16. E-Governance-Localization Issues

Mr. Pradeep Chopra, Ms. Swaran Lata E-mail : pchopra/slata@mit.gov.in

Role of Information Technology for providing better services to the public is well recognized. Government has taken initiatives in the direction of achieving e-governance which is not merely computerization of stand alone back office operations but is meant to bring a fundamental change the way government operates and provides citizen interface. Working models for e-governance are being defined so that the government can perform effectively in a networked global economy. E-governance originated in India during the 70s with a focus on in-house government applications in the areas of Defence, Economy, monitoring planning and the deployment of ICT to manage data intensive functions related to elections, census, tax administration etc. The back bone connectivity provided by NIC up to the district level has proved as a boon. Now the e-governance is focusing on providing better citizen interface by deploying ICT solutions.

The use of Indian languages so far has been in office management applications mainly in the government and the public sector, the development of Indian languages enabled operating systems by multi-nationals like Microsoft has given a further boost for deploying IT solutions in Indian languages.

One of the major parameters for successful e-gov implementation is the deployment of applications in local languages so that every citizen of the country can access it without any hassle. A limited number of Indian language based applications are developed and deployed in e-governance domain. Most of these applications still have English interface. Static reports are generated using Indian language template. Few applications are enabled at data input-output in local languages. E-governance application penetration in various States is at various levels with Andhra Pradesh, Karnataka and Madhya Pradesh government being the first few to take major initiatives in this area. Reasons of non-proliferation of applications is due to the lack of appropriate language interface.

The language technology development in India date back to the development of GIST technology in 1982 and still there is a long way to go with regard to achieving seamless Indian language technology solutions. Based on the experience and study of major applications deployed in various government domains, the following issues need to be addressed:

I. Standardization

Existing developments are being carried out on a number of vendor dependent non-standard platforms which has resulted into non-interoperability of stored data. Hence, implementation of e-gov solutions using standards is of utmost priority. The standards can be categorized as follows:

a. Encoding Standards

The national standard, IS13491:1991, Indian Standard Code for Information Interchange (ISCII) is in-vogue, however, number of vendors while using ISCII still do data storage in their own font codes. Due to the limited market size, many vendors implement their proprietary codes to protect their market segment.

In the wake of globalization, multi-nationals around the globe formed a Unicode consortium to define the Unicode standard representing the languages of the world. The latest version of Unicode iS 4.0, brought out in 2004 in which Indian languages are also represented. Most of the e-gov applications are intranet/internet based, hence, the need to migrate to Unicode is now being felt, so that data can be seamlessly ported and accessed across the platforms which is not possible using ISCII.

b. Keyboard Standards:

INSCRIPT keyboard layout has been adopted in India which is widely used, however many other variants like Remington, Phonetic, Godrej etc. are also in use, because of the availability of trained manpower on existing typewriters. Major multi-national vendors are providing inscript keyboard facility and other keyboard lay outs depending on the user requirement. Inscript keyboard being the national standard is based on ISCII characterset. The additional characters in the Unicode adopted later on need to be represented on the keyboard, hence, there is a need to extrapolate the INSCRIPT keyboard to meet the current needs.

II. Data base related issues

Data entry for large e-gov applications is decentralized and handled by partially trained data entry operators. Due to linguistic characteristics of Indian languages, similar names are entered in different phonetic styles. Indian names such as Nayak and Naik, Shiva and Siva etc., though phonetically pronounced, similarly, but

are stored differently due to variation in data entry. The searching of data base causes problems, in case full text search is employed. Though soundex i.e. phonetic technique for near match search of data is a successful technique for retrieving similar sounding words for English, however, this technique is yet to be successfully developed for searching name data bases in Indian languages. The data generated using non-standard font based and vendor dependent fonts needs to be converted to ISCII or Unicode for easy interoperability. It is observed while entering data non-permissible combination of Matra are also entered by the data entry operators which does not appear on the display as error. This causes problems while searching the data base. There is need for availability of such utilities for checking and modifying the data stored in non-permissible combinations.

III. Transliteration

Most of the e-gov and other national applications need to be developed in multiple Indian languages for citizen interface across the country. Due to the multilingual data to be handled, there is need for transliteration of stored data from one language to another. The applications developed in ISCII are easily transliterated, however, necessary utilities must be provided by the software vendors for transliteration utilities in Unicode. Before migration to Unicode such tools need to be developed and made available.

IV. Indian language technology trained manpower

The present curriculum of the technical institutions of the country does not contain any modules of localization. Most of the technical institutions are not even sensitized or have felt the need to work in this area. Due to universal preference for English language, competent manpower to develop applications in local language is not available. Software developers are not aware of various language technology products and standards resulting in use of proprietary non-standard fonts for displaying content in local language only and not ensuring interoperability.

A minor percentage of the Engineering graduates either by way of working on Indian language related projects at few research institutions or by way of annual training as a part of the curriculum at such institutes get exposed to this area. Since job opportunities in this area are limited, many of them may not even pursue later on for their professional carrier, especially because it does not empower them to get good remuneration.

V. Vendor related issues

Currently most of the Indian language support is provided through plug-ins which run on existing applications and even if applications are generated ab-nitio, these run on the operating systems, many of which are proprietary. Some of the operating systems have also been enabled for few Indian languages at the data input output level. The developers use the plug in tools for such developments which also piggyback on the specific version of operating systems. The new versions of operating systems, browsers may cause clash with the code points of plug-in tools developed for Indian languages. This necessitates modification of Indian languages products/ plug-ins which are provided as up-grades.

The following issues are of concern:

- Being sellers market and limited current domain, vendors resort to higher prices policies

- Vendors are not equipped to provide technical support across the country due to a large geographical spread.

- The problems are also caused because the vendors do not adhere to standards and sometimes due to ignorance standards are not implemented properly.

- Multi-lingual support in applications as demanded by users in some of the applications may also put constraints on vendors because of lack of linguistic support or lack of knowledge.

- The applications developed by different vendors at time could not co-exist due to non-adherence to standards while developing applications.

- The upgrades are brought out by vendors after a long delay and vendors do not feel the responsibility of providing these upgrades free of cost to the existing users. Although upgrades are available in the market with additional cost burden yet due to lack of information users are unable to get these in time and also suffer because of higher pricing policy of the upgrades.

VI. Issues related to website and e-mail

Though the presently available technology facilitates developing website in Indian langauges there are still certain technological glitches which need to be addressed for deploying multi-lingual content on the websites. The content developed using ISCII or other 8bit proprietary softwares requires either the use of dynamic font or the facility to download the font for proper display of the content which is cumbersome procedure. Over an above this, different browsers display the content by rendering in a different way and hence, the content does not reproduced with 100% correctness. The popular search engines available on the web are unable to search the Indian language content.

The web content developed in Unicode does not have search issues, however content cannot be accessed through older machines which do not support Unicode.

The present day browsers which are not integrated with the specific operating systems are not able to display the Unicode content authentically due to rendering issues.

Sending e-mails in Indian languages with an attachment is not an issue if the recipient also uses the same software which was used for typing the e-mail. The e-mail sent in Indian languages are font dependent, hence, this issue could be resolved if messaging system use Unicode as standard for sending e-mails.

VII. Translation related issues

- Standardization issue

The terminology of e-gov application even for the same application varies across the country. Hence, there is a need to built domain specific terminology and dictionaries along with their mapping of authenticated translation of terminologies Indian languages there by covering regional diversity as well.

- Translation tools

There is a need to build quick translation tools which aid the human translators to meet the large translation requirement of the e-gov domain. The e-gov application developers are from the IT field and lack linguistic background, hence raising proper linguistic support for successful implementation of the project becomes an issue.

- Linguistic support

Moreover linguists are not familiar with Information Technology which sometimes puts a constraint in making them understand the user requirements of this domain. There is a need to provide linguistic support across the Indian languages for large e-gov applications so that authenticated terminologies are used for data integrity.

Localisation of Applications & Language Technology Standards