3. BRIEF PROFILE OF VARIOUS LANGUAGE TOOLS

True Type Fonts with Keyboard Drivers

This tool installs 8 bit True Type fonts from different organizations each having its own encoding. This is also accompanied by the respective keyboard driver. With the help of these FONTS & KEYBOARD drivers, the user can type in any application running under the Microsoft Windows platform. Three different keyboard layouts are provided a) Phonetic b) INSCRIPT and c) Typewriter. Some default fonts are getting installed with each keyboard driver as well as in some cases user can install extra fonts. Keyboard drivers are provided in Marathi, Sanskrit, Hindi, Punjabi, Kannada, Malayalam, Oriya, Gujarati, Telugu, Tamil, Nepali, Bengali, Konkani and Manipuri CD in Bengali script by CDAC and other contributors like Cyberscape, IMRC, and CK technology.

Nashir, which is page layout and publishing software, is provided as a true type keyboard driver in Kashmiri, Sindhi and Urdu languages.

![Figure 1: Marathi True type keyboard driver, CDAC GIST](image_url)
Unicode Compliant Open Type Fonts

This tool installs Unicode compliant Open Type Fonts & is intended to be used along with UNICODE enabled Operating systems such as Windows 2000, XP, Fedora Core 9.0+, etc. Since it is a Unicode compatible font all the fonts provided poses the same behavior & hence no compatibility issues.

Open Type fonts will work only with the systems and applications, which are Unicode enabled. The font character is based on Unicode and can support any script. Open Type fonts can have up to 65,536 glyphs. Font files are intended to be cross-platform, and can be used without modification on Mac OS, Windows and some UNIX systems.

Open type fonts are provided in Indian languages as listed below: - Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri (Devnagari and Perso Arabic script), Konkani (Devnagari and Roman script), Maithili, Malayalam, Manipuri (Bengali and Meetei-Mayek script), Marathi, Nepali, Oriya, Punjabi, Sanskrit, Santali (Devnagari and Olchiki script), Sindhi (Devnagari and Perso Arabic script), Telugu, Tamil and Urdu.

IME based approach is given in Bengali, Sindhi (Perso Arabic), Manipuri (Meetei Mayek), Manipuri (Bengali), Kashmiri (Perso Arabic) languages.
Multifont Keyboard Engine

Multifont Keyboard Engine provides support for typing in Indian language using True Type Fonts with different encoding. This Multifont Keyboard Engine supports Phonetic, INSCRIPT, and Typewriter layout. It supports monolingual, bilingual and Unicode font types. It supports Windows 98, 2000, XP and Vista. Document made using this will have to be converted to and fro from various encoding to Unicode and ISCII for compatibility. It is available in languages like Assamese, Bengali, Hindi, Gujarati, Kannada, Telugu, Punjabi, Oriya, Malayalam, and Marathi.
Unicode Compliant Keyboard Driver
This Keyboard Engine provides support for typing in Marathi language using Open Type Fonts. This Keyboard Engine supports Phonetic, INSCRIPT, and Typewriter. It is supported Win 2000 and onwards operating systems. It is available in Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri (Devnagari and Perso Arabic script), Konkani (Devnagari and Roman script), Maithili, Malayalam, Manipuri (Bengali and Meetei Mayek script), Marathi, Nepali, Oriya, Punjabi, Sanskrit, Santali (Devanagri and Ol-Chiki script), Sindhi (Devanagri and Perso Arabic script), Telugu, and Tamil except Urdu.
**Generic font's code and storage code converter**
The Generic Fonts Code and Storage Code converter 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. Priya Informatics has provided Hindi and Marathi language code converter.
And for the remaining languages like Assamese, Bengali, Marathi, Malayalam, Kannada, Oriya and Gujarati, "Generic font's code and storage code converter" has been provided by Cyberscape. It is supported on Windows 98 onwards.
Localized version of Bharateeya (Spread Sheet, Presentations, Word processing & drawing tools)
This includes Indian language localized version of Bharateeya Open Office Suite (Impress, Writer, Calc, Base, and Draw) by CDAC. Following are certain basic features of the same.
   a. No commercial licenses or software compliance issues to worry about
   b. On all major computing platforms - Translated into over 30 languages, available on all major computing platforms (Microsoft Windows, Mac OS X X11, GNU/Linux, Solaris)
   c. The source code for the software is freely available under a full OSI certified license
   d. OpenOffice.org Version 1 is now in use by tens of millions of delighted users worldwide.
   e. The roots of OpenOffice.org go back twenty years, creating a huge wealth of experience
   f. Hundreds of thousands of users have participated in the beta testing of version 2.
   g. With a fully open development process, OpenOffice.org has nothing to hide - the product stands or falls on its reputation
   h. The software looks and feels familiar and is instantly usable by anyone who has used a competitive product
   i. OpenOffice.org is supported by a global community of friendly volunteers, only too happy to provide assistance to newcomers and advanced users alike.

WRITER
Open Office Writer OpenOffice.org’s word processor: use it for anything from writing a quick letter to producing an entire book with embedded illustrations, cross-references, tables of contents, indexes, bibliographies. Auto-complete, auto-format, and real-time spelling check make light work of the hardest task. Writer is powerful enough to tackle desktop publishing tasks such as creating multi-column newsletters, brochures – the only limit is your imagination.
CALC

Open Office Calc is powerful spreadsheet, has all the tools you need to calculate, analyze, summarize, and present your data in numerical reports or sizzling graphics. A fully integrated help system makes entering complex formulas a breeze. Pull in external data using the Data Pilot, and sort it, filter it, and produce subtotals and statistical analyses. Use previews to select from thirteen categories of 2-D and 3-D charts including line, area, column, pie, XY, stock and net with dozens of variants.
IMPRESS
Open Office Impress is the fastest, most powerful way to create effective multimedia presentations. Presentations will truly stand out with special effects, animation and high-impact drawing tools.
DRAW
Open Office Draw will produce everything from simple diagrams to dynamic 3D illustrations and special effects.
BASE
Open Office base enable us to manipulate database data seamlessly within OpenOffice.org. Create and modify tables, forms, queries, and reports, either using own database or Base's own built-in HSQL database engine. With the database tools in OpenOffice.org users can easily access and analyze data, create complex reports or do a mail merge based on a customer database. Many open source and commercial database systems are supported.
MATH

MATH is OpenOffice.org’s component for mathematical equations. It is most commonly used as an equation editor for text documents, but it can also be used with other types of documents or stand-alone. When used inside WRITER, the equation is treated as an object inside the text document. Similarly, one can also insert these into other OpenOffice.org programs like CALC and IMPRESS.

Using MATH, there are three main ways of entering a formula:

- Type markup in the equation editor.
- Right-click on the equation editor and select the symbol from the context menu.
- Select a symbol from the Selection toolbox.
Different versions of open office are provided in different languages. Manipuri (Bengali/Meetei mayek), Santali (Devanagri/Ol-Chiki), Konkani (Devanagri/Roman), Sindhi (Devanagri/Perso Arabic), Kashmiri (Devanagri/Perso Arabic) - version 3.1
Bengali, Sanskrit, Dogri, Nepali, Bodo, Maithili - version 3.0
Marathi, Assamese, Malayalam, Oriya, Urdu – version 2.0
Gujarati – version 2.2
Telugu – version 1.9
Tamil, Hindi – version 1.1.4
Punjabi – version 2.1
Kannada – version 2.0.3

Web Browser (Firefox)
Mozilla Firefox is a free and open source web browser descended from the Mozilla Application Suite and managed by Mozilla Corporation. This is a localized version of Firefox web browser, the Browser That Has It All. Firefox has security, speed and new features that will change the way you use the Web. Some of Firefox’s specialized features include pop-up blockers and advanced privacy settings. We can open more than one website in your browser window, and tabs on the screen allow you to easily switch between them.
Localized Mozilla firefox is available in following Indian languages:-
Urdu (version 1.0.6), Dogri (version 3.0.6), Malayalam (version 1.0.7), Maithili (version 3.0.6), Telugu (version 1.0.6), Nepali (version 1.0.6), Bengali (version 3.0.5), Kashmiri(version 3.5),
Manipuri (version 3.5), Punjabi (version 1.5), Marathi (version 1.0.6), Konkani (version 3.5), and Santali (version 3.5).

(Figure 12: Santali Dev Mozilla Firefox, CDAC)

**Email client (Thunderbird)**

Mozilla Thunderbird is a free, open source, cross-platform e-mail and news client developed by the Mozilla Foundation. Thunderbird can manage multiple e-mail, newsgroup and RSS accounts and supports multiple identities within accounts. Features like quick search, saved search folders ("virtual folders"), advanced message filtering, message grouping, and labels help manage and find messages. Here we are providing localized versions of Thunderbirds.

Following versions of Localized Thunderbird is given in respective Indian languages.-

Dogri (version 2.0.0.16), Malayalam (1.0.7), Maithili (version 2.0.0.16), Telugu (version 1.0.2), Nepali (version 2.0.0.16), Bengali (version 2.0.0.17), Kashmiri (version 2.0.0.22), Manipuri (version 2.0.0.22), Punjabi (version 1.5), Marathi (version 1.7), Konkani (version 2.0.0.22), Santali (version 2.0.0.22),
Multi protocol Messenger (PIDGIN)

Pidgin is formerly named as GAIM. Pidgin is an Instant Messaging (IM) program which lets you log in to multiple accounts on multiple IM networks simultaneously. This means that you can be chatting with friends on MSN, talking to a friend on Yahoo Messenger, and sitting in a yahoo chat room all at the same time. Pidgin runs on Windows, Linux, and other UNIX operating systems. Pidgin supports many features of chat networks, such as file transfer, away messages, buddy icons, custom smilies and typing notifications. We are providing localized versions of Pidgin in various Indian Languages.
Calendar Application (Sunbird)
This is a localized version of Sunbird calendar application. Mozilla Sunbird is a cross-platform calendar application, built upon Mozilla Toolkit. The goal is to provide the users with full-featured and easy to use calendar application that you can use around the world. It's entirely standalone: it doesn't require the bulk of another application, but doesn't offer any integration into other programs.
Page Layout Application (Scribus)
Scribus is an open-source Desktop publishing or page layout application released under GNU. It is designed for flexible layouts and typesetting and ability to prepare files for professional quality image setting equipments. It can also create animated and interactive PDF presentations and forms. It supports most major image formats including scalable Vector Graphics (SVG).
Content Management System (ECKO)

ECKO 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. It provides a platform for creating, using, and sharing information among and across the members of E-Communities. ECKO provides features for storing and managing information from various sources, and provides to the users in an easy-to-use and understand format. This information can then used to build knowledge repositories across various domains and would help in making useful inferences.

ECKO requires the following Open source software to be installed and configured in your system.

1. Apache web server 1.3 or above
2. PHP Interpreter 4.3 or above, configured as a module in Apache web server
3. MySQL Database Server 4.1 or above with UTF 8 format enabled.
SpellChecker

This software is used to spell check Indian language document for spelling mistakes. This tool also gives suggestions in case there are any spelling mistakes found in a text. This can be a useful utility in case of important documents which needs to be perfect in terms of spellings and other grammatical mistakes.

In Malayalam, Gujarati, Marathi, Kannada, Assamese, Marathi and Oriya languages spellchecker has been provided by Cyberscape. This requires Multifont Keyboard Engine to work properly. This spell checker is supported on Win 98 onwards.
In Hindi language Spell checker of CDAC GIST is integrated which is supported for Hindi language documents with true type fonts only and is getting added as a word macro. This utility does not support spell check for English documents.
Hyderabad university has been contributed for spell checker in Telugu language. It spell checks documents of WX Roman or Telugu script.
The spell checker uses the Tamil morphological analyzer as a building block and also has another sophisticated strategy for spell checking using frequent words and employs edit distance approach for generating and ranking alternate words for misspell words.

(Figure 21: Tamil Spell Checker, AU-KBC Chennai)

**Bilingual Dictionary**

This tool includes bilingual dictionary. The user can type English word and find Indian language equivalent of the word and vice versa. This can act as a very useful to users who know English but want to learn Indian language. Cyberscape has been contributed for bilingual dictionary in Assamese, Gujarati, Kannada, Marathi, Malayalam and Oriya languages.
Magnet technology has been contributed for Gujarati Lexicon. It gives Gujarati meaning of English word and vice versa. Soft keyboard has been provided to type text in Gujarati. It also gives the Idioms, opposite words with their meaning and grammatical format, thesaurus information like concept of the word which will give you the general class of selected word, head word, main meaning and other meaning of selected word.
Marathi dictionary from IIT Bombay has been integrated in Marathi CD. This application is supported on Windows 2000 onwards and Linux Fedora core 3 with Indian language pack. This dictionary allows the user to enter the English word and its meaning in both Marathi and Hindi. CDAC GIST has been contributed for Nepali and Hindi bilingual dictionary, CDAC Thiruvananthapuram has given contribution for Malayalam bilingual dictionary and Tamil dictionary from Palaniappa Brothers has been integrated.
In Oriya language Bilingual dictionary from Utkal University has been integrated which has the same features as above dictionaries.

**Microsoft Word Tools**
This application involves following tools as a word macro.

**Number to Word** – This is a utility which gets added in Microsoft word and converts Indian language numerals to equivalent Indian language word.
Smart converter – The smart font converter converts text of one font to another font. It also supports conversion from Unicode to other font encoding.
Sorting Tool – This is a utility which gets added in Microsoft word for sorting a table in Indian language alphabetic sequence.

Spell Checker – By using this utility you can spell check respective Indian language document.
All the above tools require Multifont keyboard engine to be installed on machine for proper functioning. All these tools are made available in following Indian languages by Cyberscape:--:
Assamese, Bengali, Marathi, Malayalam, Kannada, Oriya and Gujarati

**Microsoft Excel Tools**

This tool includes:-

- **Excel Magic** – It transliterates English text to respective Indian language text.

- **Excel Sorting** – It sorts data alphabetically in excel sheet.
• **Excel Converter** – It converts font from one font family to another font format.

All the above tools require Multifont keyboard engine to be installed on machine for proper functioning. These utilities are getting added as a macro in Microsoft Excel and are supported on Windows 98 onward. All these tools are made available in following Indian languages by Cyberscape:- Assamese, Bengali, Marathi, Malayalam, Kannada, Oriya and Gujarati

**Database Sorting Tool**

This software sorts the database tables of Indian language in alphabetical order. This can be done in ascending and descending order. The Indian language database tables as well as simple text files can be sorted. This utility can be very useful especially for sorting large databases and text files which has Indian language data. It supports only Ms Access database.

This has been contributed by Cyberscape for following Indian languages:- Assamese, Bengali, Marathi, Malayalam, Kannada, Oriya and Gujarati
Transliteration Tool

Transliteration is process of converting letters of one language to their phonetic equivalent representation in another. This tool transliterates English database tables to Indian language equivalents. It also converts database form any supported fonts to any other fonts. This also requires Multifont keyboard engine to be installed on users machine before using this application.

This has been contributed by Cyberscape for following Indian languages:- Assamese, Bengali, Marathi, Malayalam, Kannada, Oriya and Gujarati.
**Type Assistant**

This utility provides simple way to type in English and get text in respective Indian language. This tool is very much user friendly and can be mastered easily. This also requires Multifont keyboard engine to be installed on users machine before using this application.

This has been contributed by Cyberscape for following Indian languages:- Assamese, Bengali, Marathi, Malayalam, Kannada, Oriya and Gujarati.
Likhan (Decorative Keyword Maker) is entirely a new conception ever introduced for computer users. It has never been possible to make decorative and colorful titles through mere typing on computer. Now it has become 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 Gujarati and English Languages. These highly decorative fonts provided in Likhan are most effective for Books, Souvenirs, Magazines, Newspaper Headings, Brochures, Booklets, Spencer Designs, Greeting Cards, Name Boards, Wall Posters, Cloth Banners, Publicity and Packing Material, Cinema and TV titles etc.

This tool is provided by IMRC and is available in Gujarati, Kannada, Bengali, Malayalam, Marathi, Oriya, Punjabi and Telugu languages.
Gujarati Language Games/Puzzles

Quick Quiz is basically a puzzle/game in which a user has to match a word from one column to the word in the second column. Help facility is included in the tool itself. It is supported on Windows XP-2 onwards, and does not support UNIX or Linux. This application is given by Magnet Technology.
**Gujarati 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. It provides a blank canvas and a variety of drawing tools to help your child be creative. This application is provided by Magnet Technology.
Oriya language Tutor Package is basically intended for new users trying to learn this language. This has got various chapters to learn Oriya.
In Malayalam language **Compute-Easy** is a Computer based tutorial to enable self paced learning for first time computer users. Voice over in Malayalam, guided practice and highly interactive interface allows the learner to learn and practice through a scientifically designed spiral learning mode. This can also be used by institutions to train the students with minimum faculty intervention.

(Figure 39: Malayalam Tutor Package: Compute Easy, ThomsonCyber)

In Kannada language **Kannada Kali** is a simple multimedia game for learning Kannada alphabet.
Akshara Kali teaches Kannada Alphabets like SWARA, VYANJANA etc., and School childrens will get to know the proper way of writing Kannada alphabets. Animation and Pictures are provided for easy narration.
Text Editor

Padakairali is basically a Malayalam Text Processing System. It handles Text, ISCII, RTF and UNICODE documents. Conversion between the different formats and encodings is also be possible. It supports most commonly used Malayalam fonts. It contains

1. Malayalam Spellchecker.
2. INSCRIPT and PHONETIC keyboard drivers for Malayalam ISFOC fonts.
3. Unicode keyboard driver for Malayalam.
4. Converters for most commonly used Malayalam fonts.
5. Words and frequency count.
7. Numerals to Malayalam words converter.
8. Font independent Find & Replace.
Optical character Recognition System (OCR)

This software system is available for following languages Viz. Hindi, Marathi, Malayalam, Telugu, Tamil, and Punjabi.

Hindi and Marathi: - The name of the software is:"Chitrankan". Since "Chitrankan" is capable of recognizing Devnagari Script, it is available for both the languages i.e. Hindi and Marathi.

Chitrankan is basically a Optical Character Recognizer for Devnagari Script. This will help users to convert scanned text pages into editable text. Chitrankan is a very useful utility. This has been specially trained for C-DAC fonts. It the user has to scan pages having text in some different font, the system has to be trained and the result may differ.

Chitrankan ™ is the first full fledged, full functional, optical character recognition system available for multiple Indian Languages. It has been engineered to meet the requirements of the publishing industry, offices, libraries, institutions and individuals who require optical character recognition systems to convert the printed content to text data which can be stored, edited, searched and sorted on computer.

Chitrankan ™ with its advanced DSP (Digital Signal Processing) algorithms will automatically detect the noise and skew in the image and remove it. This software will also take care of the irregular and
regular images in the text while doing the recognition. This will add up to the accuracy and help retain the column and block information in the image. Combined support for English and Indian language text will benefit the users who have a requirement of Bi-Lingual OCR software.

**Salient Features:**

- Supports Bi-Lingual OCR (Devnagari and English)
- Advanced Auto Noise Removal and Skew Correction
- Supports Auto Text detection and Text Block Marking
- Supports column information restoration
- In build ISCII editor for editing the text with basic editing features like (Cut, Copy, Paste, Find and Replace)
- Phonetic and Inscript keyboard layouts supported
- Spell Checker Support for Hindi and Marathi
- Supports all TWAIN compatible scanners
- Encompasses basic image editing tools
- Encompasses image Flip and Rotate options
- Encompasses image negation
- Supports gray scale, 300 DPI Bitmap, TIFF and JPEG file formats.
- Supports ISM (Rich Text ISFOC File Export) and iLeap (ISCII file export)
- Supports Unicode
- Supports font training for different fonts
Punjabi:
Name of the software is "Gurumukhi Scan" and the version of the software is 1.0.
Gurmukhi Scan’s optical character recognition (OCR) technology accurately and easily converts scanned paper documents and files in Punjabi into editable Punjabi text in font of your choice for use in your favorite computer applications. Gurmukhi Scan is the first freely available OCR system for Gurmukhi Script. If one needs to electronically store and manipulate large amounts of Punjabi text or printed matter such as newspapers, contracts, letters, faxes, price lists or even forms and questionnaires, one will find Gurmukhi Scan program can save a lot of effort.

Gurmukhi Scan 1.0 offers a smart way to increase your work productivity by eliminating the need for retyping.

Gurmukhi Scan's optical character recognition (OCR) technology accurately and easily converts scanned paper documents and files in Punjabi into editable Punjabi text in font of your choice for use in your favorite computer applications.

The main features of the Gurumukhi Scan are:
<table>
<thead>
<tr>
<th><strong>Recognition Accuracy</strong></th>
<th>Recognition accuracy is around 97% for books, photocopies and medium degraded documents and around 99% for laser print outs and good quality documents.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Speed</strong></td>
<td>1 Minute for an A4 size page of text on a Pentium-III system</td>
</tr>
<tr>
<td><strong>Import Image</strong></td>
<td>B/W Images, 24-bit color images, 256 grayscale images.</td>
</tr>
<tr>
<td><strong>Import file format</strong></td>
<td>.BMP</td>
</tr>
<tr>
<td><strong>Output file format</strong></td>
<td>ISCII / RTF file encoded in any one of the popular Punjabi fonts.</td>
</tr>
<tr>
<td><strong>Fonts supported</strong></td>
<td>All non-decorative Gurmukhi fonts.</td>
</tr>
<tr>
<td><strong>Proofing</strong></td>
<td>On-screen verifier.</td>
</tr>
<tr>
<td><strong>Additional features</strong></td>
<td>Inbuilt Spell checking facility.</td>
</tr>
<tr>
<td></td>
<td>Automatic skew detection and correction.</td>
</tr>
<tr>
<td></td>
<td>Upside down image auto detection and correction.</td>
</tr>
</tbody>
</table>

Gurmukhi Scan is the first commercially available OCR system for 'Gurmukhi Script'. Gurmukhi Scan program can save a lot of effort if one needs to electronically store and manipulate large amounts of text or printed matter. The program for Gurmukhi Scan however has a few limitations. Some of the cases where the Gurmukhi Scan may not recognise the text accurately are:

1. Text in Italics.
2. Old Punjabi Text.
4. Low Quality / Noisy Text.

The basic steps of Gurmukhi Scan’s OCR process are:

1. Scan a paper document and store it in the .BMP format.
2. Perform OCR to convert text information into editable text characters.
3. After OCRing, both the Input Text as well as the Recognized Text appear side by side. In Gurmukhi Scan the user can proof read and edit the Recognized Text. An inbuilt Spell Checker is invoked for checking the errors in the Recognized Text.

4. Export the file to the desired location. In Gurmukhi Scan the user can save the recognized text as a .ISCI file, can save it as a Punjabi font encoded RTF file, or can place it on the clipboard.
लघिव पहेली बातधर क्यों लघिव पहेली बातधर

सागिर देखे वातधर हित दिख रहा लघिव पहेली

पीरियरे गैरिे

रुख दी बलिंदे हे पृथव पी। हित हित दिख तारे ला पृथव ताम नृष मी। माँझे जँझण दो से "सिमें बाबु ही घटता उगे उं गुं गुं गुं गुं गुं गुं गुं गुं गुं गुं गुं गुं

परदी मी, बाँझे हित ताम उगे उं पाय, हलफ उगे उं गुं पृथव, दे बीजालः

उं तें हित भवान हलफ माँझे उगे लृणा लृणा लृणा लृणा लृणा है ताम ही ही
Tamil:

This software scans and recognizes Tamil printed text. This software can be trained to recognize a particular font used in a book. Using the information specific to a particular font enhances the recognition capability. Also it reduces the recognition time. Default font information files, provided in the software may also be used for recognition. Parts of the text can be selected for recognition. Some parts of the page, like pictures can also be excluded from recognition. Multiple columns of texts in a page are also taken care of.

PonVizhi can recognise old type letters like naa and nai. They can be seen as they are. They all can be converted to the new type with just one click. The output of the text is in the Standardised TAM font format, and in the widely used RTF file format. Being the first version of this software for Tamil, it has some limitations. Letters with distortions like, broken letters, smudged letters, overlapping letters cannot be recognised in this version. Also the formatting information like bold and italics cannot be identified.

To improve the productivity, special quick user friendly editing features have been provided. A powerful Tamil Spell Checking module of PonMozhi is also included in PonVizhi. This special spell check works not only in TAM and TAB fonts, but also in the font selected by the user.

**STEPS IN USING OCR**

1. SCANNING
2. RECOGNITION
3. CORRECTION
4. SAVING
5. TRAINING
வீடியோ புரிந்திருக்கும், மேலும் புரிந்திருக்கும் ஒரு குறுக்கு. ஏனெனில் எந்த புரிந்திருக்கும் ஒரு குறுக்கும் போது அவ்வாறு செய்ய வேண்டும். என்னும் மூலம் அவ்வாறு செய்யும் போது வேண்டும் யுத்தத்தின் மூலம் என்பது. எந்தவொரு பகுதியிலும் மூலம் அவ்வாறு செய்யும் போது வேண்டும் யுத்தத்தின் மூலம் என்பது. எந்தவொரு பகுதியிலும் மூலம் அவ்வாறு செய்யும் போது வேண்டும் யுத்தத்தின் மூலம் என்பது. எந்தவொரு பகுதியிலும் மூலம் அவ்வாறு செய்யும் போது வேண்டும் யுத்தத்தின் மூலம் என்பது. எந்தவொரு பகுதியிலும் மூலம் அவ்வாறு செய்யும் போது வேண்டும் யுத்தத்தின் மூலம் என்பது.
**Malayalam:**
Nayana, the Malayalam Optical Character Recognition System, converts scanned Malayalam document images into computer editable text. The scanning resolution should be greater than or equal to 300 dpi. The software automatically detects and corrects skew ranging from -5 to +5. The layout analyser identifies text part and image part separately. The supported image formats are BMP, TIFF, GIF and JPEG. The supported output text formats are UNICODE, ISCII, RTF, HTML and TXT. The text editor supports both INSCRIPT and QWERTY keyboards.

**Telugu:**
DRISHTI AN OCR SYSTEM FOR TELUGU AND OTHER IL SCRIPTS
This document describes installation and operating instructions for the DRISHTI OCR System for Telugu developed by RCILTS (Telugu) University of Hyderabad.

SECTION 1: Installation Guide
SECTION 2: Operations Guide

INSTALLATION GUIDE
This section contains instructions for installing DRISHTI OCR system for Telugu script. The following conventions are followed throughout the document.

PREREQUISITES FOR THE WORKING OF THIS SOFTWARE: This requires a recent system that can run Fedora or at least minimally a Red Hat 9 Linux with a Pentium 4 or equivalent processor with at least about 256 MB RAM and 50-60 MB of free disk space. We assume that the user can scan any images which need to be processed and store them in the DRISHTI directories for processing.

LINUX PACKAGES TO BE INSTALLED: GIMP, Image Magick etc graphical image manipulation packages. TCL/TK packages and all the dependencies. Further it is assumed that all the usual libraries etc. for running PERL would also be installed.

1. Mount the CD-ROM provided in an appropriate mount point such as

   /mnt/cdrom or /media/cdrom

2. This document is drishti-readme.txt. The following set of files need to be there for the Drishti system:

   File Size  File Name                        Comment
   15201257   drishti-sept12release.tgz       OCR system
   1369619    fonts.tgz                      Indian Language Fonts
   2087       install-fonts                  Script that installs above
   359        INSTALL-YUDIT-README           Guide to install Yudit
   180360     Pothana2000.ttf                Free Telugu UNICODE font
   4863352    yudit-2.7.6-1.i386.rpm         UNICODE Editor shows Telugu Text

3. Create a directory for installing DRISHTI such as

   /usr/local/packages/drishti. This directory is referred as DRISHTI-HOME hereafter.
4. Add the following line to your .bashrc file or any other shell start-up file in your home directory

```bash
export DRISHTIHOME=DRISHTI-HOME for example use
```

```bash
export DRISHTIHOME=/usr/local/packages/drishti
```

5. Copy all the files listed above to DRISHTI-HOME.

Extract the DRISHTI system by giving the command:
```bash
tar xvzf drishti-sept12release.tgz
```
If successful, it would create the entire DRISHTI software system on your computer.

6. Run the script: install-fonts This will then install and expand the fonts under the package fonts.tgz

7. Follow the directions as given under the file: INSTALL-YUDIT-README and install Yudit Unicode editor and the Pothana2000.ttf fonts package needed for viewing UNICODE encoded Telugu text.

This completes the installation instructions.

**DRISHTI OCR System for Telugu RCILTS (Telugu)**

**OPERATION GUIDE**

**OPERATING INSTRUCTIONS** This prototype version of DRISHTI software included in this package contains only minimal functionality. An extended version with added functionality will be made available soon. This particular version is customized for working with the CDAC Hemalatha fonts.

Prior to starting DRISHTI, ensure that the DRISHTIHOME environment variable is set. It can be verified by typing

```bash
echo $DRISHTIHOME
```

The system should display the DRISHTIHOME directory given during installation on the screen.

**DRISHTI Interface** This section describes DRISHTI graphical user interface. Study the interface to familiarise yourself with the location of different buttons, windows and other objects.

**DRISHTI interface contains four main areas:**

- **Menu Area** Located at the top of the interface, it provides convenient drop-down menus for accessing available functions. It contains the following sub-menus:
1. FILE is a drop-down menu with functions (currently) to SCAN and OPEN images and EXIT from the software.

2. CONFIGURE is a drop-down menu with functions (currently) to choose between C-DAC and SreeLipi font families.

3. PREPROCESSING is a drop-down menu that provides noise-removal and skew-correction functions prior to performing OCR. This menu is currently not initialized.

4. LAYOUT ANALYSIS is a drop-down menu that provides functions for handling multi-column documents and those containing images along with text. Currently these functions are NOT included.

5. RECOGNITION is a drop-down menu with functions (currently) to perform OCR on the input document.

6. POSTPROCESSING is a drop-down menu that provides functions to restore document layout and perform spell-checking (currently not activated).

7. TRUTHING brings up a separate interface to analyze performance of OCR (currently not activated).

Image Region Located below the MENU AREA, this region comprises IMAGE INFO AREA and IMAGE INPUT AREA.

IMAGE INFO AREA displays information about the input image, obtained either by scanning or opening from FILE menu.

IMAGE INPUT AREA shows the input image. The input image shown is scaled down as the actual image is usually very large so that a greater portion is displayed.

Text Region Located below the IMAGE REGION, this region consists of TEXT INFO AREA and TEXT OUTPUT AREA. TEXT INFO AREA provides information such as the number of lines, words, etc. about the text recognized from the input image.
TEXT OUTPUT AREA shows the text recognized by DRISHTI system.

General Information Region Located at the bottom of the interface, this region displays information about the current working and user-settings.

MESSAGE AREA displays short messages about the system operations.

LANGUAGE shows the IL script selected for OCR. Currently it is restricted to Telugu but in the future may include other South Indian scripts and Gujarati.

KEYBOARD shows the keyboard for use to edit the text in TEXT OUTPUT AREA. Currently this feature is not activated.

FONT FAMILY shows the font family used by the recognition engine. It is changed by appropriate selection by the user from CONFIGURE menu.

MODE shows the processing mode of DRISHTI. Currently it is not activated and restricted to AUTO PROCESSING mode.

Starting DRISHTI

Change to DRISHTIHOME directory and type

wish -f drishti gui.tcl

Scanning an input image Select the SCAN item from FILE menu. It brings up an external program, xsane, to scan an image. Use the controls provided by xsane program to scan a document. Remember the name given to the document image while scanning. Once the image is scanned, exit xsane and open the image as described below.

Opening an existing image Select the OPEN item from FILE menu. It brings up a dialog box listing available images in the current directory. Change to appropriate directory and select an image. When done, click on OPEN button. It removes the dialog box and the selected image is shown in the IMAGE INPUT AREA.

TESTIMAGES

The test images are provided under the directory: TESTIMGS. Two test images are provided under this directory. Use the file open menu as described above to open one of the test images. Since the font choice is already set, then we can go to the RECOGNITION menu.

TO RUN OCR
Select the RUN OCR option from RECOGNITION menu. It starts the recognition process which may take several minutes (typically 2 U''4 minutes) for an A4 document containing 1500 - 2000 characters.

When the operation is completed, you will see ISCSI codes in the TEXT OUTPUT REGION. Automatically a conversion is done of the recognized text into UNICODE.
The recognition script brings up the Yudit window where the recognized text can be seen.

Typing Tutor

This is available for Hindi, Marathi, Tamil, Kannada, Telugu, Malayalam and Oriya languages.

Hindi, Tamil, Telugu and Marathi:

The name of the software is "Aasaan Typing Tutor". Structurally and functionally these are similar for Hindi, Tamil, Telugu and Marathi and Oriya languages. So for description we are giving Hindi. The Hindi Aasaan software is a very simple guide to learn touch typing. Today, large number of software engineers and the technical staffs are using only two fingers for typing. The “Hindi Aasaan” software has been prepared for the benefit of such people who have not had the opportunity or time to take up typing training.

By using this software, you can learn the basics very quickly, within 6 to 10 hours. Depending upon your speed, you can move to advanced lessons and be a master of touch typing within a shot period of time.

The specialty of this software is that typing exercises are provided for two different types of keyboard in single software! You will also notice that the lessons are taught in duel language (English and Hindi). Learners can choose between English and Hindi mode of typing depending on their need and interest.
Two types of Keyboard Layout:

1. Hindi Inscript
2. Typewriter (English)

1. Hindi Inscript

This Hindi Inscript keyboard is approved by the Central Government. The prime objective of this software is to make all Hindi computer users learn the easy methods of using this keyboard.

2. Typewriter (English)

English typing method is being taught here.

**Scheme**

Lessons
Test
Games
Help

**Layout**

Hindi Inscript
English
Special Lesson

Numbers
Special characters

Scheme:

By clicking the lesson menu in the scheme or by using Ctrl + L you can select and enter the Learning screen. There you will find exercises where you can choose a lesson and start the typing course. From here itself you can change the required keyboard layout by clicking "change layout" option. At the end of each drill (word, sentence and paragraph drills), a report will be displayed. WPM (Words Per Minute) and accuracy will be calculated and the level of your progress will be spelled out.

Test:
By clicking the **test** lesson menu in the scheme or by using the Ctrl +T you can select and enter the **Typing Test** screen. Here you will have the lessons for testing where you can choose and start a test. Before beginning your test, you should select the duration. At the end a report will be displayed. Also, WMP (Words Per Minute) and accuracy will be calculated and the level of your progress will be spelled out.
Games:

By clicking the **games** menu in the scheme or by using Ctrl +G you can select and enter the Games screen. Here you will find two games – **Invader** and **Keyrobics**. You can choose any one of them and start playing.
1. Invader:

The characters will be dropping down as rain drops from above. If you type these dropping characters correctly, you will score the points. If you type incorrectly, the characters will drop down. You will be loosing your chance. As your score increases, the speed of the characters dropping down will also increase. Thus, you can play the game till you loose 16 such characters.
2. Keyrobics:
The characters which appear on the screen from an unknown place will suddenly disappear. You have to keep on typing those character keys. If you type it correctly, you will score one point. If you type it incorrectly your missing count will increase. On the other hand, at the increase of your score the speed of the characters disappearing will also increase. You can play the game till your missing count reaches 25.

Keyboard Layout

You can change the keyboard layout by clicking layout menu. Green colour indicates the layout selected. When you want to start your course with Hindi typing, Hindi Inscript will be the default selected layout. When you want to start with English then English layout will be selected as default.
Special Lessons

In the Special Lessons section, Numpad exercises and Special character exercises are included.

Numpad:

By clicking the special lesson Numpad or by using the Ctrl + N you can enter the Numpad exercise screen. You can select a lesson and start your exercise. Make sure that the numlock light is on before you start with the Numpad exercise. If the numlock light is off then you should first press the numlock key. You will find this numlock key just above the num keyboard on the left corner.

Kannada:
The software system is Nudi keyboard layout tutor. It will give information about kagunitha, othakshara with example workouts.

Malayalam:
The name of the software is Type-Easy. The Multi-user Typing Tutor in Malayalam and English will enable any individuals or institutes to undergo Malayalam Keyboard Training as per Kerala Government laid down keyboard format or In script Keyboard. Type Easy is based on a 4 stage spiral learning mechanism with Keyboard familiarization, Practice, Speed Test, games and Competition. Extensive reports can also be generated apart from the facility to customize the lessons.
Oriya:
Oriya language Tutor Package is basically intended for new users trying to learn this language. This has got various chapters to learn oriya.
**Text to Speech System (TTS) System**

This software is available for Hindi, Malayalam, Kannada, and Oriya.

**Hindi:**

**Vaachak:** The name of the software is Vaachak. It is basically a Plug-In for Office applications.

**MS Word - Read out of written text**

Text in Hindi in a Word document can be read out with the help of the MS Word module of the Vaachak Office Plug-in. Through a simple, easy to use toolbar, the user can hear out all documents in MS Word through the following steps:

1. Open the Hindi Document which you want Vaachak to read out.

2. Identify the font of the text.

3. Change the font of Vaachak plug-in in font menu to the same font of the selected text.
4. Select the Tempo of the voice from tempo menu if required (125 is recommended)

5. Select the text you want Vaachak to read out.

6. Click on the play button.

If you are not able to see the Vaachak plug-in in your MS–Office and Internet explorer tool bar, please click on View>>Toolbars>>Vaachak. You will now be able to see the plug-in in your MS–Word/Excel/PowerPoint and Internet Explorer toolbar as shown above. Following figures visually depict various buttons used in the Vaachak Office Plug-in and their usage.

<table>
<thead>
<tr>
<th></th>
<th><strong>Play Button</strong>: This button will Read out the Hindi text written in the Microsoft Office documents.</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td><strong>Tempo Menu</strong>: This button will allow user to adjust the tempo of the Vaachak voice fig (a).</td>
</tr>
<tr>
<td>Transliteration</td>
<td><strong>Font Menu</strong>: This button will allow user to select the text font fig (b).</td>
</tr>
<tr>
<td>ColumnWise</td>
<td><strong>Row/Column Menu</strong>: This menu will allow user to read out of the text rowwise or column wise in the MS Excel fig (c).</td>
</tr>
</tbody>
</table>

**MS-Excel – Read out of written text**

Along with being able to read out information present in Word documents, Vaachak can also be used to help in listening to large amounts of Hindi data.
in MS Excel documents. Based on the tabular nature of the information, Vaachak can also be configured by the user to read out information in a row-wise or a column-wise manner.

**Row wise Read out**

1. If user wishes to hear information row wise he has to follow these simple steps.

2. Open the Hindi excel sheet which you want Vaachak to read out.

3. Identify the font in which your data is present.

4. Change the font of Vaachak plug-in in the font menu to the same font as your data.

5. Select the Tempo of the voice from tempo menu if required (125 is recommended)

6. Select RowWise from the Vaachak Toolbar to set row-wise readout of data.

7. Select the cells which you want to be read out.

8. Click on the play button.

---

**Column wise read out**
For column wise read of information user has to follow these simple steps.

1. Open the Hindi excel sheet which you want Vaachak to read out.
2. Identify the font in which your data is present.
3. Change the font of Vaachak plug-in in the font menu to the same font as your data.
4. Select the Tempo of the voice from tempo menu if required (125 is recommended)
5. Select Column Wise from the Vaachak Toolbar to set row-wise readout of data.
6. Select the cells which you want to be read out.
7. Click on the play button.

Text To Speech for Hindi : The name of the second software system is "Text To Speech for Hindi". This is from IIIT Hyderabad. This system can read out Hindi files. For this to work you must have a sound card with speakers.

The TTS uses advanced techniques and is able to produce natural sounding speech in Hindi, (unlike some other systems which sound robotic or machine like). The system is under constant developement.

How to Use:

(Make sure your speakers are on).
1. The Text to Speech (TTS) System is presented here has a Graphical Interface. You can start it as below:

* Goto My Computer
* Double click on the C: Drive icon
* Go to the Folder you have copied the Contents from the CD (IIITH_TTS) as above
* Double click on the icon of IIITH_TTS_DEMO
* This will open a GUI which you can use to run the TTS

2. The TTS can play a UTF8 Hindi File through a command. The contents of the UTF8 Hindi file will be spoken out.

3. To run a sample file supplied with CD do the following:

You may go to the directory where you copied the TTS (say IIITH_TTS)

* Click on the Open button and select a file from the SampleFiles Folder
* Click on the Play Button
* You can press Stop to stop Playing
* You can click on the Display Button to see the Text being spoken out.

It will open in a Notepad/Wordpad and you should be able to read the Hindi Text.
Several Sample file are provided for your testing.

**Malayalam**

The name of the software is "Subhashini". It is a Text-to-speech system for Malayalam. This software reads out malayalam text input intelligibly. This software is integrated with a text editor. The editor is capable of handling the most popular font codes and standard formats like ISCII and Unicode. The text editor supports INSCRIPT standard keyboard for the entry of Malayalam data.

**FEATURES**

- Intelligible speech output.
- Works on Windows 98/NT/2000 platform
- Onscreen active keyboard.
- Automatic handling of Numerals and Standard Abbreviations.
- Integrated text editor with both ISCII and ISFOC support.
- Can read out text encoded in popular Malayalam fonts.
- INSCRIPT keyboard support.
Kannada

The name of the software system is "Mitra".
Mitr is a Unicode based Kannada Text To Speech system(TTS). This is a stand alone application. Here you can type the text in Kannada and get the desired output.
Pravakta is also a TTS System but comes as a plugin in Microsoft Word. This is useful specially when you already have some file created and you want to read the same.

**Oriya**

The name of the software is "Priyambada". It is a Text-to-speech system for Oriya. This software reads out Oriya text input intelligibly. The editor is capable of handling the most popular font codes and standard formats like ISCII and Unicode. The text editor supports INSCRIPT standard keyboard for the entry of Oriya data.
Library Management System

It's a Kannada Library Software Management System. Books are categorized on different subjects. The search option is provided based on author, book title, category and etc., Various reports can be generated here.
Kannada Personal Utilities

- Vilasa Suchi

Vilasa Suchi is basically a Address management software. Here you can get different types of Reports.
• **Karya Suchi**

One can record their program of visiting place and other personal official programs. The parameters are like contact person, place, date, remarks and etc., and you can able to get a report of all programs.
• **Dina Suchi**

Depending on your birthday input date, it can calculate the years and months of a person. It also contain the Kannada Calendar.
Kannada Games/ Puzzles

• AnkiVinoda

Anki vinoda is a puzzle in which the player's aim is to arrange numbers in sequential order in the minimum possible moves. Help is provided in the software.

• Pada Vihara

A kannada Puzzle software based on the subjects place, country, scientist and etc. Player name and scores are generated on the screen. One can drag the answer on the screen and result will be announced with sound effect.
**Kannada Logo**

LOGO is a programming language for kids.

KanLogoHelp.chm is the Kannada help file which is available with the software. Just double click on this file from Windows Explorer to open it. If you are already familiar with LOGO language, read KannadaLogoKeywords.pdf file to know the Kannada equivalents of LOGO keywords.

To install the program double click on Setup.exe file and follow on-screen instructions.

The program uses Nudi keyboard driver. There is no need to start the Nudi keyboard driver separately to use KannadaLogo. The Nudi driver will be started by KannadaLogo itself.
Kannada Seamless E-mail Client
This is a utility for sending the emails and documents in Kannada in a very easy fashion. The user who wants to send the kannada file has to right click the icon and select "send to" -> "GIST-Mail Client". This will pop up the list of fonts available to be sent with the mail. Follow the steps given in the utility. The user at the receiving end will be getting the font as well as the document as an attachment. Similarly for sending the mails also.

**Integrated Word Processor**

Since Indian Languages forms the core of the India, there is a growing need for word processor with built in support for Indian Languages. Swarnakriti allows one to compose in Indian Languages in Hindi, Marathi, Tamil & Gurmukhi as well as in English document with basic editing features support. Swarnakriti has special features besides having common features of a Word Processor, like an integrated powerful Spell-Checker for Hindi which works on Unicode format, Dictionary based Transliteration, Calendar & Scientific Calculator.

The Software interface comprises of following main components

1. **Menu Bar**
2. Tool Bar

Hindi Language Transliteration Tool
BharateeyaOO.o INDIC TRANSLITERATOR (English To Indian Languages)
Indic Transliterator is a transliteration tool for English to Indian languages for Openoffice.org Writer. Currently, it supports transliterations from English to six Indian languages namely Hindi, Tamil, Telugu, Kannada, Malayalam and Gujarati. Transliteration is a process, in which the user actually types in a language which he is used to (English in our case) but the words will be of target language (hindi, tamil etc).

eg: jai hind !!! (typed in english but hindi words), when transliterated it will get converted into actual hindi words in devanagari script. To get this transliteration please click on the "Indic Transliterator" option in the menu as shown in the figure above and select the desired language.

The base for Indic Transliterator is ICU (International components for Unicode) by IBM. English (Latin) to Indian language transliteration is two step process in ICU. First English to InterIndic conversion and then InterIndic to particular Indian language. The mapping has been changed and unique features added to make it more user friendly. Following are mapping tables used for transliteration.

Note(for all tables):
1."/" is used to show options, eg: A/aa means both A and aa will be mapped to आ
2.Digits will be mapped to corresponding digits in Indian language.
3. Where ever possible, capitals (B,F,M,P,Q,V,W) are mapped to small letters.

1. Hindi Table
| Punjabi Morphological Analyser and Generator | 105 |

| o  | .N | च | ch | य | y | र | .r | / |
| oː | .h | ः | j | ल | l | ए | .e |  |
| आ | a  | झ | jh | य | w/v | .s |
| आ | A/aa | ः | ~n | य | sh |
| इ | i  | ट | T | ए | Sh/ss |
| ई | I/ii/E/ee | ठ | Th | S | s |
| इ | u  | ड | D | ए | h |
| ऌ | O/oo | ढ | Dh | .b |
| ओ | RRi | ण | N | .a |
| ख | LLI | त | t | ओ | .e |
| ए | e  | ए | th | ओ | .o |
| ऐ | ai | द | d | HH |
| ओ | o  | ध | dh | ओ | OM |
| ओ | au | न | n | क | .k/q |
| ख | k  | ध | p | ख | K |
| ख | kh | ध | ph/f | ग | G/.g |
| ग | g  | व | b | ज | J/.j |
| घ | gh | भ | bh | झ | .D |
| ङ | ~N | म | m | ठ | .Dh |
Working of Punjabi Morphological Analyzer and Generator

This tool provides following options to the user:

- **Know the grammatical information of a word**
  Here, the user needs to specify a word (or choose from the list), its word class (if known) and then the tool will search for that word in its database will return its root word, word class and other grammatical information depending upon the word class. Like for nouns, it will return its gender, number, and case.

- **Know all the word forms of a root word**
  Here, the user needs to specify a root word (or choose from the list), its word class (if known) and then the tool will return all its word forms (Paradigms) along with their respective grammatical information.

- **Generate word from grammatical information**
  Here, the user needs to specify a root word (or choose from the list), its word class, and other required grammatical information depending upon the word class selected. Then tool will generate the word form(s) based on that information, if possible.

The function of ‘Find similar word’ button next to list of words is explained here. If no search results are found for a word then user can click on this button to search for words in the word list similar to that given word, similar in the sense that taking into account the common mistakes made by users like omission of (S), (H), (W), and (N), and replacement of R with r and vice versa. It will also be used to handle omission of pairi bindi (L) in KL, gL, jL, PL, and fL.

Giving input in Gurmukhi script (script used by Punjabi language)

For giving input to the tool i.e. entering a word, user is provided with two options. Either he/she can directly enter the word from keyboard or he/she can take the help of on screen keyboard provided with the tool. Using the on screen Punjabi keyboard provided with the tool user can enter the word by just clicking on the desired Punjabi alphabets on the Punjabi keyboard.

The key map used for entering the words from keyboard is Phonetic based. It uses ‘Akhar’ font for displaying Punjabi words or taking input from the user. The key map used is provided below:

<table>
<thead>
<tr>
<th>A b c d E f g h i j k l m n o p q r s t u v w x y z</th>
</tr>
</thead>
<tbody>
<tr>
<td>A b c d E f g h i j k l m n o p q r s t u v w x y z</td>
</tr>
<tr>
<td>A B C D E F G H I J K L M N O P Q R S T U V W X Y Z</td>
</tr>
</tbody>
</table>
Marathi WordNet

The Marathi WordNet Browser is a Graphical User Interface to access the Marathi WordNet lexical database. The users input words in Unicode and the results are also shown unicode. Thus the software requires no font installation. The results that are displayed are the synsets i.e. synonyms of the entered word. Presently the semantic replations are not displayed.
There is 1 sense of आई as NOUN (नाम) :
1. आई, माय, माउली, माउली, मातोश्री, मातुश्री, जननी, जनमदश्री; जन्म देनारी सो, "आई हो मुलाची पहिली गुरु आहे"
SanskritPradipika aims to provide a leisurely introduction to the Sanskrit Language. The intended audience is adult learners, familiar with English. The package is meant to be a primary source for those who are reading without the benefit of a teacher, and a supplement for students embarking on a first course in the language at a University.

The software is structured in the form of a book: the Contents page is accessible from almost every screen. The page has buttons which quickly bring you to a desired Chapter, or to any screen of one; and there are more buttons at the bottom of the page to lead you to the Preface, Dedication and Acknowledgements.
This tool provides search facility in Amarakosa in many Indian languages including Sanskrit & English.

Amarakosa, the Sanskrit thesaurus developed by Amarasinha in 4th CE has influenced modern lexicographic techniques in quite the same way as Panini has done to generative linguistics. There have been attempts to put the text of Amarakosa online or in digital formats. But there has been no attempt to create a version of this work which not only lets user interactively build Unicode database but also lets them search and test and do text analysis based on Amarakosa.

The system has the following features -
- facility for online data entry/editing by language experts
- includes Sanskrit, Hindi, Kannada, Punjabi, Bangla, Oriya, Assamese, Maithili and English. Scope for other major Indian languages
- multilingual data is being stored in Unicode
- stores up to 50 synonyms with category, gender, number information and detailed glosses
- cross-referencing among synonyms
- search capability in the supported Indian language
- ontology display

**Sanskrit Language Mahabharata Indexer**

This Mahabharata index allows for various kinds of searches of
Mahabharata key words and gives detailed references. It also allows searching words in other online Sanskrit lexical resources.

This web-application has been developed using Java servlets on Apache Tomcat and RDBMS techniques using MS SQL server 2005 in unicode. The application allows three kinds of searches -

☐ Direct Search - user can enter key word in utf-8 and get all the references and details from Mahabharata
☐ Alphabetical search - user can click on a Devanagari alphabet to get the index of words beginning with that alphabet
☐ Search by the structure of the text - user can click on "parvas" > "upa-parvas" > "adhyayas" > "akhyanas" to get the index

Clicking on an indexed word will display the details for that word and also a facility to search that word in some other online lexical resources.
Sanskrit Language Sandhi Generator

This tool generates "Sandhi" between given two Sanskrit words. Basically Sandhi is a term which is used for variety of phonological processes that occur at morpheme or word boundaries. The examples include the fusion of sounds across word boundaries and the alteration of sounds due to neighboring sounds or due to the grammatical function of adjacent words.

As shown in the image below, we can enter the words to form a sandhi using the keyboard available on the website and separating these by the help of + sign. Alternatively, You can type fast using inbuilt iTRANS-Devanagari unicode converter.
Sanskrit Language Subanta Analyzer

Sanskrit is a heavily inflected language, and depends on nominal and verbal inflections for communication of meaning. A fully inflected unit is called pada. The subanta padas are the inflected nouns and the tinanta padas are the inflected verbs. Hence identifying and analyzing these inflections are critical to any further processing of Sanskrit. The subanta padas are the inflected nouns and the tinanta padas are the inflected verbs.

The results from the subanta analyzer for the input text fragment

***आमस्या आत्मकश्चः***
चपला: बालकः आग्राणाम् उद्वर्णं गच्छनि। तत्र आग्राणानि पश्चाति प्रसन्नः: च मनवनि।

are displayed as follows.

आमस्य_SUBANTA_SUBANTA आत्मकशा [आत्मकशा (स्वविलोकः) + सू, प्रथमा, एकवचन] [*_PUNCT]
चपला: [चपल (पुलिलिङ्गः) + जस्, प्रथमा, वद्वर्तचन] बालकः [बालक (पुलिलिङ्गः) + जस्, प्रथमा, वद्वर्तचन] आग्राणाम [आग्र (पुलिलिङ्गः) + आम, श्री, वद्वर्तचन] उद्वर्णं [उद्वर्ण+अम, दृशीया, एकवचन]
[गच्छनि_VERB] [*_PUNCT] [क्र_AV] आग्राणानि [आग्राण+जस्/शस् प्रथमा/दृशीया, वद्वर्तचन]
[पश्चाति_VERB] प्रसन्नः. [प्रसन्न (पुलिलिङ्गः) + जस्, प्रथमा, वद्वर्तचन] [क्र_AV] [मनवनि_VERB]
[*_PUNCT]

Those colored blue are non subanta categories and those colored red are possible errors. The default black colored ones are the subanta padas analyzed.
Sanskrit Language Subanta Generator

This tool generates nominal inflected forms in seven vibhaktis

Sanskrit Language Tinanta Analyzer
Verbs constitute an important part of any language. A sentence indispensably requires a verb to convey complete sense. Given the importance of verb and verb phrases in any linguistic data, it is necessary to develop a proper strategy to analyze them. Creating lexical resource for verbs along with other parts of speech is a necessary requirement. Sanskrit is a highly inflectional language. It is relatively free word-order language. The semantic inter-relation among the various components of a sentence is established through the inflectional suffixes.

This tool analyzes Sanskrit verbs and returns constituent information like dhatu, lakaara, person, number etc.
**Sanskrit Language Tinanta Generator**

This tool generates Sanskrit verb forms in various tense & aspects

![Sanskrit verb-form generation](image)

**Sanskrit Language Karaka Analyzer**

This tool analyzes Sanskrit sentences and labels karakas
**Sanskrit Language POS Tagger**

This tool tags the Sanskrit text with Parts of Speech (POS) and other grammatical information
Sanskrit Language Letter Pronunciation

This tool helps learning pronunciation of Devanagari letters with help of words and pictures. This is very essential and important for children and anyone who wants to learn Simple words in Sanskrit.
Sanskrit Language Alphabet Writing

This application teaches Devnagari writing. The software system is basically in the form of Animation, which shows the flow, and direction of curves to form a particular character. This is also important and useful Animated software which children and new learners can utilize.

Sanskrit Language Panchatantra Stories

This application presents animated Panchatantra story in multi-media form in Sanskrit and Hindi.
Baishakhi Linux 1.0

Society for Natural Language Technology Research (SNLTR) developed Baishakhi Linux 1.0 in collaboration with MAT3 Impex and IIT Kharagpur. This is a free Bangla Linux that has been built over Ubuntu 8.04 distribution.

All computer related decision making and office activities, such as document writing, preparing presentations, web browsing, sending and receiving emails as well as spreadsheet calculations can be carried out in Bangla using this distribution. All Bangla compound words can be viewed and written in Baishakhi Linux, and this special feature distinguishes it from the other localized Linux distributions. Even in spreadsheet application (an office suite for calculation) all types of mathematical calculations (addition, subtraction, multiplication, division etc.) can be done in Bangla including fraction number, which is also a unique feature of this distribution.
Some Salient Features of Baishakhi Linux are as follows:

☐ Based on Ubuntu-8.04-LTS (the latest)
Ubuntu Linux distribution is ranked 'one' according to 'hits per day' statistics and also it is a lightweight distribution. That's why Ubuntu is chosen as the base of Baishakhi Linux distribution.

☐ Bangla Unicode supported Fonts as default font
As the distribution is built with Unicode compliant, Unicode supported fonts are set as default font. So that all languages especially Bangla can be written and viewed properly.

☐ Default locale as Bangla Locale (bn_IN.UTF-8)
The Unicode compliant Linux distributions are generally UTF-8 encoding supported. If the locale of the distribution is extended with UTF-8, then the distribution is said to be Unicode encoded. That's why the locale is set with UTF-8 by default. Though it is UTF-8 encoded, files created with UTF-16 encoding can be opened and viewed in this distribution.

☐ Pre Activated 'Keyboard Indicator' on Gnome Panel.
For typing in more than one language, an application 'Keyboard Indicator' is to be set in Gnome Panel and the setting process is not so much easy to the Linux user of beginner
level. To alleviate this difficulty, the application has been set as default with Bangla and English language.

- Writing and viewing Bangla compound words. All the possible Bangla complex text layout can be written and viewed into open office application of this distribution. In general open office applications, all Bangla Compound words (Bangla complex text layout) cannot be written or viewed, But in this distribution Open Office application has been customized to view and write all the Bangla compound words.

- Viewing GUI Application In Bangla & English Messages of all GUI based applications are in both Bangla and English Languages to make it more user friendly.

- Installation Process can also be done in Bangla The installation process is also fully in Bangla, which is a very helpful feature for the people, who are well versed with Bangla Language but beginners to Linux use.

- Customized Open office Application Using Baishakhi Linux one can prepare documents presentations as well as spread sheet in Bangla. In case of general Open office spread sheet application, mathematical calculations like addition, subtraction, multiplication, division etc cannot be done in Bangla. But in Baishakhi Linux, Open office application has been customized to do all these operations (addition, subtraction, multiplication, division etc.) in Bangla. Sorting and searching can also be carried out.

- Customized Firefox Viewing the sites developed in Bangla language had been a common problem for Firefox. But in Baishakhi Linux, all the sites (developed in Bangla language) can be viewed properly in Firefox by default. Printing Bangla complex text layout from Firefox had also been a common problem, which has been solved in this distribution. Some special features of Open Office Application, which is integrated with Baishakhi Linux:

  - Open Office Word Processor:
  - All the Bangla compound words can be viewed and written in this application.
  - Here is an example of representing some typical complex Bangla characters as well as compound words.
Open Office Spreadsheet Application:
**Firefox Application:**
An example of viewing Bangla sites is shown below
Bengali Font Encoding Converters

The software tool performs the following encoding conversion operations:
   i) iTrans to Unicode
   ii) Unicode to iTrans
   iii) ISCII to Unicode
   iv) Unicode to ISCII
   v) ISCII to iTrans
   vi) iTrans to ISCII

This software is available for Windows Operating systems and have been developed using Programming Language JAVA 1.6.

HOW TO USE THE CONVERTER
1. Download the "ConvertersAll.zip" file.
2. Unzip the file using Winzip or Winrar software tools.
3. A new directory, "ConvertersALL" will be created containing the following files:
   a) encodingConverter.bat
   b) encodingConverter.jar
   c) Readme_ConvertersALL.txt
   d) bengISCII_UNI.txt
4. Within the directory, double click the "encodingConverters.bat" file.

5. A GUI will open.

6. From the combobox, select the type of conversion you want to perform i.e "Unicode to ISCII", "ISCII to Unicode", or "iTrans to ISCII" etc.

7. Enter the full path of the input file in the first text box (or Browse) For example : "C:\Documents and Settings\Administrator\Desktop\converter\input.txt"

8. Enter the full path of (or Browse) the output file in the second text box

9. Click on the "Convert" Button.

10. A messagebox will appear confirming the conversion task.

   * NOTE: Make sure JAVA 6 or its higher version is installed in your machine.

   ** NOTE: The present system handles only text files (created by Notepad .txt or Wordpad .txt or .rtf) format as input and output.

   *** NOTE: The bengISCII_UNI.txt, bengUNI_ISCII.txt, Uni_iTrns.txt are some of the essential files which should not be tampered.

**Baishakhi Soft Keyboard for Windows**

This is a kind of keyboard with the help of which you will be able to type Bengali very easily. Since the tool has short learning curve, the new user will be able to type in few minutes. The snapshot of the tool is given below. This will come up after you install the keyboard on a machine and then run from the start menu or by double clicking its icon.
Baishakhi Keyboard driver for Windows

This keyboard driver is IME based INSCRIPT keyboard. This can be used for typing in any application. The data created will be Unicode data. Kindly see below the image of where to find the Keyboard Shortcut to start the typing.

After the installation process, described above, two icons will appear as shown in the following figure.

![Keyboard Shortcut Icon](image)

**EN** Click on this icon for changing English (United States) to Bengali (India) or vice-versa. After clicking on this icon, you can change languages as shown in the following figure.

**Note that you achieve the same by pressing** `<Shift> + <Left Alt>` **keys simultaneously.**
To view “Baishakhi” or “Inscript” layout, click on this icon and select any one of the layouts, as shown in the following figure.

After completion of the above steps, you can type in Bengali in Microsoft Office Word, Microsoft Office Excel, Microsoft Office PowerPoint, Open Office Writer, Open Office Impress and Open Office Calc in a Windows or Vista machine.

**Note: The above set of Softwares are also available for Linux systems as detailed Below:**

3. Baishakhi Keyboard Layout for Ubuntu 7.04
4. Baishakhi Keyboard Layout for Ubuntu 8.04
5. Baishakhi Keyboard Layout for Fedora 9
6. Unicode compliant Open Type Fonts for Linux.
7. Soft Keyboard for Linux
8. Converters for Linux.
**Mahabharat Indexer:**
The application allows three kinds of searches -
- Direct Search - user can enter key word in utf-8 and get all the references and details from Mahabharata
- Alphabetical search - user can click on a Devanagari alphabet to get the index of words beginning with that alphabet
- Search by the structure of the text - user can click on "parvas" > "upa-parvas" > "adhyayas" > "akhyanas" to get the index Clicking on an indexed word will display the details for that word and also a facility to search that word in some other online lexical resources

**Urdu design guide and PASCII**
This is a PDF document which talks about PASCII (Perso Arabic Standard Code for Information Interchange) design. It describes all the characters required for representing a PA script. There is another document which describes the standardization efforts made for PA script which has information on Characteristics of Perso-Arabic languages, Standardization of Glyph Set, The 16-bit Nastaliq and Naskh font for Urdu & Kashmiri, Urdu Alphabets, Vowels, etc. There are also some documents which

**Urdu Resources**

**Urdu Encyclopaedia**
This is a Urdu Encyclopaedia which is divided into six parts. Each of the parts has entries on different and varied subjects and topics. This is quite exhaustive encyclopaedia and the total size is around 1.38 GB.

**Telugu Word Application Package**-
This is a package of different following software utilities
  - Telugu Word Generator.
  - Telugu Word Analyser.
  - CALTS Multilingual Lexical Accessor.
  - CALTS English-Telugu Bilingual Dictionary.
  - ANU Telugu-Hindi Bilingual Dictionary.

**Word processor with spell checker**
Akshara Telugu Language Word Processor with Spell Checker for Windows 98, 2000 and above. AKSHARA: Advanced Multi-Lingual Text Processor. AKSHARA encodes texts in a standard character encoding scheme such as ISCII or UNICODE. In AKSHARA attributes are included in an open XML style markup language called Extensible Document Definition Language (XDL) developed by us. This makes it easy to convert to and from various other encoding schemes thereby ensuring highest levels of portability and platform independence. AKSHARA is platform independent - you can use it on MS Windows, Linux and many other platforms. AKSHARA is also robust and reliable – you can comfortably work with large documents without worrying of silly restrictions such as line lengths. AKSHARA has been successfully used to develop a 10 Million word corpus of Telugu.

Telugu Language Publications
This is a collection of Telugu Language Publications on varied subjects such as "Right to Information Act" and "National Employment Guarantee Act 2005". These books are written in Telugu language.