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. India is a multilingual country with as many as 22 scheduled languages and only about 5% of the population is able to understand English.

Keeping this in mind, Ministry of Communications and Information Technology launched a major initiative in the year 2005 the area of e-governance, to enable more reliable and efficient services to our citizen's requirements from the government in tasks ranging from land / citizen information to obtaining passport and managing tax payments. Through this Language Technology mission, a major initiative has been taken to aggregate Indian Language fonts and software tools from various public/private players, incorporate them into user friendly tools and products and make them available free for public through CDs and web downloads. IT ministry's initiative of Technology Development for Indian Languages (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 a number of resource centres and funded projects, it has created rich human resources, linguistic resources, software tools, etc. across the country in many languages.

**Current releases**

C-DAC, GIST, Pune under the leadership of TDIL, DIT has released the CD’s for ten constitutionally recognized Indian languages viz. Tamil, Hindi, Telugu, Marathi, Urdu, Punjabi, Oriya, Kannada, Assamese and Malayalam for free mass usage. Gujarati and Bengali are in pipeline. The process of consolidation for other languages as well as release of second version of CD for some of the languages is currently underway.

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.

Broadly the contents of each language CD are as follows:

1. True Type fonts with keyboard driver
2. Multi-font keyboard engine for True Type fonts
3. Unicode compliant OpenType fonts
4. Unicode compliant keyboard driver.
5. Generic font’s code and storage code converter
6. Localized version of BharateeyaOO (Spread Sheet, Presentations, Word processing & drawing tools), Fire fox Browser, Thunderbird email client, GAIM (Multi protocol messenger)
7. Spellchecker
8. Bilingual Dictionary
9. Decorative fonts design tool
10. Transliteration Tool
11. Language Tutor
12. Text to Speech
13. Database sorting tool
14. Microsoft word tools
15. Microsoft Excel tools
16. Type Assistant
17. Content management system
18. Typing tutor
19. Games/Puzzles
20. Library Management System
21. Seamless email send / receive utility
22. Text Editor
23. OCR
24. WorldNet
25. Morphological Analyzer and Generator
26. Text To Speech Systems

**2.0 Process / life cycle**

The tools built by resource center along with contributions from other academica and private Indian language developers have been consolidated into the “free software tools and fonts”. These are available in the form of a free CD as well as free download on the internet. On
registering online free home delivery of CD is done anywhere in India.

The contributors were identified by releasing press advertisements from time to time. A performa for submitting contributions was made available. On submission these possible contributors were invited for technical deliberations.

A technical team evaluated the proposals as well as the software and negotiated finances with the possible contributors, for finalizing the list of software. Technical constraints such as size of software, usability, novelty / availability of other similar software, impact of the technology, etc. were also considered.

The software submitted by the possible contributors was vetted for basic functionality and use. The resulting test reports were submitted back for possible corrections and updates. In several cases the contributors worked very hard and updated softwares in extremely short periods of time to cater to the needs of quality and usability.

3.0 Localization an Overview

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. Before we discuss the above process in detail, we must talk about the languages and the importance of 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):

<table>
<thead>
<tr>
<th></th>
<th>Language Name</th>
<th>Population</th>
<th>Other Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Assamese</td>
<td>15 million</td>
<td>Assam</td>
</tr>
<tr>
<td>2.</td>
<td>Bengali</td>
<td>67 million</td>
<td>Andaman &amp; Nicobar Islands, Tripura, West Bengal</td>
</tr>
<tr>
<td>4.</td>
<td>Dogri</td>
<td></td>
<td>Jammu and Kashmir</td>
</tr>
<tr>
<td>5.</td>
<td>Gujarati</td>
<td>43 million</td>
<td>Dadra and Nagar Haveli, Daman and Diu, Gujarat</td>
</tr>
<tr>
<td>6.</td>
<td>Hindi</td>
<td>180 million</td>
<td>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.</td>
</tr>
<tr>
<td>7.</td>
<td>Kannada</td>
<td>35 million</td>
<td>Karnataka</td>
</tr>
<tr>
<td>9.</td>
<td>Konkani</td>
<td>1,760,607</td>
<td>Goa, Karnataka, Maharashtra</td>
</tr>
<tr>
<td>10.</td>
<td>Malayalam</td>
<td>22 million</td>
<td>Bihar</td>
</tr>
<tr>
<td>11.</td>
<td>Malayalam</td>
<td>34 million</td>
<td>Kerala, Andaman and Nicobar Islands, Lakshadweep</td>
</tr>
<tr>
<td>12.</td>
<td>Manipuri</td>
<td>1,270,216</td>
<td>Manipur</td>
</tr>
<tr>
<td>13.</td>
<td>Marathi</td>
<td>65 million</td>
<td>Dadra &amp; Nagar Haveli, Daman and Diu, Goa, Maharashtra</td>
</tr>
<tr>
<td>14.</td>
<td>Nepali</td>
<td>2,176,645</td>
<td>Sikkim, West Bengal</td>
</tr>
<tr>
<td>15.</td>
<td>Oriya</td>
<td>20 million</td>
<td>Orissa</td>
</tr>
<tr>
<td>16.</td>
<td>Punjabi</td>
<td>26 million</td>
<td>Chandigarh, Delhi, Haryana, Punjab</td>
</tr>
<tr>
<td>17.</td>
<td>Sanskrit</td>
<td>49,736</td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>Santhali</td>
<td>Santal tribal of the Chota Nagpur Plateau (comprising the states of Bihar, Chhattisgarh, Jharkhand, Orissa)</td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>Sandhi</td>
<td>2,122,848</td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>Tamil</td>
<td>66 million</td>
<td>Tamil Nadu, Andaman &amp; Nicobar Islands, Kerala, Puducherry</td>
</tr>
<tr>
<td>21.</td>
<td>Telugu</td>
<td>70 million</td>
<td>Andaman &amp; Nicobar Islands, Andhra Pradesh</td>
</tr>
<tr>
<td>22.</td>
<td>Urdu</td>
<td>46 million</td>
<td>Andhra Pradesh, Delhi, Jammu and Kashmir, Uttar Pradesh</td>
</tr>
</tbody>
</table>
In the above list, there must 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, CDAC, GIST was given the assignment of releasing localized versions of Bharateeya OO, Fire fox, Thunderbird and Pidgin softwares in all the above languages. This was taken as a challenge, and the process of looking for experts / linguists for these languages began.

**Challenges faced during Localization**

Looking at the list above, you can see 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 were very less. While there were lot of literary experts in various languages but in the computer field, very few were available. Moreover, localization was meant for the common men, who were not supposed to be fluent in English and hence we were actually required to create a whole new set of words especially for computer usage which have to very user-friendly. Hence, keeping this in mind, we had to search for competent people.

The material provided to us for localizing purposes were 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 were not complete sentences, wherein you can find tenses, verbs or clauses. They were computer commands, which we use in English. Even the basic words like file, folder, directory, document etc. were not enlisted in any dictionary available. While in some languages, it was transliterated and retained as they were 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 were words like delimiters, add-ons, plugins etc. which had become words of common usage in English, but for other languages, finding equivalent words was 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. We also had to take this 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. We were fortunate to find 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 was taught as a subject were approached, and advertisement inviting applications was released in various national and local newspapers. We presently have 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. We could find some experts, but the speed of work was slow as it was a completely new area of work for them. But now basic translations are over in all the above languages except Sanskrit and validation is going on now. Sanskrit localization has also been taken up by JNU at Delhi.

**4.0 List of Contents / Contributions**

The free software tools and fonts CD consists of various tools and technologies classified as under

- 4.1 Basic Information Processing Kit
- 4.2 Productivity enhancing tools
- 4.3 Beta tools for power user
**Bharateeya Open Office Calc**

Calc is a spreadsheet program. A spreadsheet program enables us to store and manage our numeric data, and perform calculations in it. Calc has a wide array of tools and features that enable us to perform complex calculations with just a few mouse-clicks. It can also open and save other spreadsheet application files such as MS-Excel.

**Bharateeya Office Impress**

It is a presentation application that provides us with many tools for creating presentations. Presentation software helps us to create dynamic slides and present them to the audiences through a slide show. Useful presentation software that is included in the BharateeyaOO package is impress. It is a very efficient and potent program that offers a number of features that enable us to create and modify effective and appealing presentations and run them in a slide show.

The tools include, guides for positioning objects, automatic snapping of the object to a freely definable grid or to each other, and scaling and cross-fading and many other effects.

There are also other features in BharateeyaOO - Impress, which can help us create our presentations easily and quickly. We can also export our presentation to be published on the Internet. All the necessary conversions are done automatically. The presentation you export can be viewed with any modern browser.
Bharateeya Open Office Draw

Draw is an object-oriented vector graphic drawing program. The object can be lines, rectangle, 3D cylinders or other polygons can be drawn. All objects already have set properties, such as size, color of the surface, color of contours, linked files, associated actions when clicked and much more. All of the properties can be modified at any time. The Bharateeya Draw has all UI in Indian languages. Image below shows Draw having all the menus localized in Urdu.

Thunderbird

Thunderbird is a free, open-source and cross platform mail client for most OS including, Windows, Linux and Macintosh. It is a complete email application, which is simple to use, powerful and customizable. It easily exports your existing email accounts and messages. It is similar to competing products like Outlook Express, but with additional features such as junk mail classification.

Firefox browser

The popular Firefox Browser has been provided with all the menus, status bars, error messages, user prompts, etc. in Indian Languages. The image below shows the Firefox browser, having all menus and various options, all localized in Urdu language.

5.0 Impact of these Software tools and technologies

Till 22/10/2007 approximately 293238, free software tools and Fonts CDs have been distributed to the masses. Apart from Online registrations and shipment of CDs, approximately
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.

4.1 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.

4.2 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.
4.3 Beta Tools (to update)
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.

4.4 Localized Open Source Software

Under this initiative C-DAC along with other open source partners 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)

The languages for which these are already available include: Assamese, Hindi, Kannada, Malayalam, Marathi, Oriya, Punjabi, Tamil, Telugu and Urdu. 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.

**Bharateeya Open Office Writer**

Writer is a free, open-source and cross platform (supports) word processor application, with all the menus localized. This is a word processor application, with all strings localized in Indian languages, which lets us create text documents, open existing MS-Word documents, Print, create Content, create WebPages, etc.

We can create personal letters, form letters, brochures, faxes and even professional manuals using this application. It provides a number of editing tools, which help us in editing and proofing our documents; and a number of formatting tools that help us in managing the appearance of our documents. It supports Windows, Linux as well as UNICODE. It can import as well as export files in .doc format as well as .html format. So Writer can be used as web-content creation tool.
Several contributors have partnered with C-DAC in joint development, productisation and testing of the available tools and technologies. This includes Utkal University, University Hyderabad, Kannada Ganak Parishad / Govt. of Karnataka, etc.

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Name of Contributor</th>
<th>Contact Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>C-DAC</td>
<td><a href="http://www.cdac.in/">http://www.cdac.in/</a></td>
</tr>
<tr>
<td>02</td>
<td>CyberScope Multimedia</td>
<td><a href="http://www.cyberscopemedia.com/">http://www.cyberscopemedia.com/</a></td>
</tr>
<tr>
<td>03</td>
<td>Chrome Kompal</td>
<td><a href="http://www.chromekompalinfo.com/">http://www.chromekompalinfo.com/</a></td>
</tr>
<tr>
<td>05</td>
<td>IIT, Guwahati</td>
<td><a href="http://www.iitg.ernet.in/">http://www.iitg.ernet.in/</a></td>
</tr>
<tr>
<td>06</td>
<td>Kannada Ganak Parishad</td>
<td><a href="http://www.kgpg.co.in/">http://www.kgpg.co.in/</a></td>
</tr>
<tr>
<td>07</td>
<td>IISc, Bangalore</td>
<td><a href="http://www.iisc.ernet.in/">http://www.iisc.ernet.in/</a></td>
</tr>
<tr>
<td>08</td>
<td>SoftVision Technologies</td>
<td><a href="http://www.softvision.in/">http://www.softvision.in/</a></td>
</tr>
<tr>
<td>09</td>
<td>Vasant Kannada Software</td>
<td><a href="http://www.vasantkannada.com/">http://www.vasantkannada.com/</a></td>
</tr>
<tr>
<td>10</td>
<td>Thomson Cyber Solutions</td>
<td><a href="http://www.thomsoncybersolutions.com/">http://www.thomsoncybersolutions.com/</a></td>
</tr>
<tr>
<td>11</td>
<td>Prag Informatix</td>
<td><a href="http://www.praginformatix.com/">http://www.praginformatix.com/</a></td>
</tr>
<tr>
<td>12</td>
<td>IIT, Mumbai</td>
<td><a href="http://www.iitm.ac.in/">http://www.iitm.ac.in/</a></td>
</tr>
<tr>
<td>13</td>
<td>Utkal University</td>
<td><a href="http://www.utkaluniversity.ac.in/">http://www.utkaluniversity.ac.in/</a></td>
</tr>
<tr>
<td>14</td>
<td>Punjab University, Patiala</td>
<td><a href="http://www.punjabuniversity.ac.in/">http://www.punjabuniversity.ac.in/</a></td>
</tr>
<tr>
<td>15</td>
<td>Thapar Institute</td>
<td><a href="http://www.thapar.edu.in/">http://www.thapar.edu.in/</a></td>
</tr>
<tr>
<td>16</td>
<td>Modular Infotech</td>
<td><a href="http://www.modular-infotech.com/">http://www.modular-infotech.com/</a></td>
</tr>
<tr>
<td>17</td>
<td>NIT, Jalandhar</td>
<td><a href="http://www.nitj.ac.in/">http://www.nitj.ac.in/</a></td>
</tr>
<tr>
<td>18</td>
<td>ABYEC Research Centre</td>
<td><a href="http://www.abyec.org/">http://www.abyec.org/</a></td>
</tr>
<tr>
<td>19</td>
<td>LearnJava Systems</td>
<td><a href="http://www.learnjava.com/">http://www.learnjava.com/</a></td>
</tr>
<tr>
<td>20</td>
<td>Palmuppa Brothers</td>
<td><a href="http://www.palmuppa.com/">http://www.palmuppa.com/</a></td>
</tr>
<tr>
<td>21</td>
<td>IIT, Hyderabad</td>
<td><a href="http://www.iitb.ac.in/">http://www.iitb.ac.in/</a></td>
</tr>
<tr>
<td>22</td>
<td>Prlogist</td>
<td><a href="http://www.prlogist.com/">http://www.prlogist.com/</a></td>
</tr>
<tr>
<td>23</td>
<td>Anusmrt</td>
<td><a href="http://www.anusmrt.com/">http://www.anusmrt.com/</a></td>
</tr>
<tr>
<td>24</td>
<td>CALTS</td>
<td><a href="http://www.caltoy.in/">http://www.caltoy.in/</a></td>
</tr>
<tr>
<td>25</td>
<td>Waseen</td>
<td><a href="http://www.waseen.org/">http://www.waseen.org/</a></td>
</tr>
<tr>
<td>26</td>
<td>Free Software Foundation of India</td>
<td><a href="http://www.fsfi.org.in/">http://www.fsfi.org.in/</a></td>
</tr>
<tr>
<td>27</td>
<td>Ultronix Digital Technologies</td>
<td><a href="http://www.ultronixdigitals.com/">http://www.ultronixdigitals.com/</a></td>
</tr>
</tbody>
</table>

The contributors have also been acknowledged in the respective CD and CD Covers.

7.0 Support and Feedback

We do not stop here just by developing the softwares, tools and giving to the common masses. We also consider that the help and support should be available to those people who are ready to learn and use these softwares and bring a joy amongst them by using these softwares in their own languages. For the same, we have a very strong team of dedicated people across different places of the country to provide help through various means like emails, telephones, faxes, letters, and even by providing classroom training wherever it is needed. People should not hesitate to contact us for any problem, however small it is.
Several Individuals and Organisation have expressed their satisfaction and appreciation of the free software tools and fonts CD initiative.

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/www.ildc.gov.in

Courtesy:

Ms. Swaran Lata
Director
TDIL Programme
E-mail: slata@mit.gov.in
Tel.: 011-24363525

Shri Mahesh D. Kulkarni
Chief Investigator
Email: mdk@cdac.in
Tel:020-25694000