

5.9 CoIL-Net : Tools & Support for Information Processing in Hindi Centre for Development of Advanced Computing, Pune

The CoIL-Net is initiative of TDIL programme under the agies of MC&IT encompassing a vision for all pervasive socio-economic developments by proliferating use of language specific IT based content, solutions and applications for bringing the benefits of IT revolution to the common citizen; requisite technology development and consequently bridging the existing digital divide.

Immediate Objectives of CoIL-Net programme

The immediate objectives of the CoIL-Net programme are aimed to proliferate the use of Hindi enabled IT based content, solutions and applications, along with developing requisite and associated technologies, in the states of MP, Chattisgarh, UP, Uttaranchal, Bihar, Jharkhand, and Rajasthan.

The objectives to be envisaged through

- Awareness Creation
- To evolve a Development Model
- Core Technologies, Tools and Utilities
- Content Development
- Human Resources for IT localization

CoIL-Net project was awarded to C-DAC, GIST with a specific mandate to develop tools & technologies for Hindi language, as required for the implementation of the CoIL projects at various centers. This also included giving certain tools for deployment of applications along with training.

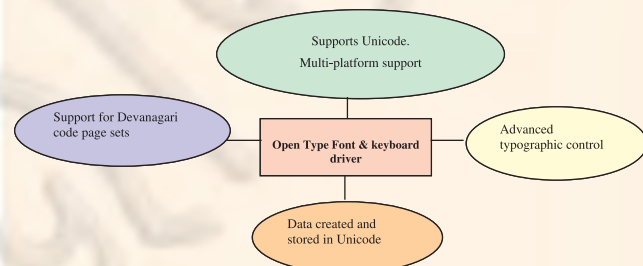
Following are the deliverables / milestone set for C-DAC, GIST

- Development of Open Type, Unicode based fonts for Hindi
- ISCII / UNICODE search engine.
- Java tools for Hindi
- Fonts plugins for supporting various browsers
- Enhanced Transliteration tool
- Dictionary Development Tool
- Training on GIST SDK/iPlugin tools.

HINDI OPEN TYPE FONT & KEYBOARD DRIVER Development

To develop Devanagari script Open Type fonts, with Unicode support & Keyboard driver for Hindi language, for promotion of Hindi in Information Technology. In all, two fonts are to be developed. The data will be stored in Unicode and then get rendered using the Open Type font rasterizer available on the Operating system. The latest operating systems are now supporting open Type fonts and these two fonts would help masses use Hindi on computers.

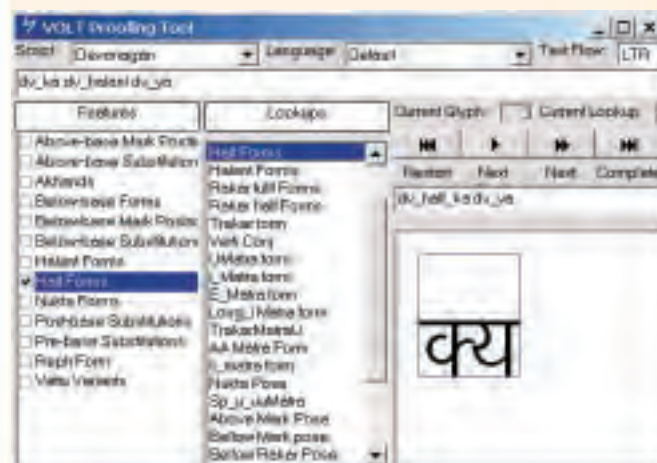
Open Type enables rendering of Unicode data, an international two-byte character encoding that covers virtually all of the world's languages. The user will also be able to view and create Unicode data using these fonts and move the data across various platforms.



Overall the usage of Hindi language on computers and in IT will get a boost.

The project module duration was 6 months & in that following activities were undertaken

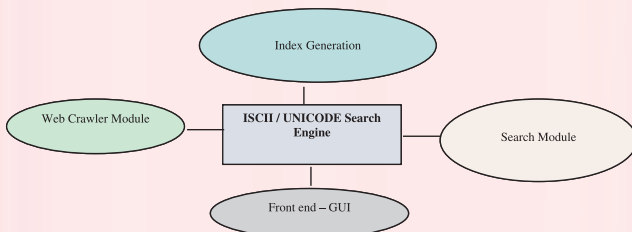
After a detailed Study, SRS preparation, the Design & Development of aesthetically good looking Open Type font for Hindi language was undertaken. The font



underwent rigorous testing & now the same is ready for deployment.

ISCII / Unicode Search Engine

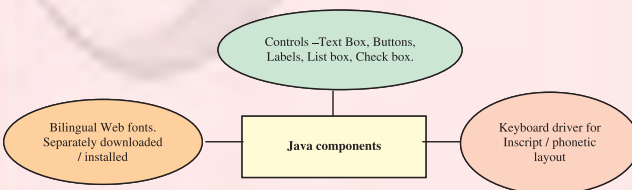
Development of ISCII / Unicode Search Engine for Hindi for searching the contents on the internet. Searching is one of the most often used web-services. This engine will help users search and locate Hindi documents and information on the internet. The search criteria and result may be displayed in Hindi.



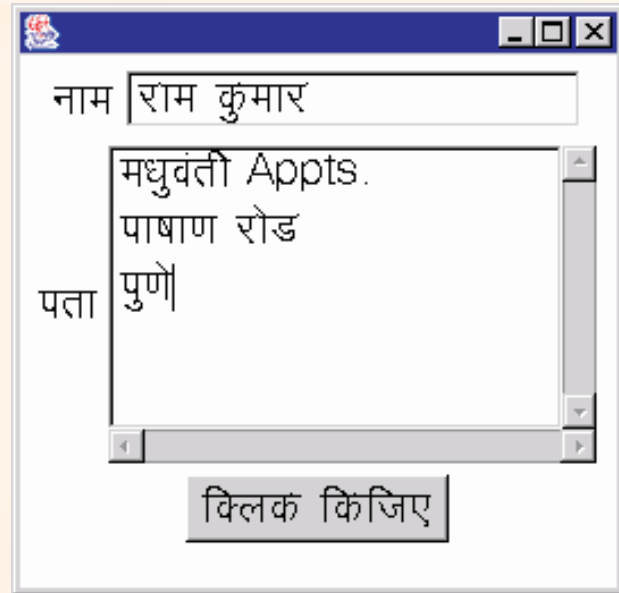
Currently an ISCII search engine is developed & requires certain fine tuning. This fine tuning will be taken care as per the requirements of the individual CoIL-Net centers while deploying their solutions.

JAVA components for Hindi

To develop platform independent 100% Pure Java components for Hindi, with Unicode support and Inscript/Phonetic keyboard layouts. These components can be used for developing desktop applications in a platform independent manner. The components designed herein will work with standard Java development IDE. These components can also be deployed for platforms, which have support for Java Runtime Environment.



The total duration of this project is 18 months which is inclusive of study, SRS & Design. Currently basic



Java components are developed & functional. Also they are parallelly undergoing field testing through the deployment of these components in the Digital Library project.

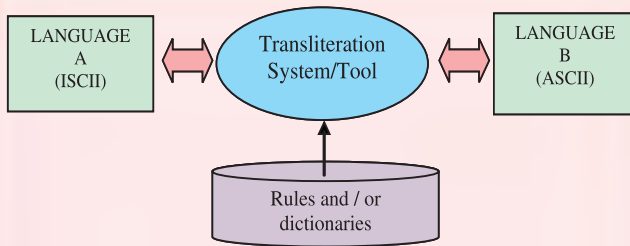
Font converters for supporting various browsers :

Development of font converter for supporting various browsers. The converter utility will be in the form of a desktop application that will allow the user to convert static HTML web-sites in various versions of ISFOC / INSFOC fonts to web version of ISFOC / INSFOC fonts for accurate rendering on browsers.

The basic converters are ready. Currently they are undergoing testing & packaged into an application that handles multiple HTML files along with the HTML parsing.

Enhanced Transliteration Tool

Enhanced Transliteration utility entails the development of a n-tier client-server based application for conversion of names from Hindi to English and vice versa. The client part of the application will be downloaded and installed on the user's machine. The server part of the application can remain on a public domain server. Requests for transliteration from the client's machine are sent to the server. The server processes the request and sends the resultant data back to the client.

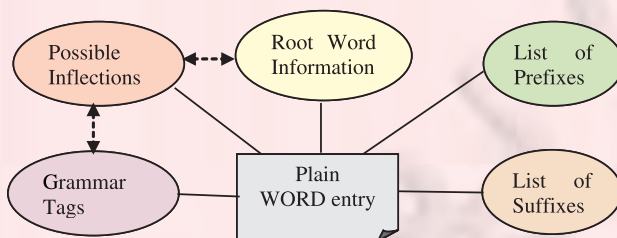


This project module includes study, SRS, design, study of current client-server technologies, study of different formats of databases. A client server based module for transliteration is being developed. The client side will get installed on the end user's system. This can connect to the server and upload data or files to be transliterated. The server will perform the transliteration and send the data back to the client.

Generic Dictionary Development Tool for Hindi

To Develop a "Generic Dictionary Development Tool" which will be primarily used to create parallel dictionaries.

Figure: Required Information with a single word entry



This tool will be in the form of a standalone desktop utility which allows the user to create dictionaries containing words. 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.



The project modules includes study, SRS, design, development, sample dictionary creation, validation & testing.

A tool for Urdu dictionaries has been made. This has to be suitably modified for this requirement. We also need to take into consideration the long term requirement of a dictionary building tool capable of handling parallel dictionaries etc.

Training

To impart training to all the CoIL-Net centers on using GIST tools. Tool learned during the training will help these resource centers to develop Indian language content on the web, hence aiding in proliferation of Indian Languages on the web. This will also ensure a uniform development in a standard procedure across all the CoILNet centres.

Training Activities

C-DAC Pune conducted training on GIST Tools – Iplugin, SDK, ISM for CoIL-Net centres, @ C-DAC Pune, Feb 2003 covering various tools & technologies developed by C-DAC.

*Courtesy : Shri Mahesh D. Kulkurni
Centre for Development of Advanced Computing
Pune University Campus, Ganesh Khind
Pune- 411007
Tel. : 020-5694092(O), 020-5694059 (F)
E-mail : mdk@cdacindia.com
mdk@cdac.ernet.in*