

20. Localization Case Implements-Banasthali Vidyapith under CoIL-Net

Dr. Aditya Shastri, E-mail : shastri@bv.emet.net

1. Introduction

Banasthali Vidyapith has been selected amongst a handful of Institutions to be a Resource Centre for IT localisation in Rajasthan by the Ministry of Communication & Information Technology (MCIT), Govt of India. In its initiative on "IT localisation for socio-economic development for the Hindi speaking states" MCIT has started the project "Content Creation of IT Localisation Network (CoIL-Net)" under TDIL programme. The project aims at bridging the digital divide by improving IT penetration and awareness level using Hindi as a medium of delivery, in which academic Institutions from the states of Rajasthan, Uttar Pradesh, Uttaranchal, Madhya Pradesh, Jharkhand, Bihar & Chattisgarh alongwith research groups from C-DAC, Pune, IGNC, New Delhi and IIT, Kanpur have been identified to develop core technologies for Hindi as well as IT localisation solutions.

The CoIL-Net project at Banasthali Vidyapith has focused on development of content and prototype localized applications in the areas of:

- e-health : (i) Health directory of Rajasthan
(ii) Ayurvedica
- e-business : Banking Information System
- e-tourism : Bilingual web portal on Rajasthan
- e-learning : (i) Vedic Literature
(ii) Distributed ITS for Hindi learning/teaching
(iii) Online course on IT Localization and disseminate them by way of holding IT localization clinics in the state.
(Figure -1)

The strategy adapted for the development is:

- I. To develop prototype applications in the above areas to demonstrate the development in above areas. The development follows the same approach all over which is
 - PC based
 - Platform independent

- Web based interactive systems
 - Bilingual at every stage on the click of mouse
 - Wherever possible Databases and interactivity both in Hindi and English
 - XMLisation wherever found appropriate
- II. To develop various tools needed for the application development such as
 - Transliteration Tool
 - Font Converters
 - Testing and evaluation tools.
 - III To disseminate the knowledge by holding trainer's training for educators in the field of **IT in the state.**
 - IV To hold IT localisation clinics to provide solutions and to discuss on issues in localisation with the local vendors and entrepreneurs in the field.

The first IT localisation clinic was held in November 2002 in Udaipur where the focus was mainly to create awareness amongst the people involved in IT applications development and in masses in general on the localisation- meaning and issues involved.

Rajasthan, inspite of having designated IT as a priority thrust area, has not been able to let its presence felt in the IT sector. The major Government Departments involved in strategic IT development in the state are

- DoIT
- Raj Comp
- NIC-State Unit

All three have on their priority agenda computerization of public utility services such as

- Road Transportation
- Electricity
- Water supply
- Medical & Health Information
- Land Records

A number of projects are implemented but localization issue is not handled in the projects in the right

perspective resulting in low penetration of the applications in masses.

Banasthali Vidyapith interacted very closely with these departments and then held the second IT Localisation in February, 2004 clinic exclusively for IT personnels in State Govt departments for proper deployment of IT localisation solutions.

In this paper we present the case study of developing the Localization Applications with specific reference to problems faced in terms of Computing platforms, fonts and encoding related issues, terminology and transliteration.

2. Localization

The process of modifying (adapting) a system or product to be used in the local environment, language and culture of a particular region or locality is termed as Localization. In fact the concept of localization leads to the process of Globalization which is defined as

$$\text{Globalization} = \text{Internationalization} + \text{N}^* \text{Localization}$$

The steps to be followed in a software localization are:

- (i) **Language enabling** - Encoding, rendering, input, cursor and edit key controls.
- (ii) **User Interface** - Menus, Dialogs, messages, documentation, Icons, layouts, Color Scheme, Interaction modes.
- (iii) **Operational** - interoperability issues, documentation standards practices.
- (iv) **Cultural** - behavioral models, rules.

In fact the complete process of Software Localization needs to follow a Software Engineering approach using an integrated tool set. The process involves:

- compilation
- translation
- localization
- testing and evaluation
- deployment and documentation

3. Case Study

In each of the applications developed at Banasthali Vidyapith in Coll-Net project or even earlier than that in the Vidyapith's Localization efforts (please see Table-1), the experience can be described as follows:

3.1 Compilation of the localized content: Compilation involves issues such as:

- defining scope
- compiling authentic content (from literature, personally gathering information)
- copyright issues

2.2 Translation: Translation requirement arises for:

- if the content is not available in local language
- translation of terms on user interfaces

There are translation tools available for English to Hindi translation. However, Banasthali Vidyapith did not use any and all translation was done by the project team in consultation with the language experts. Translation raised terminology related issues, very basic ones such as:

- whether to use Hindi term which is uncommon or to use the more common English term and write it in Devnagari - e.g.: Operating System
- whether to use standard terminology (even if not agreeing with it)
- the standard terminology is not complete (possibly can not be) and requires updations. Moreover it needs to be available online facilitating search.

2.3 Localization : Localization involved handling issues related to encoding, font rendering and other edit related issues. It also involved handling all factors listed in section 2. Being from a Hindi speaking state the issues of cultural factors in User Interface design were easy to handle and this effect can be seen in the screen layouts developed (sample screens in the figure below). But the encoding related issues are continuously popping up. Using windows based application development requires reedit the whole content as soon as a new version Windows come in the market. Only recently Unicode seems to have resolved these issues-but then encoding all content need to be converted to Unicode requiring Code Convertor. The Code Convertors provided with the product are not fool proof and the application requires Manual post edit. In fact even fonts need to be available with the application developer in a manner that as soon as a locale is selected in an application the related fonts appear automatically in the font list.

Another important Localization issue is Transliteration. Which involves both encoding and font. Not only that the map table has to be complete, the search in the table need to be efficient. Moreover, one may not want every input to be transliterated. For example, in databases, there are few fields having proper nouns and the localization means transliterating only those selected fields, How does one do that?

Many e-mail software provide 'Transliteration' facility (in fact some provide only that). But on cross server operation the results are not satisfactory (see example screens in figure -2)

Localizing web pages have to answer questions such as:

- is the website multilingual ?
- if yes strategy for correspondence both English-Hindi pages.
- design of active pages (we found it hard to design active bilingual page).

2.4 Testing and Evaluation: our experience finds this face the most difficult. In fact there are no metric present for Testing and Evaluation of localization efforts. We tried to develop - (see example screens in figure -3) for word processors.

One need to develop Benchmark for evaluation of any localized product/Application launched and then system in place to have mandatory the marketing.

2.5 Deployment & Documentation : There has to be a standard layout of Product Catalogue describing the localization issues and factors to be handled to a novice user

4. Suggestion for tools development to be available in public domain

- a subset of English-Hindi terminology available as Help in all localized applications. In fact a general shell architecture is needed which facilitates development of Application terminology glossary.
- a shell for Code-Convertor, where one enters the encoding tables of the pair of codes for which the convertor is to be designed.
- font enable engine involved with selection of Hindi locale.
- generalized transliteration engine.
- content management system for bilingual website design.
- Benchmark for localization products.
- Indian Locale support - Calendar, Calculator, Diary, Currency Convertor Games, Maps.

5. A FSM model for Software Localization

Any developer, whether developing an abinitio localized application or localizing an existing product must first prepare a checklist for compilation, Translation, Localization and testing and then can follow the following FSM model to achieve the desired output.

Table-1: Banasthali Vidyapith localization efforts

(A) Book writing:	
वनस्थली ग्रंथमाला	DOEACC ओ लेवल पाठ्यक्रमानुसार हिन्दी पुस्तकें
१. पास्कल में कंप्यूटर प्रोग्रामन	१. सूचना तकनीकी
२. कंप्यूटर से संख्यात्मक रीतियाँ	२. पी.सी. सॉफ्टवेयर
३. डिस्क ऑपरेटिंग सिस्टम	३. कोबोल में कंप्यूटर प्रोग्रामन
४. डीबेस छद्म प्लस	४. 'C' में कंप्यूटर प्रोग्रामन
५. वर्ड स्टार: एक परिचय	५. बिजनेस सिस्टम
६. कंप्यूटर संरचना एवं संगठन	
७. शब्द संसाधन वर्डस्टार के माध्यम से	

(B) Web based content Development:

वेब आधारित कम्प्यूटर विषयक पाठ्यसामग्री

१. सूचना तकनीकी
२. पी.सी. सॉफ्टवेयर
३. कोबोल में कंप्यूटर प्रोग्रामन
४. C' में कंप्यूटर प्रोग्रामन

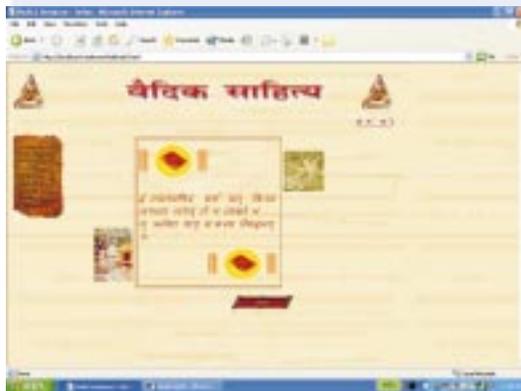
अन्य वेब आधारित अनुप्रयोग

१. ऑन लाइन अंग्रेजी-हिन्दी शब्दकोष
२. आकांक्षा - बच्चों के लिए हिन्दी में संपूर्ण दृश्य परिवेश
३. विज्ञान शब्द माला

कॉयल - नेट प्रायोजना में निर्माणाधीन अनुप्रयोग

- ◆ ई-स्वास्थ्य
 - १. आयुर्वेदिका
 - २. राजस्थान स्वास्थ्य निर्देशिका
- ◆ ई-व्यवसाय
 - बैंक सूचना तंत्र
 - मरू दर्शन
- ◆ ई-पर्यटन
 - १. वैदिक साहित्य
 - २. प्रबुद्ध शिक्षकीय तंत्र

e-Learning



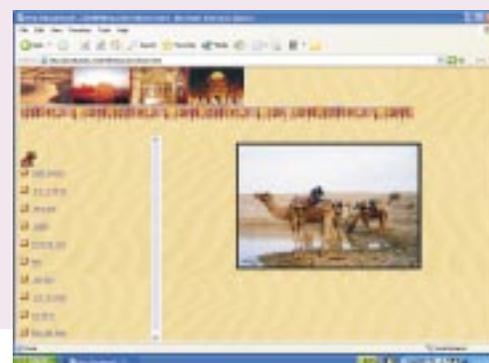
e-Health



e-Business



e-Tourism



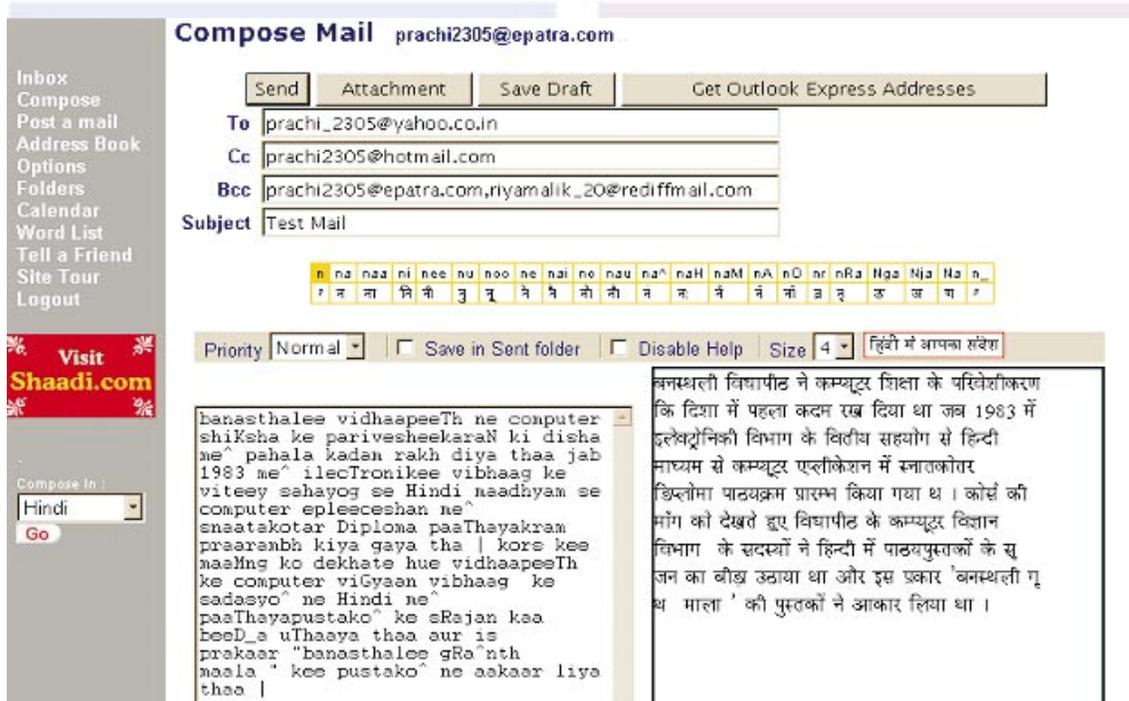


Figure -1 : CoIL-Net Project

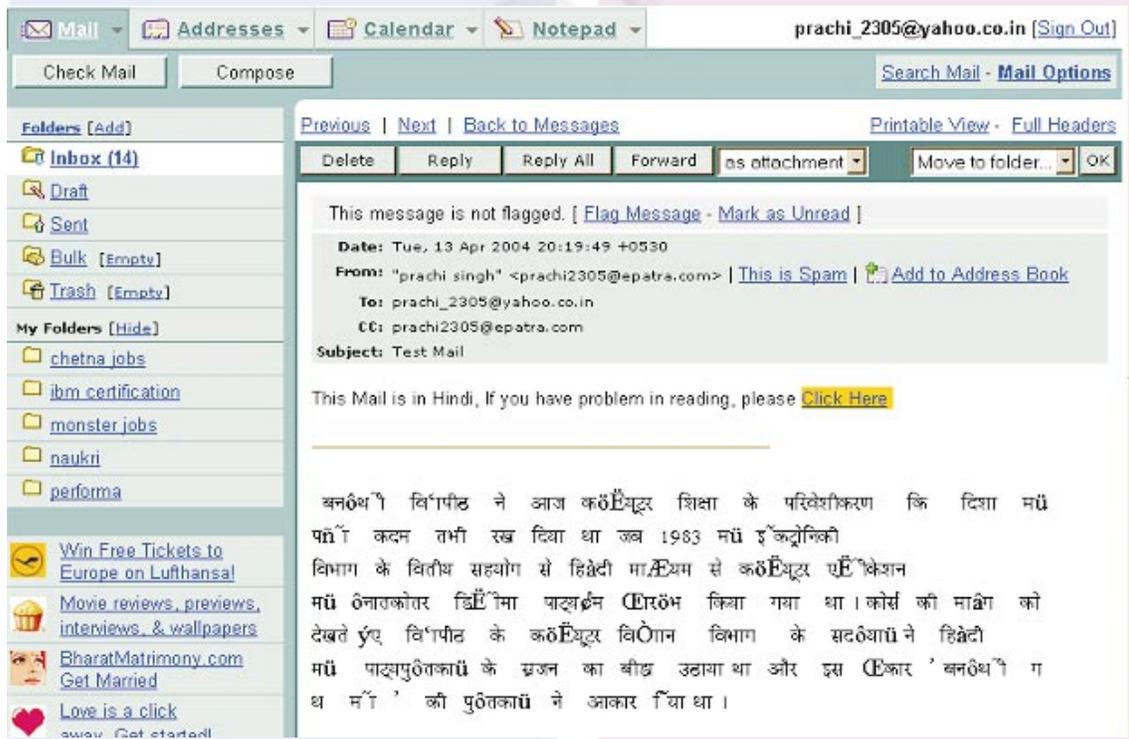


Figure-2 : epatra mail sent to yahoo mail server

Localisation Case Implements Under CoIL-Net Project

Form7

EVALUATION TOOL

SELECT THE CATEGORY

Word Processor Any Other

SELECT THE YENDOR

CDAC Any Other

SELECT THE PRODUCT

Apex Language Processor(ALP)
iLeap
ISFOC Script Manager(ISM)
ISM 2000 Office
LEAP Office 2000 Any Other

Version 4.0 Any Other

EVALUATE

Form7

Reports Help Exit

Evaluation of ISM 2000 Office Version 4.0

Functionality	Score	Feature
<input checked="" type="checkbox"/> Bilingual/Multilingual	10	Proper Typing
<input checked="" type="checkbox"/> Language Sensitive Multilingual Editor	5	Formatting Features
<input checked="" type="checkbox"/> Transliteration	4	Search and Replace
<input checked="" type="checkbox"/> Keyboard Shortcuts	5	Simple Editing Functions
<input checked="" type="checkbox"/> Number to Word Conversion	3	Table Manipulation
<input checked="" type="checkbox"/> Application Integration	4	Font Converters
<input checked="" type="checkbox"/> Text Sorting	2	Proper Text Rendering in the Browser
<input checked="" type="checkbox"/> Official Language Dictionary	4	Web Page Development Facility
<input checked="" type="checkbox"/> HTML Converter	2	Multilingual/Bilingual
<input checked="" type="checkbox"/> Shabdavali(Glossary)	0	Spell Checker Import/Export
<input checked="" type="checkbox"/> Mail Merge		
<input checked="" type="checkbox"/> Mail Facility		
<input checked="" type="checkbox"/> Facility for checking File Format		

Portability	Score	Feature
<input checked="" type="checkbox"/> Platform Independent		
<input checked="" type="checkbox"/> Encoding Conversion		

Usability	Score	Feature
<input checked="" type="checkbox"/> KeyClick on Screen Keyboard		
<input checked="" type="checkbox"/> Configurable Keyboard		
<input checked="" type="checkbox"/> Graphic Display of Keyboard Overlay	5	Various Keyboard Layouts
<input checked="" type="checkbox"/> Self Explanatory User Interface	3	

Total Score: 77

Grade: Good

Assessment: Acceptable

Start | Netnity | Project - M... | Norton AntiV... | Form7 | 11:41 AM

Figure-3 : Testing and Evaluation

Metrics Selection

- In the context of the Testing of Word Processors we have sorted the Metrics, on the basis of which Testing can be done.

These Metrics are :

1. Bilingual/Multilingual	16. Mail Merge
2. Proper Typing	17. Mail Facility
3. Formatting Features	18. Font Converters
4. Search and Replace	19. Proper Text Rendering in the Browser
5. Simple Editing Functions	20. Web Page Development Facility
6. Import/Export	21. Spell Checker
7. Language Sensitive Multilingual Editor	22. Check File Format
8. Transliteration	23. Shabdavali
9. Keyboard Shortcuts	24. Platform Independent
10. Number to word conversion	25. Encoding Conversion
11. Table Manipulation	26. Various Keyboard Layouts
12. Sorting	27. Graphics Display of Keyboard Overlay
13. Application Integration	28. Self Explanatory User Interface
14. Official Language Dictionary	29. User Configurable Keyboard
15. HTML Converter	30. Keyclick on Screen Keyboard

Marks assigned to each Tool

- Following is the result of Evaluation after Test Executions.
- Each Tool is Tested on various Metrics and marks assigned to each metric are shown in the Table

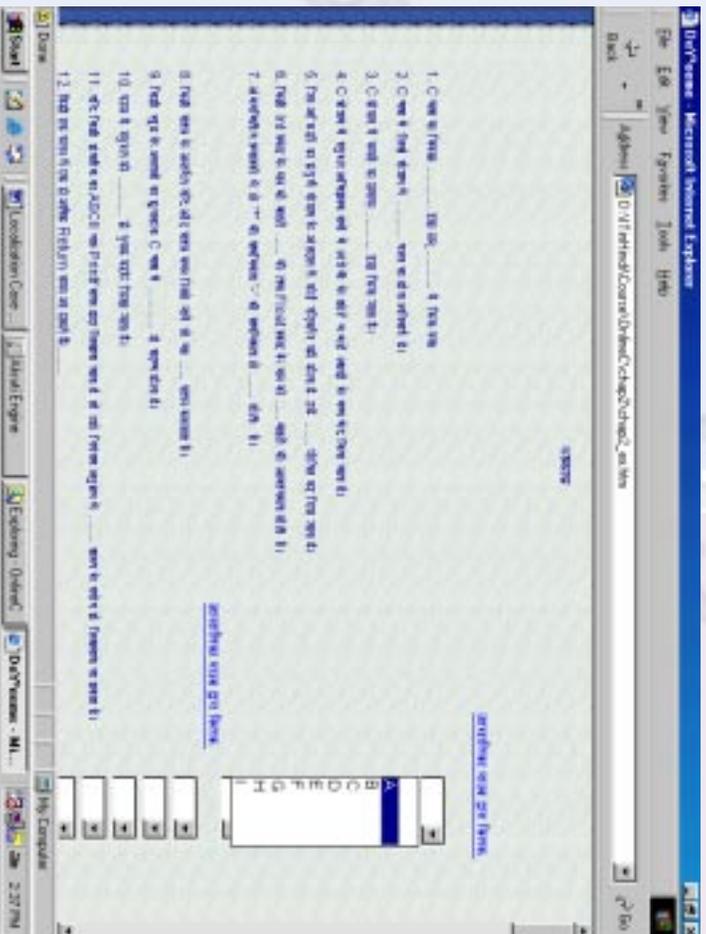
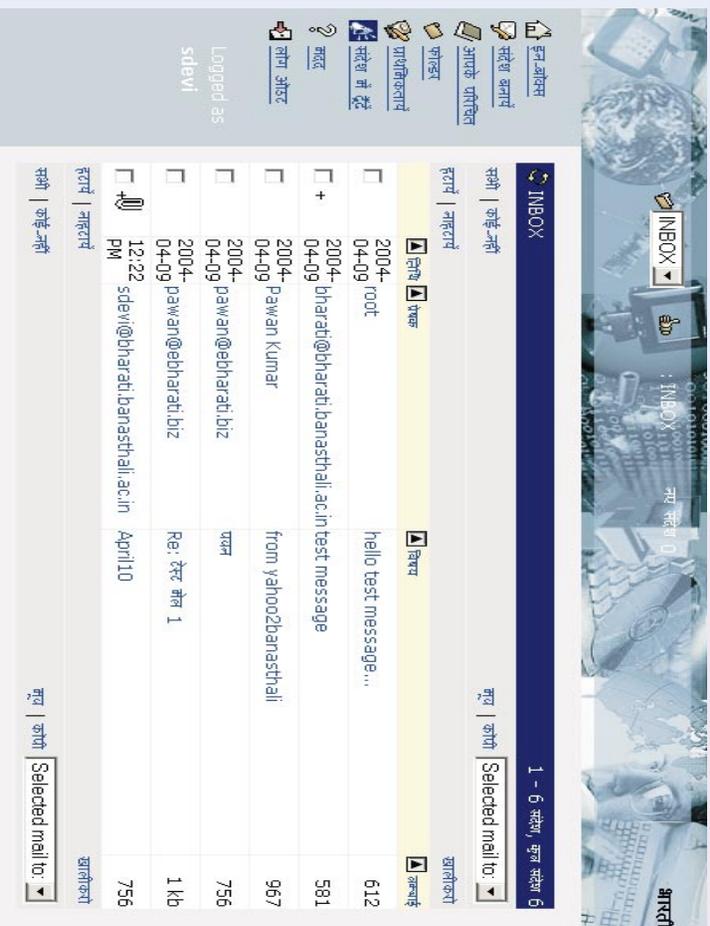
Metric	ISM Version 3.04	ISM 2000 Version 4.0	iLeap Version 2.07d	Akruti Version 6.0	Patrika Version 1.2
Bilingual/Multilingual	2	2	2	2	2
Language Sensitive Multilingual Editor	0	0	2	2	2
Transliteration	0	0	2	2	2
Keyboard Shortcuts	2	2	2	2	2
Number to Word Conversion	0	2	0	2	2
Application Integration	2	2	0	2	2
Text Sorting	0	0	0	0	2
Official Language Dictionary	2	2	2	2	2
HTML Converter	0	2	2	2	2
					Contd...

Metric	ISM Version 3.04	ISM 2000 Version 4.0	iLeap Version 2.07d	Akruti Version 6.0	Patrika Version 1.2
Shabdavali(Glossary)	0	0	0	0	0
Mail Merge	0	2	0	2	2
Mail Facility	0	0	2	0	2
Facility for Checking File Format	0	0	0	0	2
Proper Typing	10	10	10	10	10
Formatting Features	4	5	5	5	5
Search and Replace	4	4	4	4	4
Simple Editing Functions	5	5	4	5	3
Table Manipulation	3	3	0	3	2
Font Converters	0	4	0	4	0
Proper Text Rendering in the Browser	2	2	*	3	4
Web Page Development Facility	0	4	4	4	4
Multilingual/Bilingual Spelling Checker	2	2	4	2	0
Import/Export	0	0	4	0	4
Platform Independent	5	5	5	5	5
Encoding Conversion	5	5	0	5	5
Keyclick on Screen Keyboard	0	2	2	2	2
Configurable Keyboard	2	2	2	2	2
Graphic Display of Keyboard Overlay	2	2	2	2	2
Various Keyboard Layouts	5	5	5	5	5
Self Explanatory User Interface	2	3	3	4	3
Total	59	77	68	83	84

* Not confirmed (Because Evaluation copy does not allow Export in HTML Format as well as Pasting to other applications was not supported therefore creating an HTML page was not possible and thus its rendering could not be checked)

Result after Evaluation

Tool	Measured Value	Rated Level	Assessment
ISM Version 3.04	59	Fair	Acceptable/Satisfactory
ISM 2000 OfficeVersion 4.0	77	Good	Acceptable/Satisfactory
iLeap Vesion 2.07d	68	Fair	Acceptable/Satisfactory
Akruti Version 6.0	83	Good	Acceptable/Satisfactory
Shree Lipi Ankur Patrika Version 1.2	84	Good	Acceptable/Satisfactory



Localisation Case Implements Under Coll-Net Project