11.	Typing tutor	Hindi, Tamil, Marathi, Oriya, Telugu, Kannada, Malayalam	7
12.	Integrated word processor	Hindi, Urdu, Telugu,	3
13.	Dictionary English – Hindi	Hindi, Assamese, Dogri, Gujarati, Kannada, Maithili, Malayalam, Marathi, Nepali, Oriya, Punjabi, Tamil.	12
14.	Spell checker	Hindi, Assamese, Gujarati, Kannada, Malayalam, Marathi, Oriya, Tamil, Telugu	9
15.	Transliteration tool	Hindi, Assamese, Gujarati, Kannada, Malayalam, Marathi, Oriya, Urdu.	8
16.	Text to speech system	Hindi, Kannada, Malayalam, Oriya	4
17.	Microsoft word tool	Malayalam, Assamese, Marathi, Gujarati, Kannada, Oriya,	6
18.	Microsoft Excel tool	Assamese, Gujarati, Kannada, Malayalam, Marathi, Oriya	6
19.	Database sorting tool	Assamese, Gujarati, Kannada, Malayalam, Marathi, Oriya,	6
20.	E-mail Client (Thunderbird)	Bodo, Dogri, Kannada, Maithili, Nepali, Tamil, Malayalam, Assamese, Marathi, Oriya, Punjabi, Gujarati, Sanskrit, Urdu, Bangla, Manipuri, Kashmiri, Sindhi, Santali, Konkani	20
21.	Multiprotocol Messenger (PIDGIN)	Bodo, Dogri, Maithili, Nepali, Bangla, Manipuri, Kashmiri, Sindhi, Santali, Konkani, Gujarati, Sanskrit ,	12
22.	Sunbird calendar application	Bodo, Dogri, Gujarati, Maithili, Nepali, Manipuri, Kashmiri, Sindhi, Santali, Konkani	10
23.	Content Management system	Bodo, Dogri, Gujarati, Kannada, Maithili, Nepali, Oriya, Punjabi, Sanskrit, Bangla, Manipuri, Kashmiri, Sindhi, Santali, Konkani.	15
24.	Decorative font design tool	Gujarati, Kannada, Malayalam, Marathi, Oriya, Punjabi, Telugu.	7
25.	Games/Puzzles	Gujarati, Kannada.	2
26.	Type Assistant	Gujarati, Kannada, Malayalam, Marathi, Oriya, Assamese.	6

27.	Tux Paint	Gujarati.	1
28.	Scribus for Linux	Gujarati.	1
29.	Text Editor	Kannada, Malayalam.	2
30.	Library Management System	Kannada	1
31.	Tutor package	Kannada, Malayalam, Oriya, Sanskrit,	4
32.	Personal utilities	Kannada	1
33.	Kannada logo	Kannada	1
34.	Word net	Marathi	1
35.	Morphological analyzer and generator	Punjabi	1
36.	Scribus layout and publishing software	Sanskrit, Bodo, Dogri, Maithili, Nepali, Manipuri, Gujarati, Kashmiri, Sindhi, Santali, Konkani.	11
37.	Sanskritpradipika	Sanskrit.	1
38.	Online multilingual Amarkosha	Sanskrit.	1
39.	Mahabharata indexer	Sanskrit	1
40.	Sandhi Generator	Sanskrit.	1
41.	Subanta analyzer	Sanskrit.	1
42.	Subanta Generator	Sanskrit.	1
43.	Tinanta Analyzer	Sanskrit.	1
44.	Tinanta Generator	Sanskrit.	1
45.	Karaka Analyzer	Sanskrit.	1
46.	POS Tagger	Sanskrit.	1
47.	Letter pronunciation	Sanskrit.	1
48.	Alphabet writing	Sanskrit.	1
49.	Animated Panchatantra stories	Sanskrit.	1
50.	Urdu encyclopedia	Urdu	1
51.	Urdu language resources	Urdu	1
52.	Urdu language design guide and PASCII resources	Urdu	1
53.	Word Application	Telugu	1
54.	Word processor with spell checker	Telugu	1
55.	Telugu Language Publications	Telugu	1

Brief Description of Important software Tools

1. **True Type fonts with Keyboard Driver**
This includes bit True Type fonts from different organizations each having its own encoding. The respective keyboard driver also accompanies this. This is especially useful for the systems, which do not support Unicode.

2. **Multifont keyboard engine for True Type fonts**

This Keyboard Engine provides support for typing in a language using True Type Fonts with different encoding.

3. Unicode compliant Open Type Fonts
This has Unicode compliant Open Type Fonts & is intended to be used along with UNICODE enabled Operating systems such as Windows 2000, XP, etc.

4. Unicode compliant keyboard driver.
This Keyboard Engine provides support for typing in a language using Open Type Fonts

5. Generic fonts and storage code converters
This tool converts various 8-bit True Type fonts encoding to and from ISCII / UNICODE

6. Localized version of Bharateeya Open Office (Spread Sheet, Presentations, Word processing & drawing tools), Fire fox Browser
This includes localized version of Bharateeya Open Office Suite (Impress, Writer, Calc, Base, and Draw), Email client and Browser.

7. **Spell checker**
This software is used to spell check a document for spelling mistakes. This tool also gives suggestions in case there are any spelling mistakes found in Malayalam text.

8. Bilingual Dictionary
This tool includes a Bilingual Dictionary. The user can type English word and find Malayalam equivalent of the word and Vice versa.

9. **Decorative font designing Tool (Likhan)**
It is possible and easier to produce decorative and colorful titles in Indian Languages including English. This is a unique "Vector Based Tool" to compose decorative and colorful titles in vernacular and English Languages.

10. Database Sorting Tool
This software sorts the database tables in Malayalam language in alphabetic order

11. Type Assistant
This utility provides a simple way to type in English and get the text in Malayalam language. This tool is very much user friendly and can be mastered easily.

12. Microsoft Word Tool
This includes tools such as Number To Word, Smart Converter, Sorting, Spell Checker, etc

13. Microsoft Excel Tools
This includes tools such as Transliteration from English text to vernacular text, Sorting, and converters.

14. **Typing Tutor**
Typing Tutor will enable any individuals or institutes to undergo language Keyboard training at their own pace.

15. Localized version of Content Management System
ECKO and Drupal is a generic, extensible, process-driven software based framework for establishing Community Information Systems (CIS) or Social Information Systems (SIS) in the form of E-Communities.

16. OCR
Optical Character Recognition System, converts scanned language document images into computer editable text.

17. Word Processor with Spell Checker
It is basically a word Processing System that includes utilities such as spell checkers inside them.

18. **Games and Puzzles**
This software includes small games and puzzles based on language characters, numbers, and general knowledge, etc.

19. Language learning tool
It is a Computer based tutorial to enable self-paced learning for first time computer users. Voice over in a language, guided practice and highly interactive interface allows the learner to learn and practice through a scientifically designed spiral-learning mode

20. **Morphological Analyzer and Generator**
This tool automates the morphological analysis and generation activities for a language. Through this user can know word forms of word along with their grammatical information, know grammatical information including root of a word, and generate a word from a root word based on specified grammatical information.

21. Transliteration Tool
This tool transliterates English database tables to vernacular equivalents and also converts database from any supported fonts to any other fonts.

22. Localized version of Tux Paint
Tux Paint is a free drawing program designed for young children (kids ages 3 and up). It has a simple, easy-to-use interface, fun sound effects, and an encouraging cartoon mascot who helps guide children as they use the program

23. Localized version of Scribus
Scribus is an open-source program that brings excellent and professional page layout to Linux/Unix, MacOS X, OS/2 and Windows desktops

24. Localized version of Sunbird Calendar

Sunbird is a calendar application. It's entirely standalone: it doesn't require the bulk of another application, but doesn't offer any integration into other programs

25. Text To Speech System
This software reads out language text input intelligibly. This software is integrated with a text editor or as stand-alone.

26. Localized version of Logo
LOGO is a programming language for kids. We have a localized version for this software.

27. Seamless E-mail Client
This is a utility for sending the emails and documents in a language in a very easy fashion.

28. Localized version of Personal Utilities
This includes such utilities, which can be used for personal purposes such as Address management; Maintain a list of personal and official engagements, calendar applications, etc.

29. Word Nets
The Word Net Browser is a Graphical User Interface to access the Word Net lexical database. The users input words in Unicode and the results are also shown Unicode

30. Library Management System
Basically it is a library management system where Books are categorized on different subjects. The search optionis provided based on author, book title, category and etc.,
Various reports can be generated here.

Chronological Release Status

As on date free Language CDs containing software tools and fonts have been released for all the 22 scheduled Indian languages along with their alternate scripts, if any. The chronological list of release is given below:

Releases in the Year 2005:
- Tamil - 15th April 2005

- Hindi - 20th June 2005

- Telugu - 29th October 2005

Releases in the Year 2007:
- Marathi – 24th January 2007

- Malayalam – 24th January 2007

- Oriya – 24th January 2007

- Kannada – 24th January 2007

- Punjabi – 24th January 2007

- Urdu– 24th January 2007

- Assamese – 24th January 2007

Releases in the Year 2008:
- Gujarati – 1st May 2008

- Sanskrit – 25th November 2008

Releases in the Year 2009:
- Bodo – 21st February, 2009

- Dogri – 21st February, 2009

- Maithili – 21st February, 2009

- Nepali – 21st February, 2009

- Bengali – 08th September, 2009

- Manipuri - 08th September, 2009

- Kashmiri - 08th September, 2009

- Konkani - 08th September, 2009

- Santali - 08th September, 2009

- Sindhi - 08th September, 2009

Localization of Bharateeya Open office in all 22 Scheduled Indian languages

Localization is the process of adapting a product or service to a particular language, culture, and gives the desired local "look-and-feel." A successfully localized service or product is one that appears to have been developed within the local culture.

It is important to understand the importance of languages and availability of all data including the IT related data in all languages existing in the world today so that progress reaches equally to all sections of society.

Language is a way of communication between two people or between a group. Ethnologue lists 6,912 living languages in the world today though it may never be determined exactly as there may be many more civilizations which are yet to be listed.As a large and linguistically diverse country, India is listed to have around 427 languages languages out of which 22 languages are officially recognized. IT field is developing at a massive speed but all the data is available mostly in English

To show the necessity of availability of localized versions of computer material, we are listing below our 22 official languages, with the places they are spoken along with the number of speakers (wherever data was available) :

Indian Language Chart

Sr. No	Language / No of Speaker	Script	Family	Official language	Where Spoken
1	Assamese (15 million)	Bangla (Modified)	Indo-Aryan	Assam	Assam
2	Bengali (67 million)	Bangla	Indo-Aryan	Tripura and West Bengal	Andaman & Nicobar Islands, Tripura, West Bengal,
3	Bodo	Devanagari Bangla (Modified)	Tibeto-Burman	Assam	Assam
4	Dogri	Devanagari, Perso-Arabic	Indo-Aryan	Jammu and Kashmir	Jammu and Kashmir
5	Gujarati (43 million)	Gujarati	Indo-Aryan	Dadra and Nagar Haveli, Daman and Diu, and Gujarat	Dadra and Nagar Haveli, Daman and Diu, Gujarat
6	Hindi (180 million)	Devanagari	Indo-Aryan	Andaman and Nicobar Islands, Bihar, Chandigarh, Chhattisgarh, Delhi, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Rajasthan, Uttar Pradesh and Uttaranchal	Andaman and Nicobar Islands, Maharashtra Arunachal Pradesh, Bihar, Chandigarh, Chhattisgarh, the national capital territory of Delhi, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Rajasthan, Uttar Pradesh and Uttarakhand.
7	Kannada (35 million)	Kannada	Dravidian	Karnataka	Karnataka
8	Kashmiri (56,693)	Perso-Arabic, Devanagari	Indo-Aryan	Kashmir	Jammu and Kashmir
9	Konkani (1,760,607)	Devanagari, Roman, Malayalam, Kannada	Indo-Aryan	Goa	Goa, Karnataka, Maharashtra
10	Maithili (22 million)	Devanagari	Indo-Aryan	Bihar	Bihar
11	Malayalam (34 million)	Malayalam	Dravidian	Kerala and Lakshadweep	Kerala, Andaman and Nicobar Islands, Lakshadweep
12	Manipuri (1,270,216)	Bangla, Meithei-Mayek	Tibeto-Burman	Maithili	Manipur
13	Marathi (65 million)	Devanagari	Indo-Aryan	Maharashtra	Dadra & Nagar Haveli, Daman and Diu, Goa, Maharashtra
14	Nepali (2,076,645)	Devanagari	Indo-Aryan	Sikkim	Sikkim, West Bengal
15	Oriya (30 million)	Oriya	Indo-Aryan	Orissa	Orissa
16	Punjabi (26 million)	Gurumukhi, Shahmukhi	Indo-Aryan	Punjab	Chandigarh, Delhi, Haryana, Punjab
17	Sanskrit (49,736)	Devanagari	Indo-Aryan		
18	Santali	Devanagari OI (Ciki)	Munda		Santhal tribals of the Chota Nagpur Plateau (comprising the states of Bihar, Chattisgarh, Jharkhand, Orissa)
19	Sindhi (2,122,848)	Perso-Arabic, Devanagari, Gujarati, Roman	Indo-Aryan		
20	Tamil (66 million)	Tamil	Dravidian	Tamil Nadu and Pondicherry	Tamil Nadu, Andaman & Nicobar Islands, Kerala, Pondicherry
21	Telugu	Telugu	Dravidian	Andhra Pradesh	Andaman & Nicobar Islands, Andhra

22	(70 million)				Pradesh
22	Urdu (46 million)	Perso-Arabic	Indo-Aryan	Jammu and Kashmir	Andhra Pradesh, Delhi, Jammu and Kashmir, Uttar Pradesh

In the above list, there may be a negligible percentage of people conversant in their local language as well as English. Hence, if the modern IT technology, wherein all data is available on a finger-tip, needs to reach the masses, the process of localizing all the data available is very important.

Realizing the need for localization, TDIL,DIT targeted release of localized versions of Bharateeya OO, Fire fox, Thunderbird and Pidgin software in all the above languages. This was a big challenge.

Challenges faced during Localization

Looking at the list above, it is evident that while there was lot of speakers for some popular languages, some languages have very less number of speakers which could not even be listed. Even amongst those, people who were computer savvy, conversant in English and also the local language are very less. While are lot of literary experts in various languages but in the computer field, very few are available. Moreover, localization was meant for the common men, who are not supposed to be fluent in English and hence it was actually required to create a whole new set of words especially for computer usage which have to very user-friendly.

The material for localizing purposes was a set of 60000 basic computer strings in English and 2 lakh strings for advanced users, which needed to be localized (not translated) in all languages. Moreover, these strings are not complete sentences, wherein one can find tenses, verbs or clauses. They are computer commands, which are used in English. Even the basic words like file, folder, directory, document etc. are not enlisted in any dictionary available. While in some languages, it has been transliterated and retained as they are in English, linguists/ experts of certain languages set about the task of creating words in their own language for every word in IT terminology. Then, there are words like delimiters, add-ons, plug-ins etc. which had become words of common usage in English, but for other languages, finding equivalent words is a tough job. This was a mammoth job, involving lots of discussions and mutual agreement. These localized words were just the consolidated

opinion of some experts, but if these words were to reach the common man, dictionaries were necessary. Simultaneously, a glossary was also developed for all languages for this purpose.

Within every language, dialects were different in the cities of the same state. It has to be taken into consideration, while making the strings. For instance, Hindi spoken in different parts of our northern belt varies a lot from city to city. Hence, this was a continuous experimentation process to find words commonly acceptable to all.

Pune, being the hub of education, where lot of students from all over India come to pursue their higher studies and do PhDs hence some local experts for Sindhi, Urdu, Assamese, Oriya , Marathi, Bengali etc. and through their help, experts already working in this field were enlisted from local areas. Various universities, wherein linguistics is taught as a subject were approached, and advertisement inviting applications was released in various national and local newspapers. This led to creation of an expert team of 70 free-lancers working in various languages from different parts of India.

The real challenge was to find experts in languages like Dogri, Maithili, Santhali, Bodo and Sanskrit. Lots of literary work has been done in these languages but no terminology or even basic dictionaries were readily available in these languages.

Localized Open Source Software

Under this initiative CDAC under directions of TDIL, DIT has undertaken the task of localizing the following Free and Open Source Software:

- BOO which is an entire office suite and includes
    - Writer (word processor),
    - Calc (Spreadsheet),
    - Impress (presentation tool),
    - Draw (drawing tool)
- Thunderbird (E-mail client)

- Firefox (Browser)
- GAIM (Multiprotocol messenger)

These are available free and are data compliant with MS-Office products. Supports Linux as well as windows and the data generated is cross platform.

BHARATEEYA OPEN OFFICE SUITE

Open Office package is a very handy and useful set of application software that enables us to complete our day-to-day tasks quickly and efficiently. OO is an Office automation package, which contains a set of applications capable of performing a number of tasks like word-processing, spread sheets, etc. It is platform independent i.e. OO can work on Windows and Linux. Also being open source software, it is free software, unlike Proprietary Office suite

The BharateeyaOO suite mainly consists of a word processor application (Writer), a spreadsheet application (Calc), a presentation application (Impress), a drawing application (Draw). These applications have interfaces similar to those of the Common and widely used Office applications and those who are familiar with the other office packages such as, MS Office, Star Office and Lotus Smart Suite would find it easy to work with the BharateeyaOO package. The data generated is cross-compatible i.e. Existing Office data can be opened and changed using OO and vice versa. BharateeyaOO is UNICODE compliant and requires Indian language enabled OS. Indian Language input is possible using Open Type fonts (for languages supported in the OS) as well as True type fonts for others.

The BharateeyaOO suite, is the customized version of Open Office, with all the menus, status bars, error messages, user prompts, etc. localized for Indian languages. There is a Help menu in each of the BharateeyaOO applications, which contains help options in Indian Languages that you can use to get help about the various tools and commands of the application.

Training on Indian Language Tools and Technologies

In order to further proliferate the use of the Indian language tools and technologies, training of various government departments such as Election Commission of India, Department of Information Technology, etc, Central Institute of Indian Languages (CIIL) , has been undertaken. Several PACE centers also offer low cost training to the masses on the tools and technologies offered under the initiative. These courses are targeted mostly at Indian language Office automation and content creation tools such as BharateeyaOO.o.

CD Distribution and Impact

The Department of Post has made special provisions and resources for delivering the free CDs. Department of Post has done commendable work by ensuring that the impact of "free software tools and fonts" initiative reaches all corners of India.

Till date approximately 7 lakh free software tools and Fonts CDs have been distributed to the masses. Apart from shipment of CDs, approximately 42 lakh downloads have happened from the website. Other media of distribution include magazines, pre-bundled software with OEM, etc.

Apart from individuals, several organizations such as Election Commision Of Indian, State Bank of Hyderabad, etc and e-governance initiatives and state and central government departments are migrating from using proprietary expensive software to using the free software tools available in CD, thereby saving the country's exchequer lot of money.

Support

The need for continued support to the users is well understood and and it is being provided through various means like emails, telephones, faxes, letters, and even by providing classroom training wherever it is needed. In near future an automated helpdesk would be setup so the round the clock support could be provided.

Appreciation

Several Individuals and Organisation have expressed their satisfaction and appreciation of the free software tools and fonts CD initiative. Some appreciation and request letters are as below:

Outcome

The project has helped in creation of database of all the Indian language developers, solutions and application providers which has become a ready reckoner for projects to be implemented at National level. This initiative also provided inputs about user needs thus gave direction for research to be undertaken.

Goals of Phase-II

This was the first effort of its kind in the country and hence was directed towards consolidation of whatever language processing tools were available in the country and awareness creation to the public about availability of such language processing technology. To support the legacy systems True Type Fonts (TTF) were provided besides Open Type Fonts .The goal of phase-II is to provide more and more free open source tools, platform independent open type fonts.

Some Useful Reference Links and acknowledgements

1. Technology Development of Indian Languages- http://tdil.mit.gov.in/
2. Center for Development of Advanced Computing- www.cdac.in
3. Indian Languages Data Center- www.ildc.in