Challenges of Multilingualism and Possible Approach for Standardization of e-Governance Solutions in India

Swaran Lata[1]* and Somnath Chandra[1]

ABSTRACT

In this paper we have addressed the major challenges and issues involved in the multilingualism aspects towards standardization of e-governance solutions in India. The paper also investigates the benefits of adopting open standards and open source software in implementing multilingual e-governance solutions.

Keywords: Multilingualism, Standardization, UNICODE, Rendering, Open standard

1. Introduction

India is a multilingual country with twenty-two constitutionally recognized Indian Languages and with many variations of dialects. The success of e-Governance initiatives depends upon penetration of ICT and its applications at grass-root level and core objective of e-governance is to provide efficient citizen centric services and better access to knowledge and information. In India only six percent people are well conversant with English causing 'knowledge Divide' amongst the citizens. Language is the primary vector for communicating knowledge, thus the opportunity to use one's language on global information networks such as the Internet will determine the extent to which one can participate in the emerging knowledge society and have access to information. Considering the multilingual and multi-script diversity in India thus it is imperative that, e-Governance applications need to be implemented with language framework and all the proposed e-Governance applications need to be ideally multilingual or at least bilingual (English and Official Language of the State). However, Indic scripts are complex in nature and technological solutions for accurate representation of these languages for input, storage display and retrieval requires in-depth technical and linguistic investigations. Adoption and adherence of global standards like UNICODE and W3C may mitigate the problems to a great extent, however, specificities and linguistic uniqueness of each Indian Languages and scripts also needs to be addressed in tandem. In this paper we have discussed challenges and possible approaches to overcome various problems in the development and deployment of e-Governance solutions in the perspective complexity of Indian Languages and scripts.

2. Multilingualism perspective in e-Governance

In the earlier section we have already mentioned that language interface is very much essential for success of the e-Governance applications. In this section we shall elucidate our views.

2.1 Absence of local language citizen centric applications

In a recent survey, the Economic Intelligence report -2005 of United Nations shows that, India secures a

[1] Department of Information Technology, Ministry of Communications & Information Technology, New Delhi 110 003, India

* *Corresponding Author:* (E-mail: slata@mit.gov.in, Telephone: +91-11-24301240)

low position in the e-readiness ranking among the 65 participating countries. [The e-readiness score is 4.25: 2006 compared to 4.17 in 2005]. The salient features of the report are presented in the Figure 1.

The report also mentions that, though India had made great progress in the Software development sector, absence of development and deployment of citizen centric applications is the major cause for such low e-readiness score. This causes non-availability of information and knowledge to the citizens causing 'knowledge divide'. Moreover in India only 2% of the literate population can understand English, causing further inaccessibility. The existing Knowledge and Digital Divide induced by low accessibility has also been recently pointed out by UN, ITU and other international bodies (Figure 2). In each of these analysis, it has been pointed out that absence of language tools, interface and citizen centric G2C and C2C applications in local languages are major impediment for such kind of 'knowledge divide'. The linguistic diversity coupled with lack of G2C applications in local language has aggravated the problem in India. In-spite of the fact that, various state governments has taken initiative in deploying e-Governance applications, most of the applications are developed in English causing limited accessibility to the people in various state governments. This picture has also been reflected in the recent e-readiness survey report conducted by Department of Information Technology, Ministry of Communications & Information Technology. The salient features of the report are depicted in Figure 3.

It is evident from the above Figure 3 that there exists large disparity amongst various states in terms of e-readiness. The states which take proactive role in the development of citizen centric applications are ahead of other states in terms of e-readiness. However, there would be acceleration in the process of access of knowledge and information if these applications are available with local language interface.

2.2 Absence of State Governments directive on the usage of standardized language tools in development of E-Governance applications

State Governments are main user and beneficiaries of the E-governance applications as they provide citizen centric services of various kind. It is therefore imperative that, E-governance applications pertaining to a particular state needs to be in bilingual with both English and state language interface and preferably trilingual to include the Hindi, being the national language. The technical initiatives in making E-Governance a reality need to be backed by strong motivation at state and national level and also close coordination between state and central level. Also, the commonality of various problems in different states needs to be addressed in a holistic manner. However, lack of awareness and directive from the state governments on the use of state official language results in limited development and deployment of E-Governance solutions in the respective states. Therefore, State Governments needs to adopt the usage of local language in e-Gov solutions and accelerate the process of deployment of such solutions. Presently such approach is not much visible in various states resulting in non-availability of information and knowledge to the citizens. Additionally, software tools and linguistic resources for many officially recognized languages are not available in present day and the development such tools and resources needs to taken up in urgent basis to minimize the existing gaps between various states in terms of e-preparedness.

3. Requirements of E-Governance applications

A typical architecture for e-governance application is depicted in the Fig.4. It is well known that, most of the E-Governance solutions are Internet/ Intranet based applications except for a few stand alone applications. The web based services are portals as single-point access interface networked with enterprise backend applications, databases, and legacy systems and also to external systems, which basically involves seamless interaction between repository of data and the content management and access systems. The issues that primarily affect the e-governance applications are, interoperability between the heterogeneous systems and data repositories, absence of standard taxonomy, absence of compliance to best practices for seamless web-accessibility and lack of internationalization of the multimodal application interface.

2006 e-reactiness rank (of 68)	2005 rank	Country	2006 e-reactiness score (of 10)	2005 score	2006 e-reactiness rank (of 68)	2005 rank	Country	2006 e-reactiness score (of 10)	2005 score
1	1	Denmark	9.00	8.74	35	32	S. Africa	5.74	5.53
2	2	US	8.88	8.73	36	34	Slovakia	5.65	5.51
3	4	Switzerland	8.81	8.62	37	35	Malaysia	5.60	5.43
4	3	Sweden	8.74	8.64	38	40	Lithuania	5.45	5.04
5	5	UK	8.64	8.54	39 (tie)	37	Latvia	5.30	5.11
6	8	Netherland	8.60	8.23	39 (tie)	36	Mexico	5.30	5.21
7	6	Finland	8.55	8.32	41	38	Brail	5.29	5.07
8	10	Australia	8.50	8.22	42	39	Argentina	5.27	5.05
9	12	Canada	8.37	8.03	43	41	Jamaica	5.03	4.82
10	6	Hong Kong	8.36	8.32	44	42	Bulgaria	4.86	4.68
11	9	Norway	8.35	8.27	45	43	Turkey	4.77	4.58
12	12	Germany	8.34	8.03	46	46	S. Arabia	4.67	4.38
13	11	Singapore	8.24	8.18	47	44	Thailand	4.63	4.56
14 (tie)	16	N.Zealand	8.19	7.82	48	45	Venezuela	4.47	4.53
14 (tie)	14	Australia	8.19	8.01	49 (tie)	50	Peru	4.44	4.07
16	15	Ireland	8.09	7.98	49 (tie)	47	Romania	4.44	4.19
17	17	Belgium	7.99	7.71	51	48	Colombia	4.41	4.18
18	18	South Korea	7.90	7.66	52	52	Russia	4.30	3.98
19	19	France	7.86	7.61	53	49	India	4.25	4.17
20	--	Bermuda	7.81	--	54	--	Jordan	4.22	--
21	21	Japan	7.77	7.42	55	53	Egypt	4.14	3.90
22	20	Israel	7.59	7.45	56	51	Philippines	4.04	4.03
23	22	Taiwan	7.51	7.13	57	54	China	4.02	3.85
24	23	Spain	7.34	7.03	58	55	Ecuador	3.88	3.83
25	24	Italy	7.14	6.95	59	56	Sri Lanka	3.75	3.80
26	25	Portugal	7.07	6.90	60	58	Nigeria	3.69	3.46
27	26	Estonia	6.71	6.32	61	57	Ukraine	3.62	3.51
28	27	Slovenia	6.43	6.22	62	60	Indonesia	3.39	3.07
29	28	Greece	6.42	6.19	63	63	Algeria	3.32	2.94
30	--	UAE	6.32	--	64	62	Kazakhstan	3.22	2.97
31	31	Orile	6.19	5.97	65	59	Iran	3.15	3.08
32 (tie)	29	Creach Rep.	6.14	6.09	66	61	Vietnam	3.12	3.06
32 (tie)	30	Hungary	6.14	6.07	67	64	Pakistan	3.03	2.93
34	32	Poland	5.76	5.53	68	65	Azerbaijan	2.92	2.72

Figure 1: e-readiness report for the Asia-pacific region countries

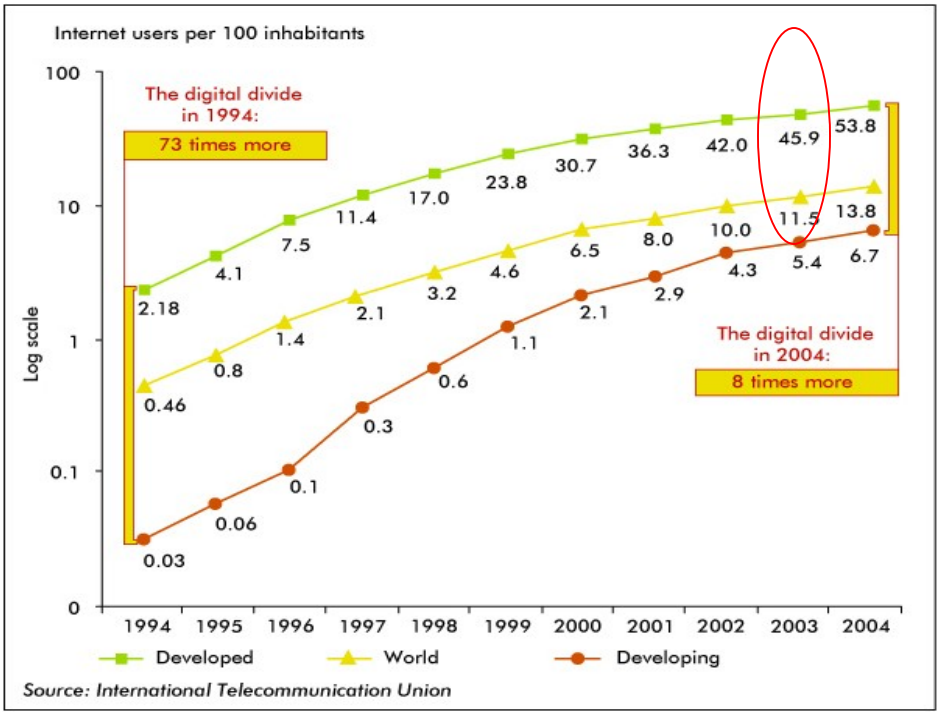

Figure 2: Digital Divide between Developed and developing nations

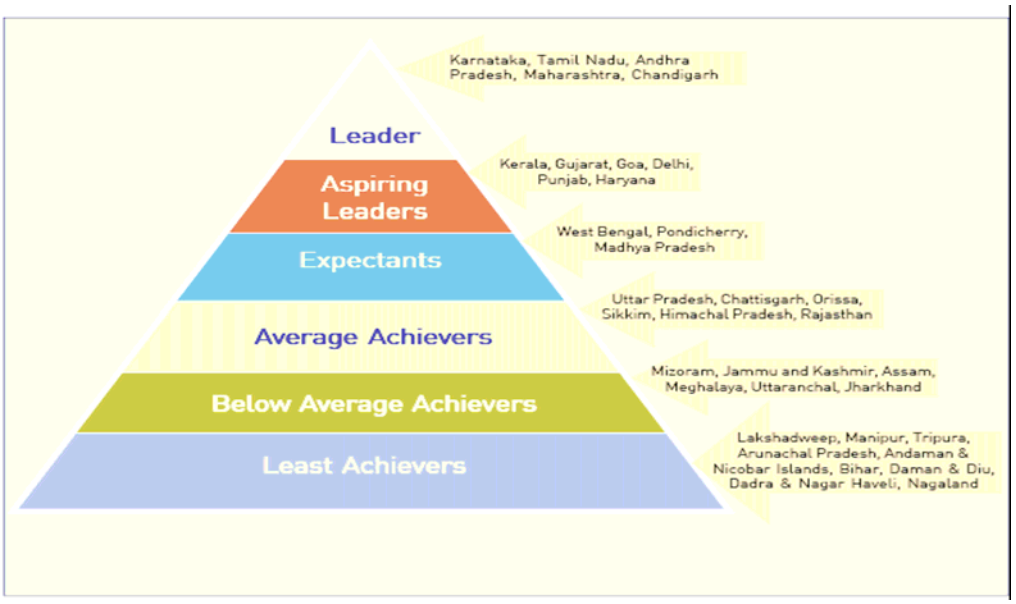

Figure 3: e-readiness amongst various states in India

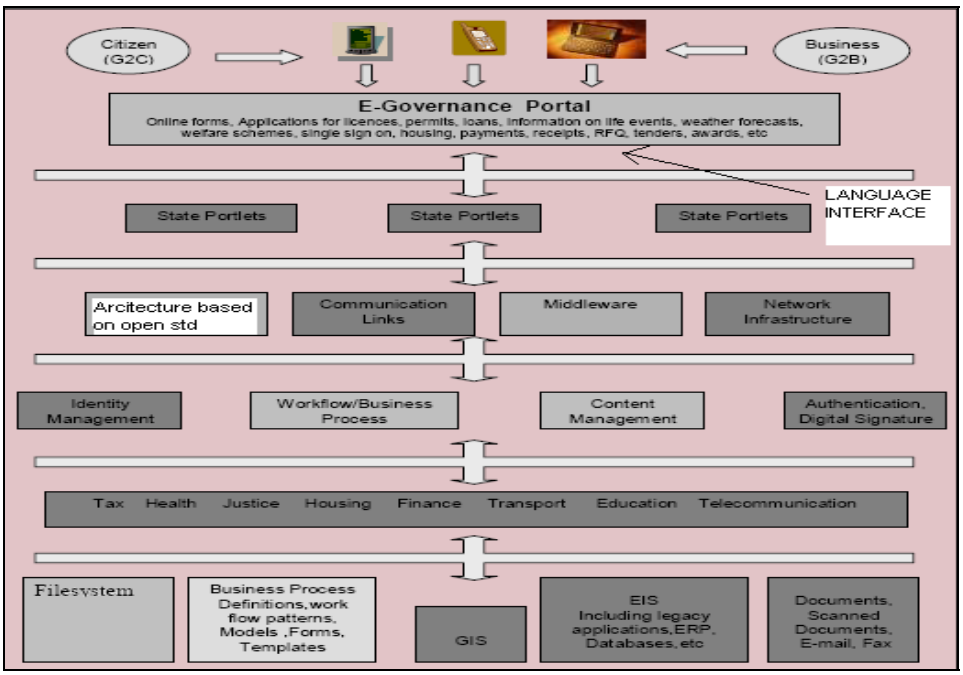

Figure 4: A typical E-Governance aaplication

Therefore, development of e-governance application services requires a paradigm shift from the present approach of ad-hoc solutions by integrating various application interfaces to a unified solution based on adoption of standards from *ab-initio* at every level of system development starting from various software components, databases and application interface. A study of the web-access of the e-governance portals shows that although the time and effort required for building a standard solution requires extra time and effort, the accessibility of web application increases by an order of magnitude compared to a non-standard one. This is presented in the Fig. 5 below. In the next section we shall list out the major challenges in respect of multilingualism for developing standard E-governance solution.

4. Challenges due to Plug-in Approaches:

4.1 Constraints Due to Data Lock-in:

Vendor lock-in is often used in the computer industry to describe the effects of a lack of compatibility between different systems**.** Different companies, or a single company, may create different versions of the same system architecture that can not interoperate. Manufacturers design their products so that replacement parts or add-on enhancements must be purchased from the same manufacturer, rather than from a third party. Due to Vendor Lock-in, data available in one platform would be difficult to be migrated to another platform. The problem of vendor lock-in results in non-uniformity in running legacy applications, already in place. To overcome the data lock-in problem use of open domain solutions may be adopted.

4.2 Transition from ISCII to UNICODE

The national standard, IS13491:1991, Indian Standard Code for Information Interchange (ISCII) has traditionally not been fully complied by a number of vendors to hold on their captured market segments. Some of these vendors still do data storage in their own their own font codes. Vendors are now migrating to UNICODE standard in the wake of globalization. However, there are several issues with respect to migration to UNICODE which are being addressed.

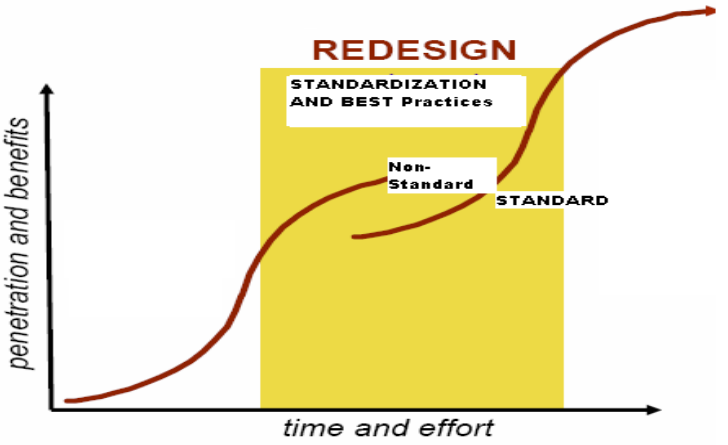

Figure 5: Accessibility index due to standardization

4.3 Problems due to improper rendering:

Although, UNICODE canonical conformance and normalization algorithms take care proper rendering of complex characters, some of complex conjunct characters are not properly reproduced during display due to improper rendering algorithms developed by the OS developers. The ISCII standard does not have such exhaustive canonical conformance mapping algorithms causing improper display. In this context, UNICODE canonical conformance mapping algorithm is more exhaustive than ISCII, however there are many complex conjuncts that need to be properly addressed by the rendering algorithms.

We have presented here some of the vertical conjuncts that are not properly displayed in WIN XP system

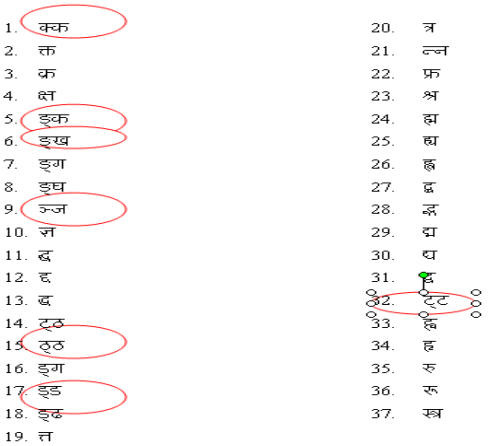

Figure 6: Improper display of conjuncts

Additionally, many of the words which are pronounced phonetically in a particular way, are stored in a different pattern in the UNICODE. The inputting of such words by any inexperienced person may lead to improper storage and display of such characters and words. Some of the examples are placed below:

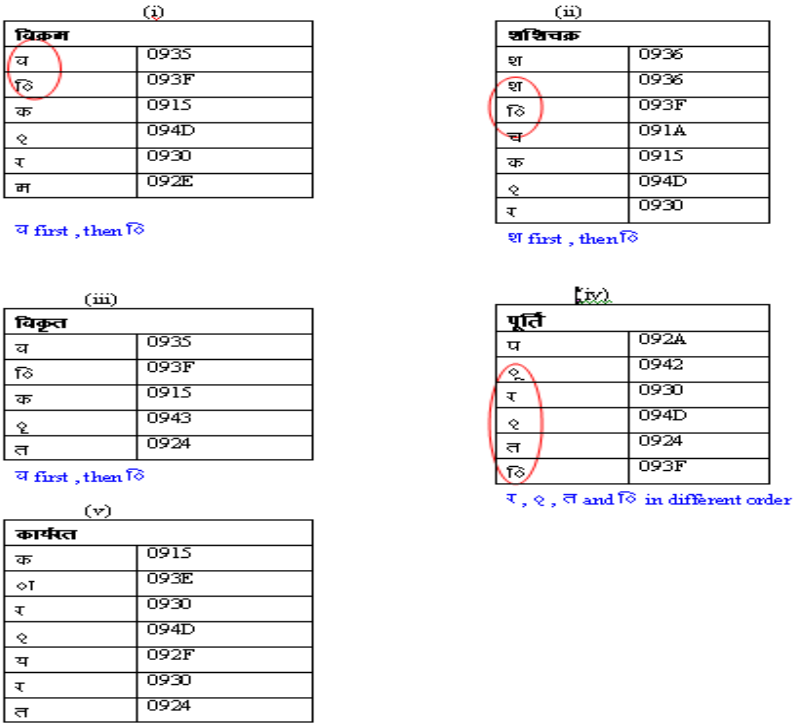

Figure 7: Different order for storage at UNICODE and display problem

Such rendering issues need to be addressed for one-to-one mapping correspondence between storage and display.

In addition to the discrepancies mentioned above, rendering Unicode text in Indian Languages can become application dependent. Even different applications running under the same Operating system do not have a standard API they can fall back on to get some uniformity. Some of the examples are presented in Figure 8.

It is evident from the Fig.8 that, text sent from a web server is handled differently by browsers and noticeable is the characters which could not be rendered.

4.4 Absence of Standardization in Key-Board:

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. In this context, we need to focus on one point that, Key-board layout is independent of storage code having multiple keyboard entries for a particular character should map to a single storage space. Specifically, if for example, a particular character in Indian language inputted through three different layouts should map to the specified UNICODE location for that particular character. A key board manager involving all three types of Key-board namely (i) Type-writer, (ii) INSCRIPT and (iii) Phonetic may be integrated as plug-ins with applications for ease of users. Hence standard implementation of keyboard driver in compliance with UNICODE is a challenge which needs to be addressed.

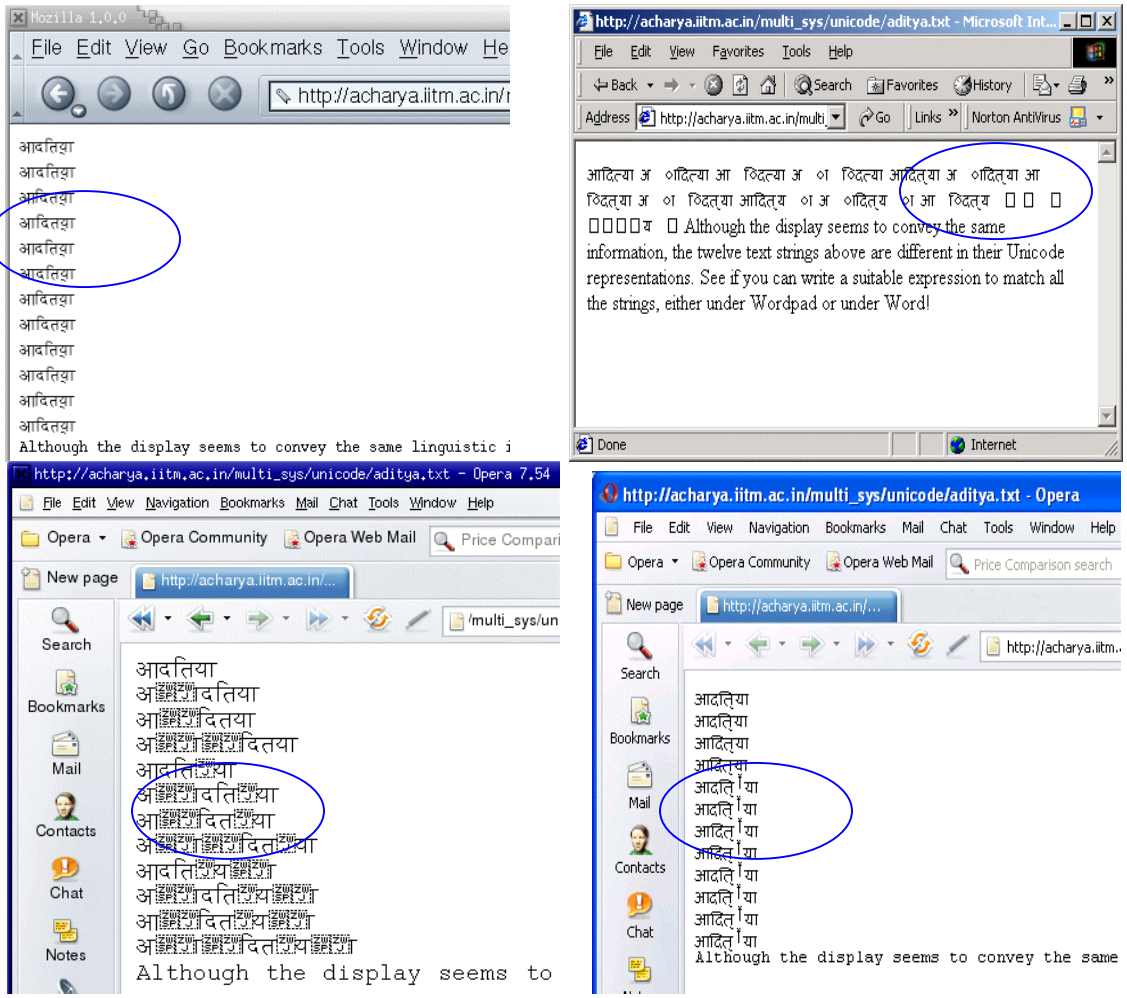

Figure 8: Rendering problem in different browsers

4.5 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. ITRANS is one of the popular schemes. ITRANS provides choice and uses capital letters also. The scheme is not suitable for searching the contents on the Internet. There is a need to evolve a Case-insensitive transliteration scheme to facilitate searching on web. The transliteration of Indic scripts needs to developed in conformance with ISO 15919 standard which is UNICODE compliant.

4.6 Information Retrieval

Various National E-Governance applications like National ID of Election Commission, Census data application etc. require search and retrieval capability switching between one language to another language.

Absence of searching and retrieval mechanisms causes non-uniform access of required information. Use and Adoption of existing standard search engines with proper language interface in Indian Languages is one of key requirement for national level E-Governance applications. In this context, we mention that, adoption of W3C for Web interface and UNICODE for storage may solve many of information retrieval problems.

4.7 Challenges related to Web and E-mail

Though the presently available technology facilitates developing website in Indian languages 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 8-bit proprietary software's 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, which will facilitate correct display of e-mail in default font.

4.8 Challenges related to Terminology Standardization:

The terminology for e-governance application though does not look so important, yet it can prove to be crucial if uniform terminology across the languages is not adopted. Hence, there is a need to build domain specific terminology and dictionaries along with their mapping of authenticated translation of terminologies Indian languages there by covering regional diversity as well. A parallel terminology dictionary in Indian languages would be a very useful tool for this purpose.

5. Need for Internationalization of the web-solutions

Future web based applications need to be developed in such a way that, applications should be interoperable for seam-less access of knowledge. World Wide Web Consortium (W3C) develops interoperable technologies along with specifications, guidelines, software, and tools. Adoption of W3C Internationalization Standards is necessary for developing inter-operable e-governance solutions in India. The Internationalization Activity, I18n, W3C Working produces specifications such as the Character Model for the World Wide Web, Web Services Internationalization Usage Scenarios, and Ruby Annotation Markup. It also provides upfront input to reviews Last Call Working Drafts on a wide range of topics, including Unicode character normalization, international typographic requirements, script issues in text-to-speech implementations, internationalization and localization requirements for schemas, usage scenarios and requirements for the internationalization of Web services, implementation of international resource identifiers, and many more. The Commonality of the E-governance solutions across various states can be exploited by developing a truly internationalized solution which can be localized for individual states in terms of the local needs such as adoption of local terminology and practices, based on W3C internationalization guidelines (I18N). Thus customized solutions for each state may be rolled out with incremental effort. It is to emphasize here that, localization process is not an inverted pyramid and if the solutions developed based on the plug-in approach for Indian languages the resultant solution may lead to constrained access to the service. As an example, we present below the XML tag-set modification using multi-layer schema customized for a particular Indian languages (Bengali) as per local requirement. The customized XML schemas for Bengali are shown in the Figure 9(a) and Figure 9(b).

```
B:\goutam-w3c-its\poscat1.xsd" xmlns:cont="http://www.kolkatacdac.in
/w3ci18ncd" xmlns:scat="http://www.kolkatacdac.in/w3ci18nsc"
xmlns:pcat="http://www.kolkatacdac.in/w3ci18npc">
<cont:content_domain name="agriculture">
 <scat:sentence_cat name="demonstrative">
   <pcat:pos_cat name="noun" meaning="farmer"> চাষি </pcat:pos_cat>
   <pcat:pos_cat name="verb" meaning="said"> বললেন </pcat:pos_cat>
   <pcat:pos_cat name="pronoun" meaning="I"> আমি </pcat:pos_cat>
   <pcat:pos_cat name="verb" type="missing_auxiliary" meaning="am"></pcat:pos_cat>
   <pcat:pos_cat name="adjective" meaning="a"> একজন </pcat:pos_cat>
   <pcat:pos_cat name="adjective" meaning="ordinary"> সামান্য </pcat:pos_cat>
   চাষি <pcat:pos_cat name="punctuation" type="sentence_final" meaning="."> । </pcat:pos_cat>
 </scat:sentence_cat>
```

Figure 9a: Bengali Schema for XML

```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE বইটেবিল SYSTEM "C:\Documents and Settings\Subham
Pahari\Desktop\xml test documents\bookdtd.dtd">
<!-- xml documents starts from here -->
<বইটেবিল>
<বই>
<বইনাম>সাহারায় শিহরণ </বইনাম>
<লেখক>জটায়ু</লেখক>
<দাম>৮৫.৫০</দাম>
<সংস্করণ>১,বৈশাখ ১৯৯২</সংস্করণ>
```

Figure 9a: Bengali Schema for XML

6. Possible Solution – Adoption of Open Standards and Open Source

In the section 3 we have presented the architecture of the e-governance service and associated problems due to absence of standards. We have seen that, most of the e-governance application services require seamless access and display of data, which implies interoperability of various components involved in the process. Traditionally, e-governance software are built and deployed by vendors without adhering to the standards and best practices for development of solutions. Also criticality and real time applications of various software components may often lead to poor access to the service. This problem is more relevant in the perspective of development of multilingual e-governance services especially in the G2C sector. Here lies the importance of the Open Standards. As per the ITU-T definition, "Open Standards are standards made available to the general public and are developed, approved and maintained via a collaborative and consensus driven process. "Open Standards" facilitate interoperability and data exchange among different products or services and are intended for widespread adoption. Open standards thus, ensure that different components of a large project can be built independently. They can be built in geographically different places, at different times, by different agencies and yet the components can interface, interact and transact in a consistent and robust manner, without loss of information. Let us now focus on why the open standards are important for developing the e-governance solutions. From the core business perspective, a cornerstone of Governance is to maintain records and documents. Therefore, information technology solutions must be able to address the basic requirement of maintaining records and documents for long periods of time without arbitrary "expiry of licenses" and "need for version upgrades". Government should be able to exercise its sovereignty on its records and documents without dependence. Also, multiple technologies could be allowed to continue in the e-governance space, open standards in the areas of architectural frameworks, design, coding, data, security, etc. would ensure uniformity in the e-governance space. The benefits of the open standards and associated open source have already started penetrating the e-governance sector. In India, states like Kerala, West-Bengal has already declared the policy of developing e-governance solutions based on open-standard and open source platforms

Considering the fact, that e-governance applications needs to be multilingual for wider access, the adoption of open standards are more required. UNICODE is an open standard. Adoption of UNICODE ensures uniformity of storage and data access. Similarly, for web-enabling of e-services services the portals and web applications should be built on the open standards and recommendations of W3C. All the standards of W3C for multilingual and multi-modal interfaces such as HTML, XHTML, XML, CSS, and VOICE XML 2.0 are open standards. Also, W3C offers a host of validation services using these standards. Since these standards and open, adopting these open standards and localization process following W3C guideline may easily ensure interoperability between storage, display and access to service even in the multilingual paradigm.

7. Concluding Remarks

In this paper we have discussed major challenges and bottlenecks for successful E-governance implementation. It has been shown that lack of local language interface is a major detrimental effect for wider proliferation of E-Governance applications in India. For successful deployment of E-Governance applications in multilingual domain, various standardization aspects related to Input mechanisms, Storage and Retrieval, and Output and Display mechanism need to be addressed in a national perspective. It is also necessary that open-standards to be in place and adopted for seamless access and interchange information and Moreover, various research aspects for futuristic tools such as Cross-Lingual Information Retrieval between Indian Languages and W3C compliant Indian Language Web-Browsers need to be initiated in an urgent basis.

Acknowledgement

The authors acknowledge support provided by Department of Information Technology to investigate various issues related to E-Governance. The authors also wish to thank team of NISG India for useful discussion.

References

1. *United Nations E-Governance Readiness Report 2006*: From E Government to E-inclusion.
2. *United Nations E-Governance Readiness Report 2005*.
3. *National E-Readiness Report 2004*, National Informatics Centre, Department of Information Technology
4. http://www.unicode.org
5. http://www.W3C.org
6. http://egovstandards.gov.in.

About the Authors

Swaran Lata is presently, Director and Scientist 'F' at Department of Information Technology, Ministry of Communications & Information Technology. She has obtained B.Tech (Electronics & Communication) with Gold Medal from Thaper Institute of Engg. & Technology and M.Tech (Computer Technology) from IIT Delhi. She joined Govt. of India in 1982 through Indian Engg. Services and worked in Directorate General of Technical Development, Ministry of Industry and made contributions towards the growth of Electronics industry in the country. She is member of various international standards organizations e.g. UNICODE and World Wide Web Consortium (W3C) and represents Govt of India to ensure representation of Indian languages in the international standards. She is also members of various expert committees and Working groups in the area of Localization and Internationalization. She is also member of Working Group on "Localization of Applications and Language Technology Standards for E-governance".

Somnath Chandra is presently working as Scientist 'C' in Department of Information Technology, Ministry of Communications & Information Technology. He has obtained his B.Tech, M.Tech and Ph.D. degree from University of Calcutta and IIT, Kharagpur respectively. Since 2002, he is working in the area of Language Technology and made significant contribution in the area of Multilingual Access of Digital Library and Cross-lingual information access system.