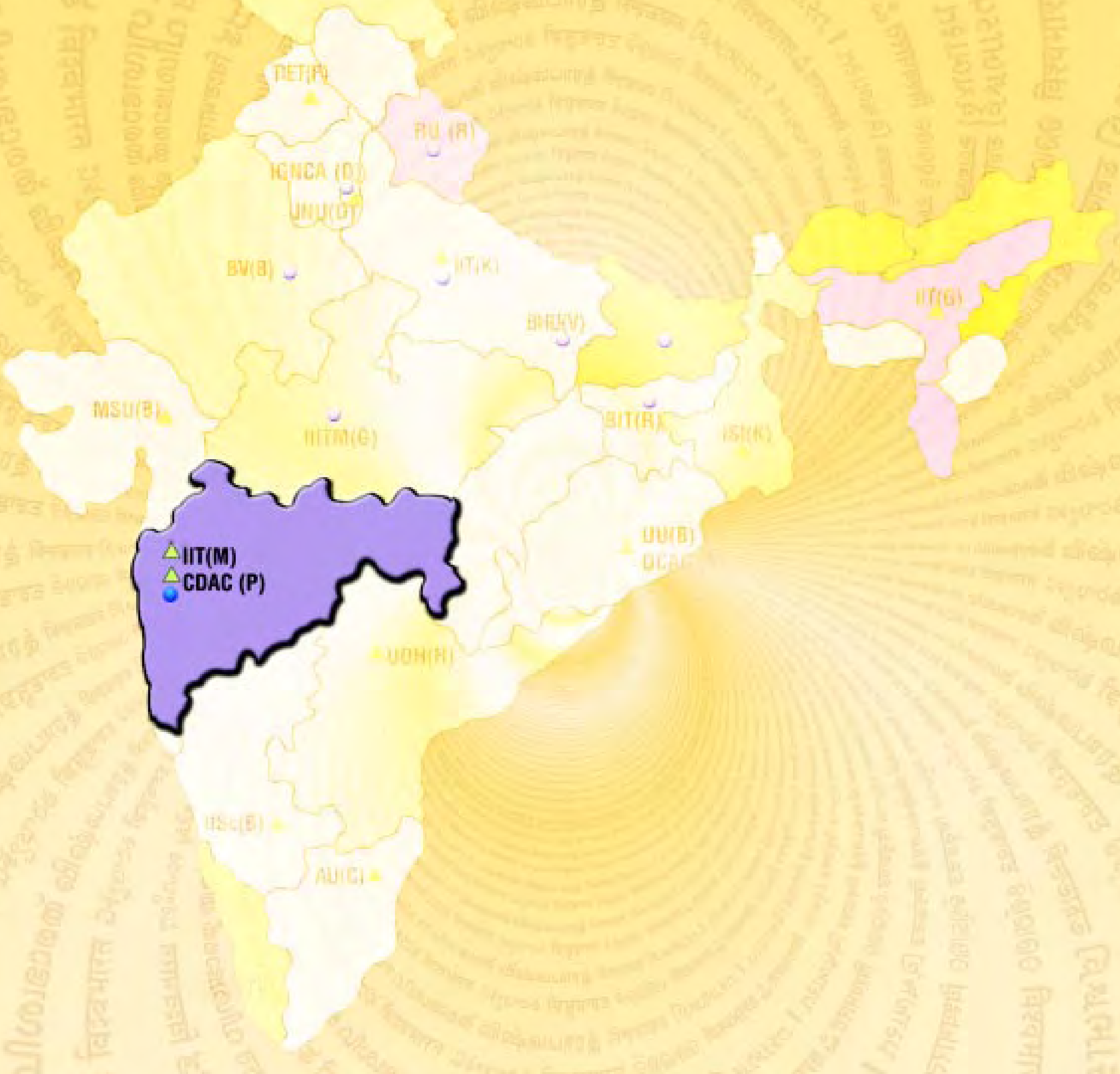
Resource Centre For
Indian Language Technology Solutions – Urdu, Sindhi, Kashmiri
C-DAC, Pune

Achievements



Centre for Development of Advanced Computing
Pune University Campus, Ganesh Khind, Pune-411007
Tel. : 00-91-20-5694000-2 E-mail : mdk@cdacindia.com
Website : http://.parc.cdacindia.com

RCILTS-Urdu, Sindhi & Kashmiri Centre for Development of Advanced Computing, Gist, Pune

Introduction

"Digital unite & Knowledge for all the vision of TDIL"

MC&IT wishes to make a difference to the quality of life of people of India through its TDIL mission, and is investing significant amount of resources towards fulfilling this wish. One such effort was the establishment of "Resource centers" for design, development & deployment of tools & technologies for Indian languages.

"Dissolving the language barrier the vision of GIST".

C-DAC has pioneered the Graphics and Intelligence based Script Technology (GIST) which facilitates the use of Indian languages in IT. GIST is geared for the Internet enabled world where all activities are gradually going online. In its endeavor to stay abreast with technologies worldwide, GIST has been adopting the latest concepts to be able to stay tuned with the changing IT scenario. C-DAC, GIST through its continuous R&D efforts has to its credit many innovative products & continues to be leaders in the language technology field.

MC&IT, through its TDIL Programme has awarded C-DAC, GIST with the project "Resource centre" for development of tools & technologies for Perso-Aarbic languages.

Overall Goal of the Resource Centre

To empower people of India through the use of Information Technology solutions in Indian languages.

Project Purpose

To improve the quality of life of people of India by enabling them to use information technology in Indian languages.

To develop new products and services for processing information in Indian languages.

To conduct research in computer processing of Indian languages.

Broad areas of work under the Perso-Arabic Resource center

GIST workshop for usage of GIST tools for organizations concerned with development & deployment of Indian language processing systems.

Creation of a network of persons and organizations working on IT applications in Perso-Arabic(PA) languages namely Urdu, Sindhi & Kashmiri.

Development of Language tools & applications in PA scripts.

Conduction of workshop to discuss on standardization of PA scripts data & font representations.

Deployment of the developed technology, products, & services through a country-wide network.

Services and Knowledge Bases

- Creation of a web site for language tools, solutions and knowledge available in PA language.
- Imparting training for the usage of PA.
- language applications on IT.
- Establishments of Translation & Subtitling services in PA languages.

Products

- High quality PA fonts.
- Dictionary, Spell Checker tools in PA languages.
- Word processor package on PA languages.
- Transliteration tool between Hindi & Urdu.
- Prototype of Pocket translator with Urdu support.
- Product for Subtitling, Character Generator, Teleprompter, DVD in PA scripts.

Activities Undertaken Under The Project

GIST workshop was conducted at C-DAC Pune from 17th to 22nd July 2000, for usage of GIST series of Indian Language Technologies, tools and products for Resource Centers concerned with development and deployment of Indian Language solutions. A team of 23 members from 8 Resource centres have attended.

1. ER & DC, Trivendrum
2. Jawaharlal Nehru University, New Delhi
3. Indian Institute Of Technology, Kanpur
4. Orrisa Computer Application Center, Bhubneswar
5. Thapar Institute Of Engg. And Technology, Punjab
6. Anna University, Chennai
7. M. S. University, Baroda
8. Indian Statistical Institute, Calcutta

Covering over five days and total of 12 interactive sessions, the subjects covered were

1. Building dictionary & Designing spell checker
2. Keyboard driver
3. GIST software development kit
4. Multilingual web tools
5. Script code, Font storage & Design.

During the workshop training kits were distributed which contained work manuals on following topics;

1. Designing Indian Language Spell Checkers,
2. Myths about Standards for Indian Languages,
3. Enabling Indian Languages on Web,
4. ISCII- Foundations & Future,
5. GIST Software Development Kit,
6. Using GIST SDK in VB,
7. Internet First Steps: Upgrading an Existing ActiveX Control.
8. Inscript Keyboard layout,
9. Typing in Phonetic Keyboard.

Two CDs on GIST s/w and brochures on various language products & fonts type face catalogue were distributed:

1. GIST based S/W Developer CD.
2. GIST Products & Channel Information.

A visit was organized to various departments of C-DAC, GIST Laboratory, National Multimedia Resource Centre, and National PARAM Super Computing facility.

Evolving Storage, Font & Inputting Standards

The main task before C-DAC was to evolve standards for storage, fonts & inputting for the Perso-Arabic languages. Development of USCII (Urdu standard for Information Interchange) was carried out well before awarding of the project. GIST Terminals designed & developed by C-DAC are based on these standards. The support for the GIST terminals was limited to only the Naskh implementation of the fonts.

Accordingly with lots of efforts and valuable inputs from experts all over, C-DAC drafted PASCII standards for storage, keyboard standards & font standards. These proposed standards are published in the TDIL magazine – Vishwa Bharat for comments / suggestions from experts all over India.

1. The PASCII standard - The Perso-Arabic Standard for Information Interchange

There are some peculiarities that can be seen in Perso-Arabic scripts:

These scripts join letters with each other, and therefore letters have different forms as per their position in a ligature (the positions are beginning, middle, ending).

Some shapes do not have middle shape i.e. they do not join at both the ends. For example alif, vao, daal, etc. Every letter also has a standalone form

1.1 Characteristics of Proposed Standard

Its an 8-bit standard. Supports letters for Urdu, Arabic, Sindhi, Kashmiri.

Defines PA alphabets in the upper ASCII leaving lower ASCII free for English. Bilingual support possible.

Defines numerals other than ASCII numbers (48 to 57).(This may help supporting both Arabic Numerals 0-9 and language specific numerals).

Maintains the order of alphabets for perso-arabic languages. Khate-kasheed is given lower value than alphabets. This will sort the words correctly even though they have khate-kasheed in between.

Alphabets for different languages are placed in their ascending order. Letters like “bhey” are not provided for URDU but kept for languages like SINDHI. The URDU may make use of “be” and “choTi-hai” for that.

Minimal erabs are provided. Tanveen ,for example do-zabar, can be formed with the help of two consecutive zabar.

Superscripts

Place for superscripts like khaRa-alif is provided.

Place for superscripts for ARABIC is provided.

Place for superscripts like “re-ze”, “ain”, etc. is provided.

Numerals are placed after erabs and super scripts. (This is provided only to support display for language specific numerals and standard numerals i.e. the ASCII numerals are available).

Only a few control characters are defined.

1.2 Standardization of Urdu Keyboard

Overview

There is no standard keyboard available for URDU language. There are vendor specific keyboards available which all differ in their layouts. There are lot of keyboards designed by different companies who provide Perso-Arabic support in their applications. Unfortunately, every company has its own Keyboard layout that entirely differs from the other’s Keyboard Layouts. For example Microsoft has its own Keyboard layouts for Perso-Arabic languages.

Drawbacks of Available Keyboards (in view of URDU language)

No standard keyboard available.

All vendor specific keyboards differ in their layout.

Most of the keyboards have been designed for Arabic.

Although Arabic and Urdu have similar script, they are both different languages, and hence a Keyboard designed for Arabic may not be that useful for Urdu language. CDAC has taken this care of languages to design the keyboard for URDU, SINDHI & KASHMIRI and tried to come up with an optimum solution.

(Details of these standards can be found in TDIL Vishwabharat magazine.)

1.3 Standardisation of Fonts

Characterstics of Perso-Arabic languages:

Urdu has traditionally been written in the Nastaleeq script. Although the script employs the basic letters of the language, the rendering of these letters in a word is extremely complex. The reason for this complexity is that Urdu text has traditionally been composed through calligraphy, a medium whose precepts are based on the aesthetic sense of the calligrapher rather than on any formula. So great is the variation in calligraphy that many times it is difficult to recognize the letters in a constituent word. This is because, in their calligraphed form, the individual letters partially or completely fuse into each other thereby losing their identity. A degree of fusion is purposely introduced to make the resulting fused glyph visually appealing.

Another characteristic of the Urdu is the existence of diacritics. Diacritics, although sparingly used, help in the proper pronunciation of the constituent word. The diacritics appear above or below a character to define a vowel or emphasize a particular sound. They are essential for removal of ambiguities, natural language processing and speech synthesis.

A Nastaliq style font – designed, to be used for printing and display. The font is based on traditional Nastaliq style written in India

Arabic Font – A high quality Arabic Font is designed to compliment religious text in Urdu.

Considerations for Digital Font Design

Following is a group-wise list of characters that were taken into consideration while designing fonts for the Perso-Arabic languages.

1. Alphabet.

2. Numerals.

3. Special Characters.

4. Diacritics.

5. Religious and linguistic Symbols.

6. Control characters.

The 16-bit Naskh & Nastaliq Font

The Fonts developed by CDAC are 16-bit and are also defined in the User Area of the Unicode range. The ASCII range is not used and can be used for different purposes (can be used to English support for example).

Includes

- all the basic shapes.
- all the starting shapes and variations.
- all the middle shapes and variations.
- all the ending shapes and variations.
- levels for erabs (short vowels).
- complete ligatures.
- Beginning ligatures.
- Middle ligatures.
- Ending ligatures.
- missing character glyph.

Glyph Standards decided for FONT Development

Urdu & Kashmiri- Naskh & Nastaliq script

Sindhi - Naskh script.

- Basic letters (Vowels, Consonants).
- Full shape letters.
- Beginning.
- Medial.
- Final.
- Ligatures.
- Graphic components.
- Numerals, Signs and Symbols, etc.

2. Script Core Technology & Rule Engine

Because of the complex nature of the Nastaliq script which is written right to left and top to bottom diagonally, it was required to make certain databases, and use them along with the rule engine, to make display possible. Following are some tools which were developed to create these databases.

2.1 Glyph Property Editor

In Urdu and all other Perso Arabic languages each character has many different shapes depending upon the position of that character, 1) the stand alone 2) the starting position 3) the middle position 4) the ending position etc. So according to the position of the character and the letters with it its joining that particular shape is to be displayed. This is an application that allows to store the data and the compositional information for various shapes in a font file for proper alignment of characters for display. (Urdu requires a larger number of glyphs than other scripts) This takes care of the horizontal and vertical movement for Nashtaliq script.

2.2 Rule Engine

A module that containing rules for each shape. The rules indicate how and which shape will join with other shape or shapes based on the succeeding and preceding shapes that occur in Nastaliq text composition. This facilitates more accurate and faster display of the shape composition in the Nashtaliq text composition.

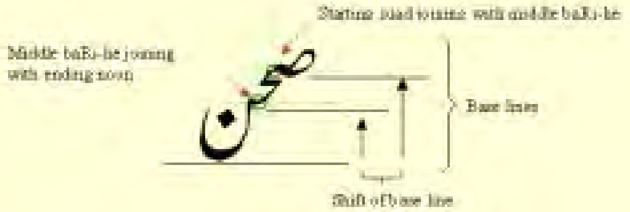
2.3 Testing version of Urdu Editor

A simple text editor is ready to test the Rule Engine and check the various rules formed for displaying the glyphs. This editor allows to write the text in left to right, and handles all keyboard mappings from ASCII to USCII, and Urdu text is displayed.

2.4 Font Glyph Editor

In Nastaleeq script (for Urdu), the shapes join from right to left, and within a ligature they also shift vertically. This forms a ligature. So in a ligature, shapes join right to left and top to bottom. The vertical positioning information cannot be put in the font itself. Hence it was decided to make a utility, which will let one set this information in separate file.

Example of Nastaleeq joints: ("sehan" => suad baRi-he noon):


Starting suad joining with middle baRi-he
Middle baRi-he joining with ending noon
Base lines
Shift of base line

Purpose

To select a glyph from the given font and character, and define on it following information:

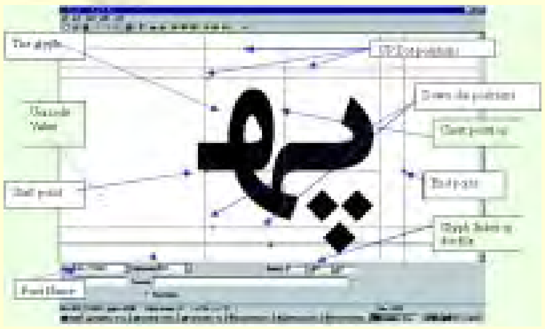
1. Starting point (SP): the point where a glyph ends. This is generally on baseline or zero.
2. Ending point (EP): the point where a glyph joins with another glyph on its starting point (SP of another glyph). This point is less than or equal to the glyph width.
3. Dot points (UP and DP): dot points where a dot or diacritic mark will be positioned on the glyph.

2.5 The Font Editor Utility

The Font Editor utility lets us place the dot information for each glyph. The file thus generated is used for retrieving dot information at run time. We'll call it a dot file.

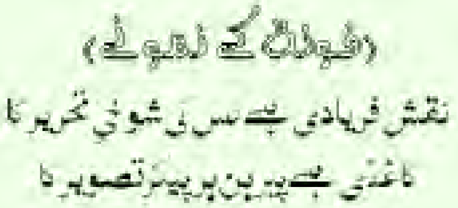
The following figure depicts a glyph with dot and caret information. As shown in the figure, the utility helps one define following for a glyph:

1. Font Name.
2. Unicode value of the glyph.
3. Glyph Index in dot file (this is the unique id of the glyph).
4. Start and End points.
5. Dot positions.
6. Caret position.



Fonts : following are samples of fonts developed

(Ghalib)

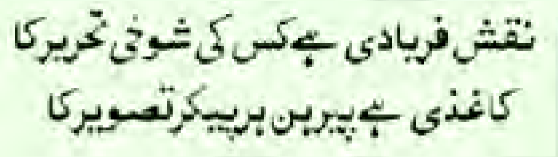


«غالب کے نمونے»

نقش فریادی ہے کس کی شوخی تحریر کا

کاغذی ہے پیرہن ہر پیکر تصویر کا

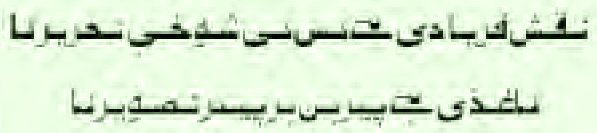
(Ghalib Bold)



نقش فریادی ہے کس کی شوخی تحریر کا

کاغذی ہے پیرہن ہر پیکر تصویر کا

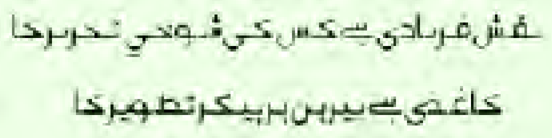
(Imaad)



نقش فریادی ہے کس کی شوخی تحریر کا

کاغذی ہے پیرہن ہر پیکر تصویر کا

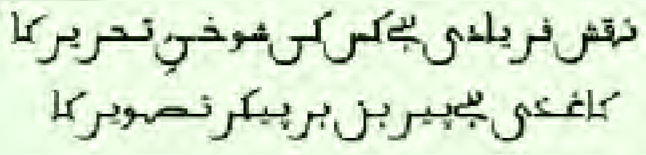
(Kamran)



نقش فریادی ہے کس کی شوخی تحریر کا

کاغذی ہے پیرہن ہر پیکر تصویر کا

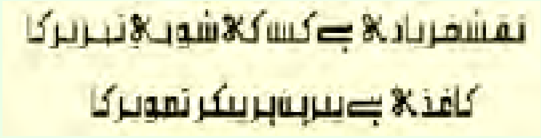
(Maghribi)



نقش فریادی ہے کس کی شوخی تحریر کا

کاغذی ہے پیرہن ہر پیکر تصویر کا

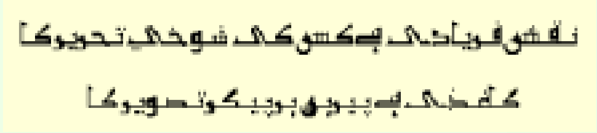
(Makhdoum)



نقش فریادی ہے کس کی شوخی تحریر کا

کاغذی ہے پیرہن ہر پیکر تصویر کا

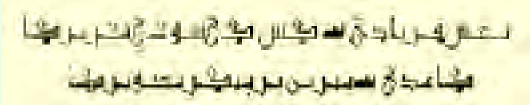
(Murasilat)



نقش فریادی ہے کس کی شوخی تحریر کا

کاغذی ہے پیرہن ہر پیکر تصویر کا

(Riyadh)



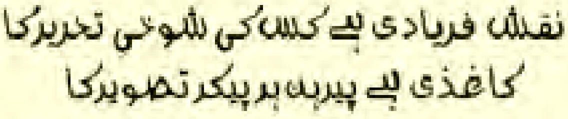
نقش فریادی ہے کس کی شوخی تحریر کا

کاغذی ہے پیرہن ہر پیکر تصویر کا

(Rubai)

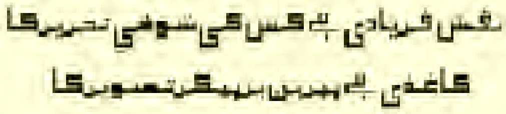
نقش فریادی ہے کس کی شوخیِ تحریر کا
کاغذی ہے پیرہن ہر پیکرِ تصویر کا

(Safi)



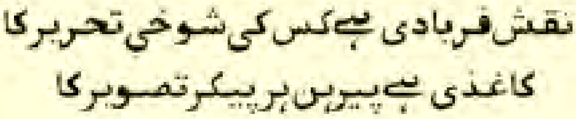
نقش فریادی ہے کس کی شوخی تحریر کا
کاغذی ہے پیرہن ہر پیکر تصویر کا

(Saleem)



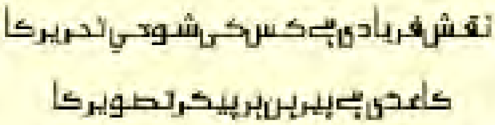
نقش فریادی ہے کس کی شوخی تحریر کا
کاغذی ہے پیرہن ہر پیکر تصویر کا

(Shibli)



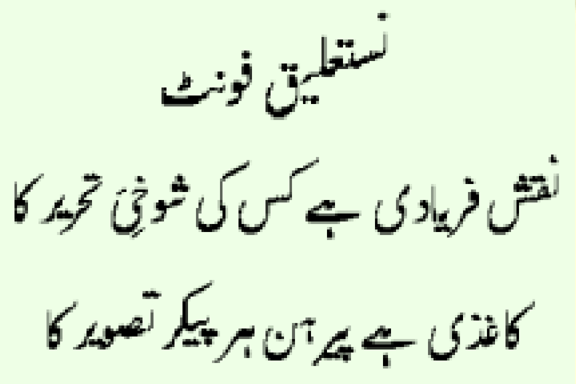
نقش فریادی ہے کس کی شوخی تحریر کا
کاغذی ہے پیرہن ہر پیکر تصویر کا

(Shuja)



نقش فریادی ہے کس کی شوخی تحریر کا
کاغذی ہے پیرہن ہر پیکر تصویر کا

(Abid)



نستعلیق فونٹ

نقش فریادی ہے کس کی شوخیِ تحریر کا
کاغذی ہے پیرہن ہر پیکر تصویر کا

3. Devanagari/Gurmukhi to Urdu Transliteration "UTRANS "

A computer based transliteration allows a user to convert a given text (say, in storage format - ISCII) of language 'A' to text in another language 'B' (in some storage format, say ASCII) on the basis of the phonetic rules that govern languages A and B.

A software (automatic) tool providing transliteration from one language to another is an effective tool which reduces the amount of efforts required to create large databases of names in multiple languages.

Once a database in one language is created, the tool can be used to generate the same data for other languages. This results in considerable savings in time and money.

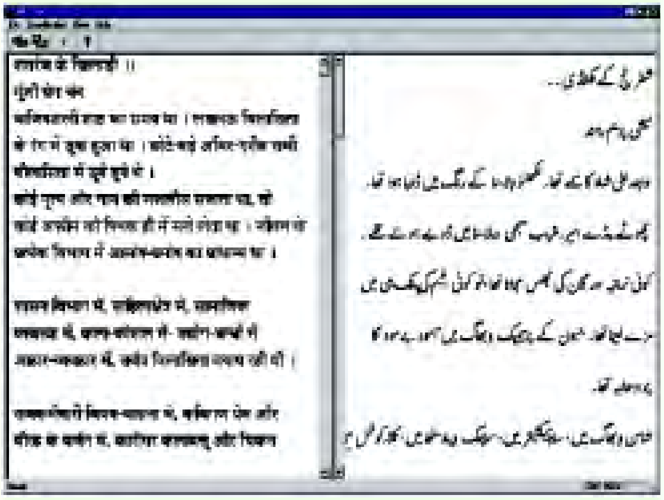
It is a useful tool for applications which require to convert database of names, addresses, phone numbers or electoral rolls from one language to another.

This can also be useful for converting entire databases from one language to another without re-entering the data. This utility can convert data in English from a database directly to ISCII.

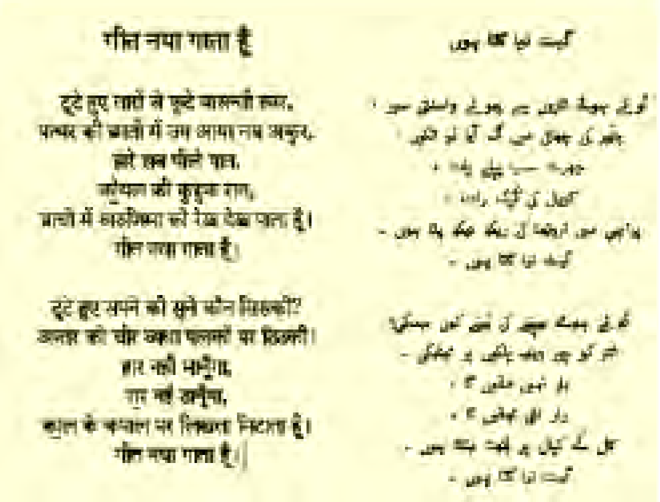
This will facilitate automated transliteration across databases.

A software tool for transliterating text from Devanagari & Gurumukhi (ISCII) to Urdu has been developed. This tool is also available as a separate module, and also been implemented in the Text Editor Test Application, which is developed.

3.1 Following is a sample of Hindi text Transliterated into Urdu (UTRANS)



Using this automated tool – UTRANS, we have transliterated the book **"Meri Ekyavan Kavitayen" written by our Hon. Prime Minister Shri. Atal Bihari Vajpayee.**



गीत नया गाता हूं

टूटे हुए तारों से फूटे बासन्ती स्वर,
पत्थर की छाती में उग आया नव अंकुर,
झरे सब पीले पात,
कोयल की कुहुक रात,
प्राची में अरुणिमा की रेख देख पाता हूं।
गीत नया गाता हूं।

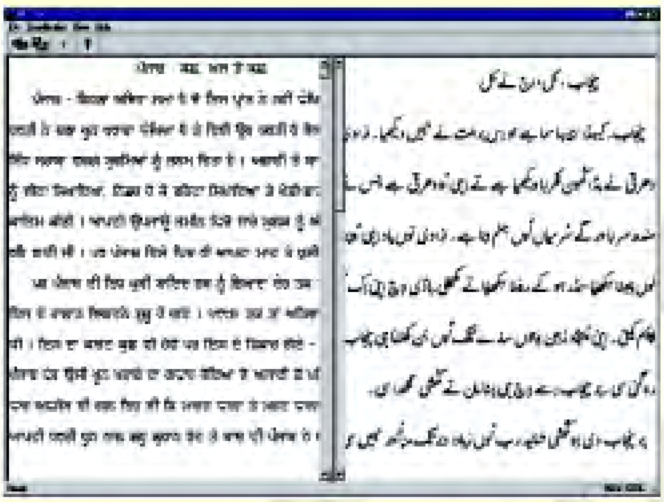
टूटे हुए सपने की सुने कौन सिसकी?
अन्तर की चीर व्यथा पलकों पर ठिठकी।
हार नहीं मानूंगा,
रार नहीं ठानूंगा,
काल के कपाल पर लिखता मिटाता हूं।
गीत नया गाता हूं।

گیت نیا گاتا ہوں

ٹوٹے ہوئے تاروں سے پھوٹے واسنتی سور،
پتھر کی چھاتی میں اُگ آیا نو انکر،
جھڑے سب پیلے پات،
کوئل کی کہک رات،
پراچی میں ارونیما کی ریکھ دیکھ پاتا ہوں۔
گیت نیا گاتا ہوں۔

ٹوٹے ہوئے سپنے کی سنے کون سسکی
انتر کی چیر ویتھا پلکوں پر ٹھٹھکی۔
ہار نہیں مانوں گا،
رار نہیں ٹھانوں گا،
کال کے کپال پر لکھتا مٹاتا ہوں۔
گیت نیا گاتا ہوں۔

3.2 Following is a sample of Punjabi text Transliterated into Urdu (UTRANS)



4. Dictionary Development Tool

Natural language processing tools such as lookup dictionaries, thesaurus, spellchecker, grammar checkers, part of speech taggers, machine aided translations require language specific dictionaries. Several other domains such as speech processing, expert systems require backend language dictionaries to be able to process and generate valid outputs. Each of these applications require dictionaries to be created to cater for the specific requirements.

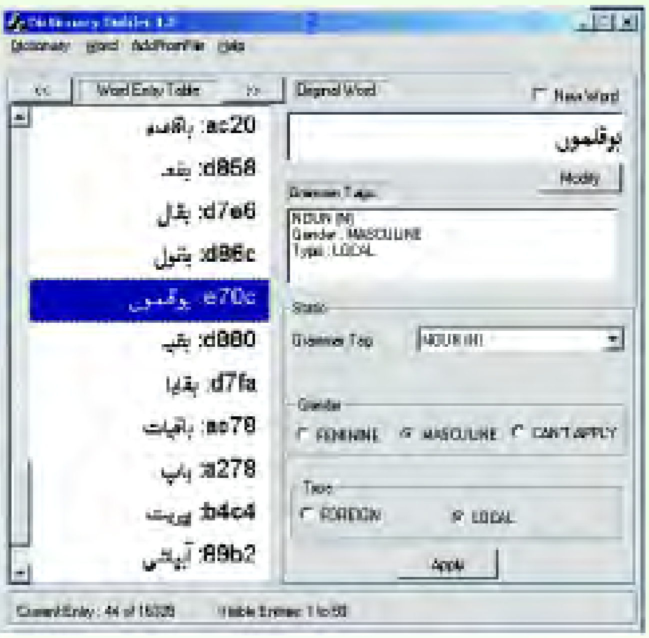
The Generic Dictionary Development Tool will allow users to create dictionaries specific to application domains. As part of this development the format of a generic dictionary should standardized. The tool will provide facilities for obtaining views of dictionaries that are subsets of the generic dictionary.

This tool, though lot of refinement is required, is in the form of a standalone desktop utility which allows the user to create dictionaries containing words for Perso-Arabic languages. The user is provided with an interface to define inflection and root word rules, add suffix, prefix, grammar tags, domain for a given word entry.

In order to build spellchecker & domain specific dictionaries following dictionary creation work was under taken and completed

1. Official Hindi dictionary - Hindi 50,000 words, equivalent 50,000
2. Urdu to English : urdu words 17000, grammer tag 17000
3. Hindi Urdu Shabdh Kosh Hindi words 10,000, Equivalent 10,000
4. Dictionary Urdu words : 10,000 grammer tag 10,000
5. Dictionary of Urdu & Hindi names : 3200, Equivalent : 3200
6. Technical dictionary Hindi words : 1500, equiva lent : 1500
7. Muslim names Dictionary : Hindi 1500, equivalent : 1500
9. Hindi - English - Urdu Hindi words : 8000, equivalent : 8000

Following is the typical GUI of the dictionary development tool for URDU.



باقاعدہ :ac20
بقعہ :d858
بقال :d7e6
بتول :d86c
برقلموں :e70c
بقیہ :d880
بقایا :d7fa
باقیات :ac78
باپ :a278
بیویت :b4c4
آپاتی :a9b2

Original Word: برقلموں

Grammar Tags:
NOUN (M)
Gender : MASCULINE
Type : LOCAL

Grammar Tag: NOUN (M)

Gender: ○ FEMININE ● MASCULINE ○ CAN'T APPLY

Type: ○ FOREIGN ● LOCAL

Current Entry : 44 of 15335   Visible Entries: 1 to 50

5. Nashir - Wordprocessor– Publishing Made Easy

The Nashir is designed to be easy to use and create documents in Perso-Arabic languages, and at the same time powerful enough to layout complete newspapers and magazines in Urdu, Sindhi, Kashmiri, Arabic and Farsi.

Each document of the Nashir consists of a number of pages. On each page of the document, you can place items like text blocks, graphics, etc.

The Nashir is kind of an Word processor & also best suited for publishing segment. Spellchecker support is added.

Apart from the base dictionary for spellchecker, various domain specific dictionaries addition is planned.

Salient Features of Nashir

1. Support Nastaliq True Type fonts. (presently 2 fonts).

2. Supports Naskh fonts. (presently 12 fonts).

3. Fonts for Sindhi, and Kashmiri added.

4. Supports C-DAC & Phonetic Keyboards.

5. User defined keyboard support available.

6. Drawing objects provided.

7. OLE Automation supported.

Other Features

Bilingual Support

Supports Urdu, Sindhi & Kashmiri along with English.

Transliteration

Transliteration engine (uTRANS) has been implemented in the Nashir. One can insert an ".aci" (ISCII file) file into Nashir and see the transliterated version in Urdu script (Naskh or Nastaliq). Rule based transliteration developed for Hindi & Punjabi.

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.

Keyboards

Phonetic keyboard supported. A Floating key board will be supported soon.

Tool tip

Tool tips are provided for the user

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.

Kerning

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

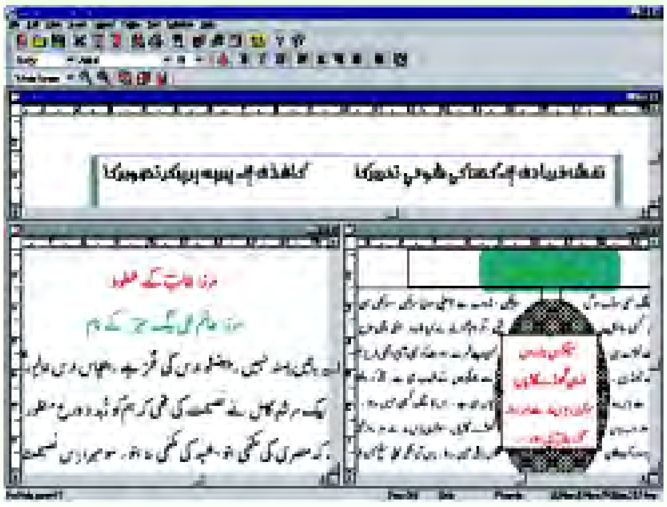
Variable Font Sizes & Colors

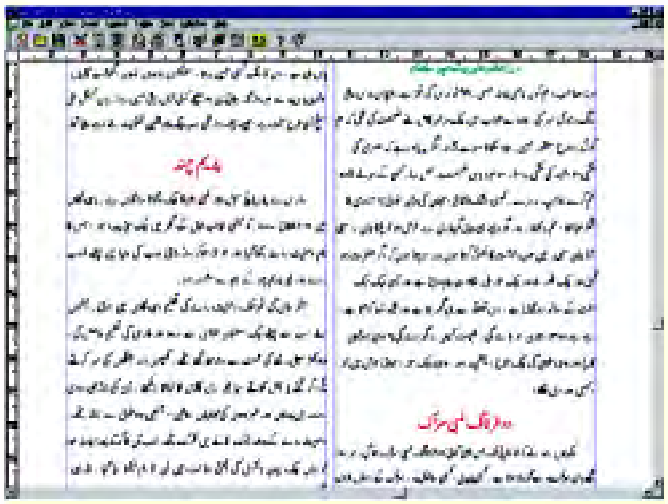
Use variable font sizes and color.

Others

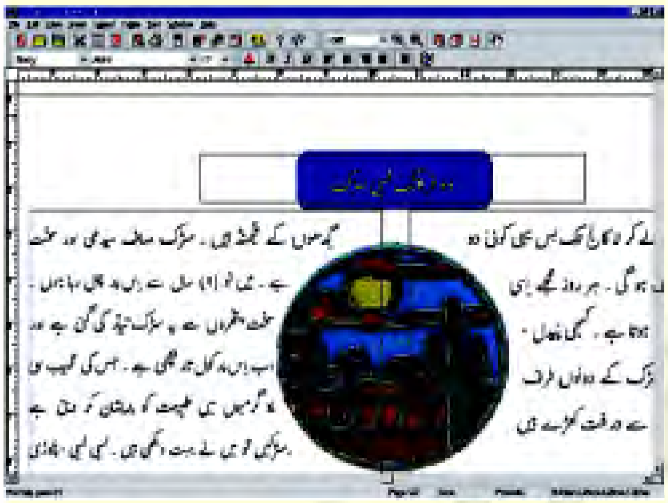
The Nashir has a number of other features like Horizontal and Vertical rulers in the GUI, dynamic font settings for the Urdu and English fonts. Indents and paragraph settings, page settings, etc.

Nashir Screen Shots





Following screen shot shows "Text wrapping around object" feature of Nashir :



Spellchecker Support in Nashir

The full version of Nashir is now equipped with the spell checker facility.

The Spellchecker program works as a word-level error detection and correction (single or multiple correction) tool. Following consideration were given while designing the spellchecker module

- Percentage of invalid words pass through it
- Average suggestion rate
- Intended suggestion is given
- Domain the spellchecker is serving
- Dictionary structure

Error Detection/Correction Approaches

1. Non-word error detection
2. Isolated word error detection
3. Context based word correction

Non-Word Error Detection

Typographic errors (typing mistake)

Cognitive errors (invalid entry by user)

Phonetic error

Real word error

1. Detection: Block invalid words to pass through it
2. Correction: Suggest alternate valid words that can pass through it

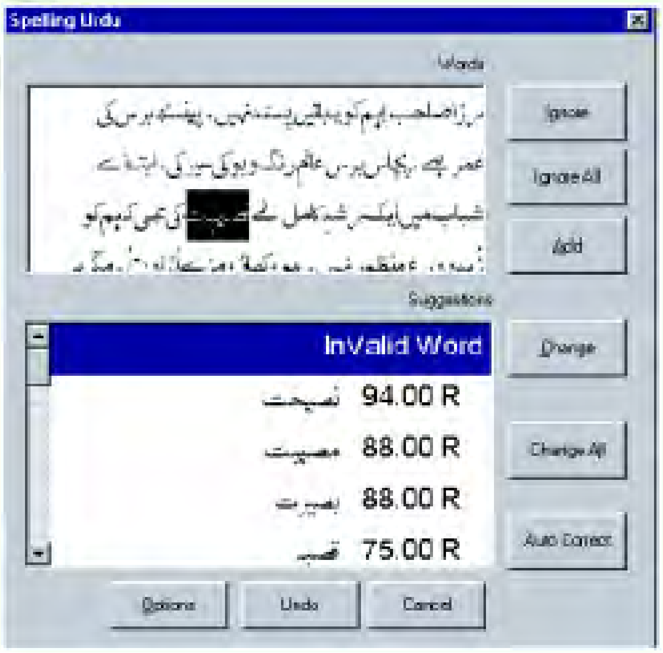
Error Detection Techniques

Lookup Dictionaries

- Structure : Hash, Tries, etc.
- Dictionary size
- Inflection, euphony, assimilation: (Morphological Analyzer)
- Domain

Bi-gram / Tri-gram approach (based on conditional probability approach)

A two dimensional matrix having conditional probability state of letters occurring next to each other. Bi-grams are searched and if a bi-gram is not in the dictionary, then the words are termed incorrect.



Spelling Urdu

InValid Word

نصیحت 94.00 R

مصیبت 88.00 R

بصیرت 88.00 R

نصیب 75.00 R

6. The Urdu SDK Controls

The GIST Urdu SDK is a software development tool for Urdu Language on MS Windows.

This is a set of software components that enables software developers to facilitate the use of Urdu scripts with MS Windows applications. Unlike other solutions, the significant feature of GIST Urdu SDK is that it enables the MS Windows application to process PASCII data directly.

The GIST SDK uses ActiveX Technology from Microsoft and provides a seamless, transparent and self contained Indian language layer for data entry, storage, retrieval and printing in Indian scripts for your MS Windows applications.

These controls are bilingual and also support English language. The package contains editing control that support the complex scripts like Nastaliq. Following is a listing of controls that are part of the Urdu SDK controls package version 1.0 which are developed by C-DAC, GIST:

- UEdit
- UListbox
- UCombobox
- UStatic
- UButton
- UCheckbox
- URadiobutton

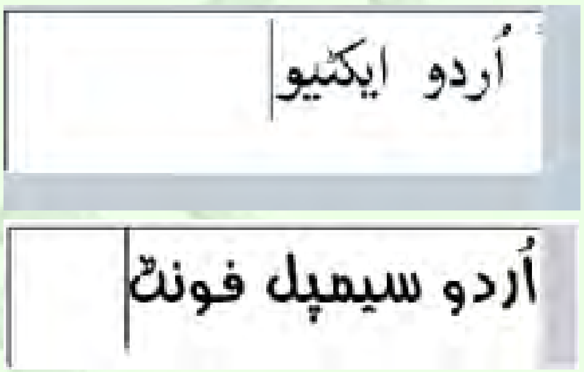
With the above set of controls, one can create all sorts of applications that support Urdu language.

The UrduEdit OCX is multi-line text editor control that supports Perso-Arabic Languages. This version (1.0) of the control supports URDU as the primary language. The control supports Naskh as the default script, but can also be used to depict Nastaliq script. The control has a number of features to provide with most of the functionalities needed for a Text editing control.

Features

1. Read & Write in Naskh / Nastaliq script
2. Bilingual text support i.e. English and Naskh
3. Copy paste text glyphs into external applications
4. Use for database applications
5. Transliterate (import) Hindi (ISCII) documents into Urdu
6. Use as simple Editor with fixed font and size
7. Use Rich Edit text facilities (e.g. multiple colors, font sizes, multiple fonts)
8. A number of Naskh fonts provided along with the control

Sample Urdu Active-X control in a form



اُردو ایکٹیو

اُردو سیمپل فونٹ

Applications using GIST URDU SDK

Following are some examples of set of applications where the UrduEdit OCX can be useful:

DatabaseApplication

The UrduEdit Control can be connected to any type of database, and store text as data. You can store (or load) text from a database from (or into) the control. Typical applications that can be created are mail merge, report generators, and all types of business applications where documents are created from information stored in a database.

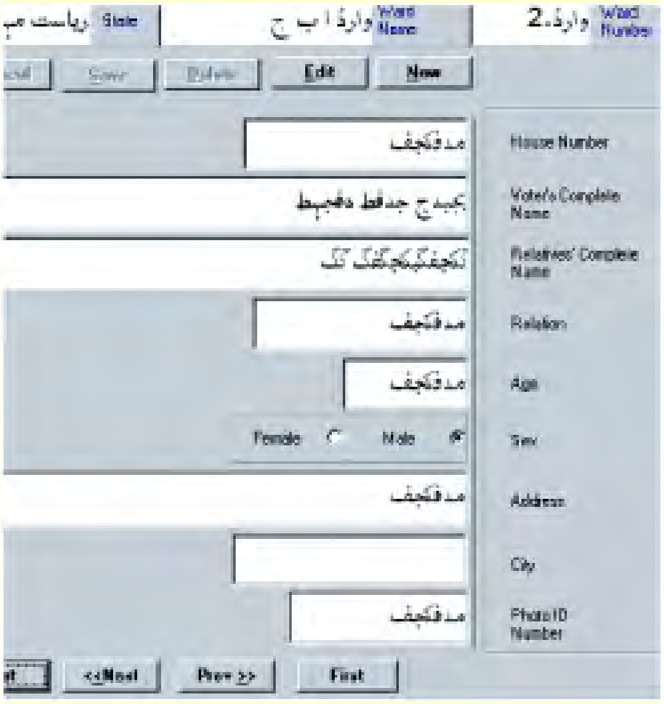
Desktop Applications

With UrduEdit Control you can easily create applications that need text fields for Perso-Arabic languages. The control can be used as single line or multiline control. You can also use Rich Edit feature of the control to display text in different fonts, colors, and sizes.

Web Applications

Create user controls with Visual Basic, use UrduEdit Control in web pages.

Sample Database application form using Urdu Active-X Control



State: ریاست جب
Ward Name: وارڈ ا ب ج
Ward Number: وارڈ 2
House Number: مدفکجف
Voter's Complete Name: بجیدج جدفط دفجبط
Relatives' Complete Name: ککجگفکگجگکگ کگ
Relation: مدفکجف
Age: مدفکجف
Sex: Female ○ Male ●
Address: مدفکجف
City:
Photo ID Number: مدفکجف

7. Pocket Translator

Multilingual Pocket Translator was developed keeping into mind the foreigners traveling to India. Even it is useful for persons traveling within India (Inter State). Similar products available worldwide but not for Indian Languages.

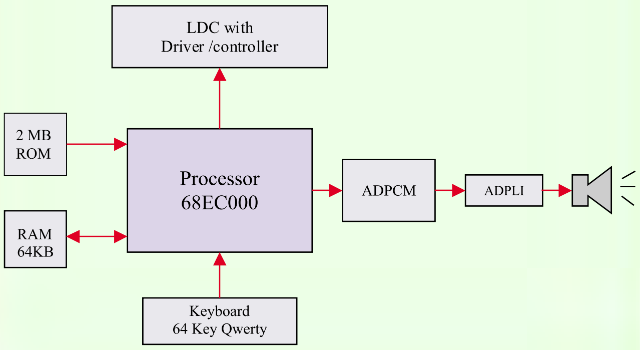
Features

1. English to Urdu & vice a versa translation.
2. Fixed messages for 5 different categories such as Social, Travel, Emergency, shop & restaurant, both for Urdu & English. 500 messages in each catagory.
3. Messages can be browsed in Urdu & English.
4. Speech output for Fixed messages only.
5. Single Toggle key for selection of various categories.
6. Arrow keys for scrolling through the messages/dictionary.
7. Script toggle key for changing the script.
8. 4 lines for English message display.
9. 2 lines of Urdu message display.
10. Talk key for speech output.
11. English - Urdu dictionary.
12. Displays nearest matching word in English if the typed word is not available in Dictionary.
13. Translation of the same to Urdu by pressing Script key.

Hardware Details

1. Developed on Motorola 68EC000 (PQFP package) processor running at 10 MHz.
2. 2 MB of EPROM for program, fonts, fixed messages, dictionary & speech o/p.
3. 16 KB of SRAM for Temporary variables.
4. 64 keys Matrix QWERTY type keyboard
5. 100 x 32 Dot matrix LCD display panel.
6. 4-bit ADPCM decoder for speech output.

Block Diagram of Pocket Translator



8. Multiprompter

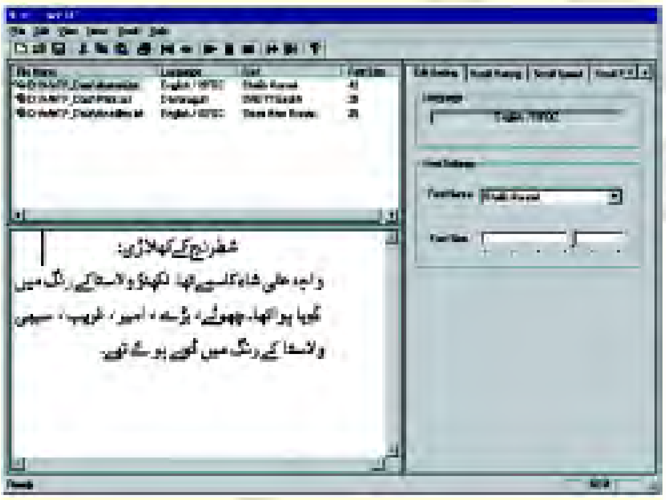
An ideal system for news reading and documentary production. Support for Naskh & Nastaliq scripts provided.

Read news or your dialogues without tears. You look into the camera and get to se the text rolling a a speed that suits your reading pace. And that too in India, Arabic or European languages. IT can have multiple mechanical attachments with optical glass for mounting on camera tripods. On-line skipping of stories.

Salient Features

Provide support to control the srcolling speed. a min. to 9 min. max and 0 for pause.

1. Support spellchecker.
2. Support Indian or European languages
3. Upgradeable to News Romm Automation System.
4. Works under Windows environment.
5. Available in two models : Desktop & portable based on Laptop.



شطرنج کے کھلاڑی:

واجد علی شاہ کا سمے تھا لکھنؤ ولاسثا کے رنگ میں

ڈوبا ہوا تھا۔ چھوٹے، بڑے، امیر، غریب، سبھی

ولاسثا کے رنگ میں ڈوبے ہوئے تھے۔

Following is the screen shot of the GUI (Graphical User Interface) of the Teleprompter application running on the windows platform. Used for creation of the news and documentaries which are then prompted on a special equipment having monitor, mirrored glass & camera.

9. LIPS Advanced Creation Station & LIPS for DVD

LIPS is a revolutionary, multilingual system meant for Electronic as well as fused subtitling. It consistes of highly productive software and cost effective hardware to create subtitles in Indian as well as foreign languagess along with the time codes.

LIPS for DVD allows you to create DVDs with multilingual subtitles in various Indian as well s foreign languages. It converts the LIPS subtitling files into a DVD compatible format. This system supports Diakin and Sonic formats. Generates the header indicating the background and foreground colors, edge color, the in-time and out-time as well as the subtitling image file.

Saient Features

1. Supports all Indian languages inclusive of URDU, Sindhi & Kashmiri.
2. Naskh & Nastaliq support provided.
3. Support TIF and BMP formats
4. Support different fonts and font sizes
5. Edge color for subtitles can be specified
6. Allows positioning of subtitles on the video
7. Provision for increment and decrement of in and out time
8. Compatible with PAL as well NTSC formats

Subtitling

10. Perso-Arabic Web Site Development

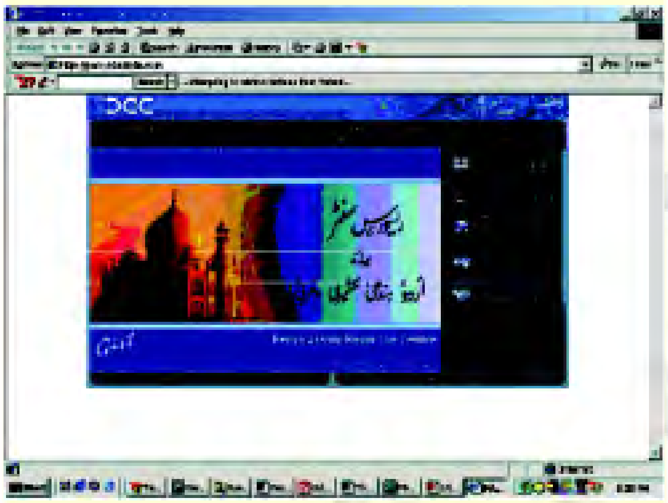
Under the Perso-Arabic resource centre one of the major deliverables was to create the web site in PA scripts & especially in the Nataliq script. Following was the method adopted for design & development of the website named http://parc.cdacindia.com

Designed a special activeX control (UEditWeb) for web that could display Nastaliq.

This control supports HTML like tags to set various attributes.

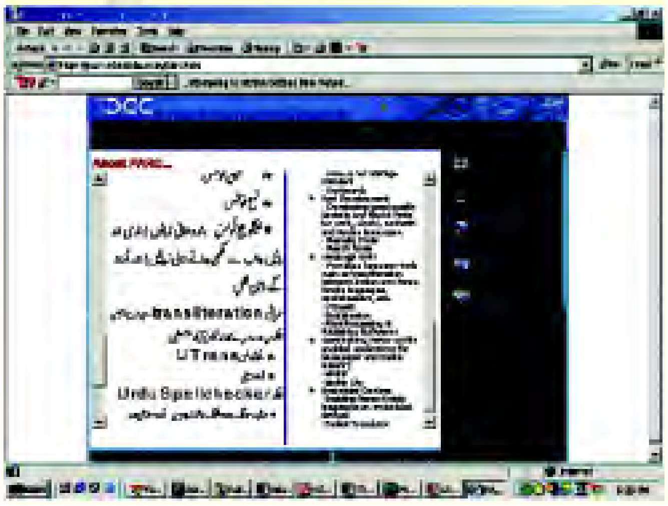
Converted USCII data into UEditWeb format.

UEditWeb data was stored in files which were put in Database.



Server side scripts were written to generate Urdu content for display.

Ten books will be added to make the site content rich.



Following screen shots shows the main page & the page containing information related to the resource centre website.

Courtesy : Shri. M. D. Kulkarni
Chief Investigator - Resource Centre &
the Resource Centre Team, GIST
Centre for Development of Advanced Computing
(C-DAC), Pune University Campus,
Ganeshkhind, Pune 411 007
Tel : 00-91-20-5694000/01/02, 5694092
Fax : 91-20-5694059
E-mail : mdk@cdacindia.com