5.3 Multilingual Text Editor

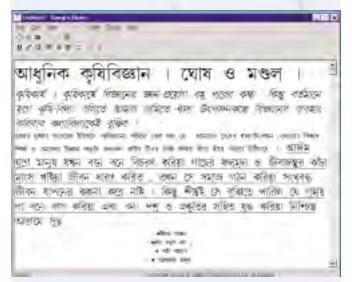
5.3.1

- Name of the Technology: Bangla Editor
- Nature of Technology: Knowledge tool (Bangla Word processor)
- Level: (Product / Technology / Sub-system)
 Product
- Product including Basic block diagram, Algorithm used, O/S used, Front-end / user nterface, and Specification of the Technology / Product: A Bangla text editor with the following feature has been developed where text document in Bangla can be created, saved (both in font code and ISCII code) formatted (bold, italic, size variation, underline etc.,) with different alignments (left, right, center). The online spellchecker mentioned above has been attached with it.

Supporting platform(s): Windows 95/98/2000/NT

Memory requirements: 2MB or above

 Representative Snapshot / screenshot of the Technology / Product: Screenshot of Bangla Editor:



Scalability / Portability / Expandability: Portability: The system can run in WIN9X platform.

- Readiness of Transfer of Technology (ToT): Yes
- Availability of documentation: Yes
- Testing of the Product / Technology : Locally tested
- ▶ IPR / Open-source: Not open source
- Potential Beneficiaries: Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.
- **User-agency tie-up**: None
- Name and address of the Resource Person:
 B. B. Chaudhuri, Professor & Head,
 Computer Vision & Pattern Recognition Unit
 Indian Statistical Institute
 203 Barrackpore Trunk Road
 Calcutta 700 108, INDIA
 E-mail: bbc@isical.ac.in

Web: http://www.isical.ac.in/rrc-bangla



5.3.2

- Name of the Technology: Likhari The Bilingual Punjabi English Word Processor
- Nature of Technology: Knowledge tool (Software for Punjabi Language)
- Level: (Product / Technology / Sub-system)
 Product
- Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product: "Likhari - The Bilingual Word Processor " supports word processing under the windows environment and allows typing and processing in Punjabi Language through the common Remington, Phonetic and Alphabetical Keyboard layouts. It has MS-Word compatible features and commands. It provides a number of features that make the use of Punjabi Language on a computer easy and provides a number of tools to increase the efficiency of the user. These tools include Bilingual Spell Checker with suggestion list, onscreen keyboard layouts with composition reference for Punjabi Language typing, bilingual search and replace, sorting as per the language alphabetical order, onscreen font conversion utility for more than 60 commonly used Punjabi fonts and technical glossary.

Likhari is the only word processor for Punjabi language that addresses the word processing requirements typical for Punjabi language, such as Punjabi Spell Checker, Punjabi Dictionary, virtual Punjabi keyboards, Punjabi sorting and Punjabi font conversion utilities.

Features of Likhari:

- Very simple user interfere
- Online active Keyboards for users who do not know how to type in Punjabi.
- Choice of Keyboard layouts with composition reference.
- Bilingual Spell Checker for Punjabi and English.

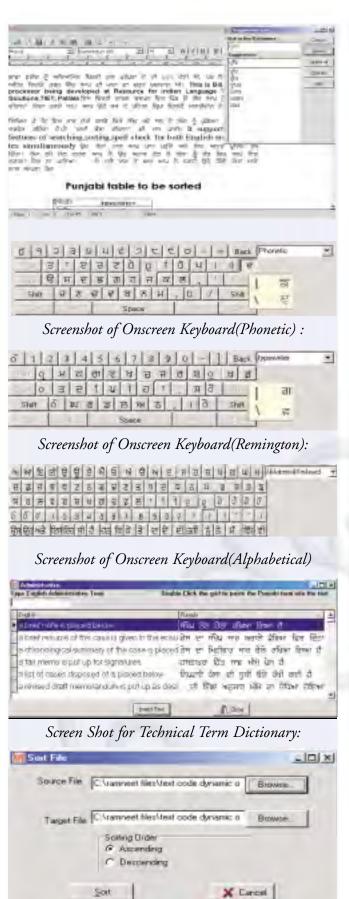
- Bilingual Search and Replace.
- Support for sorting the text in English or sorting of words in Punjabi as per the language alphabetical order.
- Support for almost all popular Punjabi fonts and Keyboard layouts.
- ISCII Compatible.
- Support for features like Tables, Word, Count, Print and Print Preview
- Online Technical glossaries.
- Support for .ISCII, .TXT, .DOC, .RTF and .HTML etc formats.
- Extensive help at various levels to make it easy for the user to learn.
- Supports more than 60 different Punjabi fonts. Enhancements in Likhari as compared to contemporary Indian Language Word Processors:
 - Wide ranges of fonts commonly used by typists are supported in contrast to the forced usage of the available fonts in contemporary softwares.
 - Spell Checking with suggestion list in Punjabi and English.
 - Sorting alphabetically in Punjabi as per the language rules.
 - Onscreen Keyboard in Remington, Phonetic and Alphabetical layouts for Punjabi.
 - Onscreen font conversion utility for more than 60 most common Punjabi fonts.
 - Technical glossary.
 - Find and Replace in Punjabi Language for any font.

Platform: IBM Compatible PC with 486 or Pentium CPU

Operating System: Windows 95,98,2000, XP

Programming Interface: Delphi

Representative Snapshot / screenshot of the Tchnology / Product: Screenshot of Likhari:

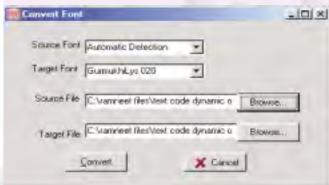


Screen shot for Sorting in Punjabi

Screenshot for Find & Replace in Punjabi



Screen Shot for Font converter utility:



- Scalability / Portability / Expandability: Portable among any of the Windows environment Win 95, 98, Window-ME, Win-2000, Win-NT.
- Readiness of Transfer of Technology (ToT): Yes
- Availability of documentation : Yes
- Testing of the Product / Technology: Testing done
- IPR / Open-source : IPR obtained
- Potential Beneficiaries: Typists, Linguists, Students, Teachers, Punjabi Learners, Educationists, Office Personnel, Corpus Developers etc.
- User-agency tie-up: Solicited
- Name and address of the Resource Person:
 Dr. R.K. Sharma, Coordinator
 Resource Centre for Indian Languages Technology
 Solutions Punjabi (RCILTS Punjabi),
 School of Mathematics and Computer Applications
 (SMCA), T.I.E.T, Patiala 147004 Punjab (India).
 E-mail: rksharma@tiet.ac.in
 Web: http://punjabirc.tiet.ac.in/rcentrei.asp



5.3.3

- Name of the Technology: Word Processor (Oriya, English and Hindi)
- Nature of the Technology: Knowledge tool (Application oriented)
- Level: (Product / Technology / Sub-system): Product
- Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / **Product:** The Word Processor (WP) avails facilities like other word processors such as MS-Word, LEAP- Office, ANKUR-Patrika etc. English phonetic transliteration module allows the user to type in phonetic English to get a corresponding output in Indian language script (Oriya-"ORTTSarala", Hindi-"DV-TTYogesh") that enables the users to type the way they speak. It also avails Spell checking for English, Hindi and Oriya language with a facility of adding new words as per user requirements into the existing dictionary. For Oriya language, it also provides auto correction for phonetically similar characters, Oriya Grammar Checker for detecting the grammatical mistake in the sentence with appropriate suggestions. It also allows integration of graphics with text, transliteration between Indian scripts (Oriya, Hindi) and help option for phonetic typing etc. The WP is developed using Java, Java Swing for both Windows-98/2000/ NT and the Linux O/S.
- Representative Snapshot / screenshot of the Technology / Product:



- Scalability / Portability / Expandability: All
- Readiness of Transfer of Technology (ToT): Yes
- Availability of documentation : Yes
- Testing of the Product / Technology : Under progress
- ▶ IPR / Open-source: Not open source
- Potential beneficiaries: Common man, Offices and Educational Institutions
- User-agency tie-up: Solicited
- Name and address of the Resource Person:

Dr (Ms) Sanghamitra Mohanty RC-ILTS-Oriya, Dept. of CSA, Utkal University, Bhubaneswar, Orissa

E-mail: sanghami@rediffmail.com Web: http://www.ilts.utkal.org

5.3.4

- Name of the Technology: The Perso-Arabic Word Processor - NASHIR
- Nature of the Technology: Knowledge tool (Application oriented)
- Level: (Product / Technology / Sub-system): Product
- Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product: The Perso-Arabic (mainly Urdu, Sindhi and Kashmiri) word processor (Nashir) is designed to be easy to use and create Perso-Arabic documents, and at the same time powerful enough to layout complete newspapers and magazines in Perso-Arabic languages like Urdu, Sindhi and Kashmiri. Each document of the Perso-Arabic word processor (Nashir) consists of a number of pages. On each page of the document, you can place items like text blocks, graphics, etc.

The Perso-Arabic word processor (Nashir) is going to be supported with the Urdu Spellchecker, a prototype of which has been developed by CDAC. Following is a list of features supported.

- Supporst Nastaliq True Type font. (Urdu and Kashmiri scripts)
- Supports Naskh fonts. (Urdu, Sindhi and Kashmiri scripts).
- Supports phonetic keyboards for Urdu and Kashmiri
- Inscript Keyboards designed for Urdu, Sindhi and Kashmiri
- Multiple script support (write in Urdu, Sindhi, Kashmiri, English in the same document)
- Multiple text attributes supported
- Table Object support
- Column layout in pages supported
- Search & Replace Perso-Arabic data
- Implements PASCII storage

- OLE object insertion suported
- Drawing objects like line, circle, rectangle, etc provided
- Provides Text-wrap around objects
- Exports text as Images, HTML pages, etc. More Features
- Bilingual support

Supports Urdu, Sindhi, Kashmiri and English languages.

• Transliteration

Transliteration engine (uTrans) has been implemented as a plug-in. With the help of uTrans, it can insert an aci file transliterated into Urdu document.

Spellcheck

Spellchecks Urdu document with the help of Urdu Spellchecker (prototype) implemented by C-DAC. The Spellchceker uses a tagged dictionary of about 20,000 entries.

• Save As HTML

The user can save the document as HTML page, and thus Naskh as well Nastaliq scripts can be viewed on the Internet.

Table

A Table control is provided to feed tabular information in the document. Cell selection and copy paste operations supported.

• Keyboards

C-DAC has designed phonetic as well as inscript keyboards for the Urdu, Sindhi and aKashmiri languages.

Kerning

Kerning feature is provided to manually adjust a text wherever required. Both Horizontal and Vertical kerning is possible.

• Multiple Text attributes

Easily change a text block attribute by simply selecting it and choosing an attribute. Define



Style templates to directly apply a set of attributes.

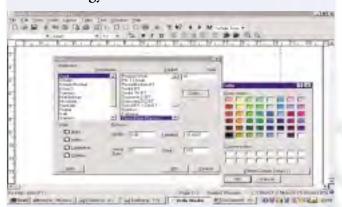
• Master Page

A master page is provided with each document to provide the user with the header, footer and similar settings. Whatever is designed on Master page appears on the pages.

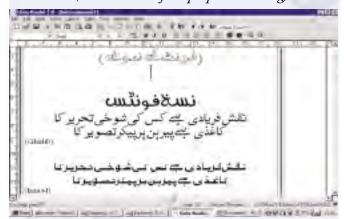
• Others

The Perso-Arabic word processor has a number of other features like Horizontal and Vertical rulers in the GUI, dynamic font settings for the Urdu and English fonts. Indents and paragraph settings, page settings, etc.

Representative Snapshot / screenshot of the Technology / Product:



(Urdu Nashir font properties dialog)



(Sample document in Urdu Nashir)

- Scalability / Portability / Expandability: All
- Readiness of Transfer of Technology (ToT): Yes
- Availability of documentation : Yes
- Testing of the Product / Technology : Under progress

- IPR / Open-source: To be decided by C-DAC Pune
- Potential beneficiaries: Common man, Offices and Educational Institutions
- User-agency tie-up: Solicited
- Name and address of the Resource Person: Shri Mahesh Kulkarni, Group Leader, GIST Group, Centre for development of advanced computing, Pune 411 007

E-mail: mdk@cdacindia.com Web: http://pune.cdacindia.com

5.3.5

- Name of the Technology: Malayalam Text Editor
- Nature of Technology: Knowledge Tool
- Level: Product
- Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product : A basic Malayalam Text Editor "STYLEPAD" is developed. It incorporates all the facilities available in Notepad together with provision to save Malayalam documents in ISCII format and read ISCII files. The editor incorporates a facility to display the standard words against some non standard words coming in the text file. This checking can be on line or off line. It also has a provision to alert the user when the length of the sentence entered goes beyond a specified length (word count). A numeric to word conversion algorithm is also included in the package,

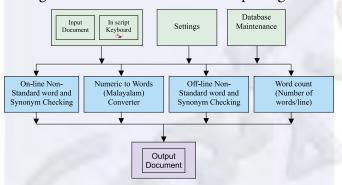
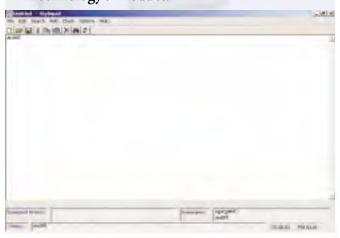


Figure 5.1 Block diagram of Stylepad

Representative Snapshot / screenshot of the Technology / Product:





- Portability/Expandability/scalability: Addition of Unicode keyboard driver and Unicode font will make the package UNICODE compatible.
- Readiness of Transfer of Technology (ToT): The Product is ready for Transfer of Technology.
- Availability of documentation: Design /TOT documents available.
- Testing of the Product/Technology: Third party testing not done.
- Potential beneficiaries: T.V Channels and publication industry. The package is presently used by the T.V Channel, Jeevan T.V.
- User-agency tie-up: JEEVAN T.V
- Name of the resource person:

R. Ravindra Kumar Additional Director & Coordinator RCILTS-Malayalam, C-DAC, Thiruvananthapuram.

Email: ravi@erdcitvm.org.

Web: http://www.malayalamresourcecentre.org



5.3.6

- Name of the Technology: Tamil Word Processor (Chathurangam)
- Nature of the Technology: Knowledge tool
- Level: (Product/Technology/sub-systems): Product
- Technical Description :
 - Description

Chathurangam is one of the tools in Tamil Office suite. It consists of four modules viz. Workbook module, Worksheet module, Expression Module and Chart Module. It handles the Tamil data representation on the Spreadsheet on different types of Charts.

Workbook module This module allows the user to save and print the workbooks. Each workbook has three work sheets. More number of worksheets can be added. The worksheets can be deleted and renamed.

Worksheet module This module enables the user to do the edit operations on the worksheet like cut, copy, paste, find and replace. The rows or columns can be inserted / deleted. The background color or the fore color of the cell, the font name, size and style for the text can be modified. The file properties like file location, size and created, modified and accessed dates, attributes of cell, text and sheet are handled. This module also provides a popup menu to access frequently used operations. The list of recent files accessed can be displayed.

Expression Module This module evaluates the expression assigned to a cell, which can be given in terms of other cell numbers or numerals. Sum, Count, Maximum, Minimum and Average are the mathematical functions supported in expression module.

Chart Module The charts are visually appealing and make it easy for users to see comparisons, patterns, and trends in data. This module

represents the data in the form of pie, horizontal and vertical bar charts. It is saved as jpeg(.jpg) file format. It provides a facility for the user to modify the title and scale of the charts.

• User Interface User Interface is developed in Java. The user interface has 'File', 'Tools', 'Calculations', 'Insert', 'Chart' and 'Help' menus. The file operations are given under the file menu. The tools menu helps the user to edit the data in the cells. The calculation menu provides the facility to do the mathematical functions. Insert menu helps to insert row, column or data. Chart menu helps to view the data in pie, horizontal and vertical charts.

Specification

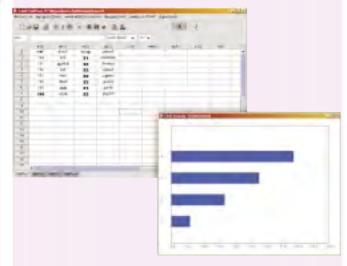
Chaturangam was developed using Java under Windows2000.

It requires J2RE 1.3.0 of above.

It requires any Tamil Keyboard driver to type Tamil.

At present, the required Font encoding scheme is TAB. Code Converter for other font encoding schemes can be provided.

Representative Snapshots



- Scalability/Expandability: As the package is modular in nature, it can be expanded to a fullfeatured Spreadsheet software.
- Portability : Yes.

- Readiness of the Transfer of Technology: Yes
- Availability of documentation: User and

Technical manuals are available.

- **Testing of the product :** Features have been tested by many developers and feedback from the users is implemented.
- ▶ IPR/Open source : Open source
- Potential beneficiaries : Public, Companies and Government offices
- User agency tie-up : Solicited
- Name and address of the Resource Center
 Dr. T.V. Geetha & Dr. Ranjani Parthasarathi,
 Resource Centre for Indian Languages Technology
 Solutions Tamil,
 School of Computer Science and Engineering,
 Anna University, Chennai 600025, Tamil Nadu.
 E-mail: rp@annauniv.edu

Web: http://ns.annauniv.edu

5.3.7

- Name of the Technology: Arangam (A Presentation tool in Tamil)
- Nature of Technology: Knowledge tool
- **Level**: Product
- Technical Description of the Technology
 - Description

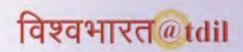
Arangam is 'A Presentation Tool for Tamil' with bilingual (English/Tamil) interface. It is one of the packages in 'Tamil Office Suite'. It helps in combining text, pictures and images and developing an appealing presentation to the audience. This software is written under the Java 2 platform to maintain compatibility among various Operating Systems. It uses extensively the Swing library that is part of the standard JFC (Java Foundation Classes).

Arangam, the main class makes all initializations to the GUI. In particular, it constructs the menu bar (with JMenuBar), the toolbar (with JToolbar) and the working area (with JPanel). All actions are activated using the event model of Swing. It also initialise a new instance of the Slideshow class. This class holds all information about the current slideshow like the size of the window (in pixels), the current view mode (one slide or slide show), etc.

The application permits to create and save files in the file system in the local server/computer or in the network. Once a file is created, slides can be added or deleted. Slides can be edited with common operation like adding text boxes, pictures and general shapes. The Slides are viewed as:

- Edit Mode (Slide)
- Preview Mode (Slide Sorter)
- Presentation Mode (Slide Show)

The edit mode is the common one and allows working in one selected slide. Preview mode gives a preview of number of slides (there are different possibilities 2x2, 3x3 or 4x4 slide in one page).



Presentation mode is the presentation of the slides one by one in the screen in big sizes.

A slideshow is an extended JPanel that can hold other JPanels. A member variable of the slideshow is an array of Slides. When adding a slide to the slideshow a new (extended) JPanel is added to the array.

The over-all container is the class Slideshow that contains a java vector of Class Slide that contain the Components - Image (JLabel), Textbox (JTextPane) and Shape (JLabel). The ConnectComponent is a general component that corresponds to the interface ConnectComponent. All the Objects used are java built in containers. Each of the data structures have member variables to hold their state

Each ConnectComponent as Image, Text and Shape responds to mouse clicks, dragging, moving and releasing and activates the corresponding functions of the slide it is in. The toolbars and menu catch the user events and activate the corresponding functions in the Class Slideshow that activate the function in the requested Slide. There is a top down design of the event catching - Component -> Slide -> Slideshow and the other direction Slideshow -> Slide-> Component.

The event handler object directs all the events generated from the menu bar, the toolbar, operation given from the Graphical user interface, operations from the keyboard.

- Block Diagram: Not applicable
- User Interface:

This will describe the Graphical User Interface for the Arangam application. A menu, an "Action" and a "View mode" Toolbar compose the GUI. The Upper Toolbar manages common file operations (New, Open, Save a File), editing operations (Cut, Copy and Paste), normal Presentation operation like adding a slide, a picture, a text box or a shape and format operation. There are buttons in tool bar to perform common functions such as changing the font size and font type in text editor. There is also a

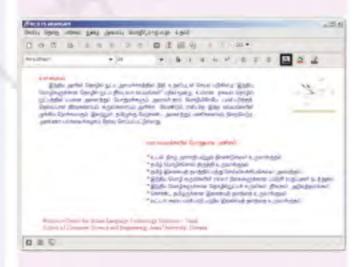
slide number indicator window that shows the current slide number; Using this, it is possible to jump automatically to any slide. Near the windows there are two arrows that permits to navigate through the slides. The lower toolbar has shortcuts for operation in view bar. Three buttons for the three mode of work compose the Lower Toolbar (Edit, Preview and Presentation).

• Specifications of Technology: Software components

The system was developed using Java on Windows 2000 platform.

Minimum requirement for the tool

- 1. J2re1.4.0
- 2. Any Tamil Keyboard driver to type in Tamil
- 3. Operating System: Windows/Linux
- Representative Snapshot:



- Scalability/Expandability: The software can be expanded to a full-featured Presentation tool.
- Portablility: Yes.
- Readiness of Transfer of Technology: Yes
- Availability of documentation: User and Technical manuals are available.
- Testing of the Product / Technology: Features tested by many users.



- Open-source: Yes.
- Potential beneficiaries: Public

Web: http://ns.annauniv.edu

- User-agency tie-up: No
- Name and address of the Resource Person:
 Dr. T.V. Geetha and Dr. Ranjani Parthasarathi
 RCILTS Tamil,
 Anna University,
 Chennai.
 E-mail: rp@annauniv.edu

5.3.8

- Name of the Technology : Palagai (Palagai Tamil Word Processor)
- Nature of Technology : Knowledge Tool (Language Technology Product)
- Level: (Product / Technology / Sub-system) : Product
 - Technical Description

Palagai provides basic facilities for word processing both in Tamil and English. There are two versions of the word processor available. One with a spellchecker and a grammar-checker, and another without these.

The word processor does not have any special file format. Files of both rich text format and text-only formats can be created or edited. Files of type ".doc" and ".pdf" can be opened as a text files by converting them into text-only format. HTML files can be viewed with this word processor. E-mails can be sent using this word processor if the current system is connected with the SMTP server. Email attachments are also handled.

• User Interface

Palagai supports both Tamil and English. All menus, menu items and etc are in both Tamil and English. The user can select the language of the user interface by a toggle key. The user can process their document with the appropriate menu and menu item from the menu bar. Freuently used operations are provided in the tool bar. Format bar contains the button to do the format operations.

• Specification

Palagai was developed using Java. It requires J2RE 1.3 or above to run. It requires any keyboard driver to type Tamil. At present the required font encoding scheme is TAB. Code converters for other font encoding schemes can be provided.

- Representative Snapshot / screenshot of the Technology / Product:
- A Scalability: Yes



- Portability: Yes
- Readiness of Transfer of Technology (ToT): Yes
- Availability of documentation: User and technical manuals are available.
- Testing of the Product / Technology: Features were tested by many users
- ▶ IPR / Open-source : Open-source
- Potential beneficiaries: Public and Government organizations.
- User-agency tie-up : No

Web: http://ns.annauniv.edu

Name and address of the Resource Person:
Dr. T. V. Geetha, Dr. Ranjani parthasarathi,
Resource Centre for Indian Languages Technology
Solutions – Tamil,
School of Computer Science and Engineering,
Anna University, Chennai - 25
E-mail: rp@annuniv.edu

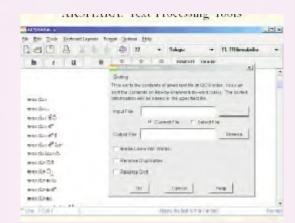
5.3.9

- Name of the Technology: AKSHARA: Advanced Multi-Lingual Text Processor
- Nature of the Technology: Knowledge tool
- Level: Product is ready for technology transfer
- Technical Description

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.

AKSHARA is an advanced text processing tool-dictionaries, morphological analyzers, spell checkers, OCR systems, TTS systems, text processing tools including searching, sorting etc. are part of AKSHARA. Several text processing tools, Telugu spell checker and Telugu TTS have already been integrated. We would also be happy to integrate any of the other dictionaries, spell checkers, TTS etc. that other centres may have developed. Full support for Regular Expressions and Finite State Machines is being integrated.

Representative Screenshot



- Operating System used: LINUX, Windows 98, 2000, XP
- Scalability / Portability / Expandability : Available
- Availability of documentation: Yes



- ▶ Testing of the Product / Technology : Testing has been done
- ▶ IPR / Open Source : IPR lies with University of Hyderabad and Dept of Information Technology
- **Potential beneficiaries :** Users of telugu softwares public and government organisations.
- Name and address of the Resource Person:
 Dr. K. Narayanamurthy
 Department of CIS, University of Hyderabad
 Hyderabad-500046, A.P. India

5.4 Transliteration Tools

5.4.1

- Name of the Technology: Transliteration from Gurmukhi To Shahmukhi
- Nature of Technology: Knowledge Tool (Software for Punjabi to Urdu Transliteration)
- Level: (Product / Technology / Sub-system)
 Product developed in collaboration with CDAC, Pune
- Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product : Punjabi Language is used in both parts of Punjab in India and Pakistan. In East Punjab (India) Punjabi is written in Gurmukhi script. This script was created by Guru Angad Dev Ji. This is written from left to right. In West Punjab (Pakistan) Punjabi is written in Shahmukhi script. This is written from right to left like Urdu and Persian. Both Shahmukhi and Gurmukhi have been in use simultaneously for writing in Punjabi Language. A transliteration program was required to break the barrier between the Punjabi Language written in these two scripts which can convert Gurmukhi Text to Shahmukhi and vice versa. In this direction the Punjabi resource Centre has collaborated with the Urdu resource Centre established at CDAC, Pune to develop a computer program which will automatically convert the Gurmukhi text to Shahmukhi. Using this software, a collection of short stories by Mr. K.S. Duggal in Gurmukhi has been converted to Shahmukhi.

Platform: IBM Compatible PC with 486 or Pentium CPU.

Operating System: Windows 95,98,2000, XP.

Programming Interface: Delphi

Representative Snapshot / screenshot of the Technology / Product: Screenshot of Gurmukhi Input Text : Screenshot of Shahmukhi Output Text :