Issues & Challenges in Developing Multilingual Applications for Mobile: Indic Languages Perspective

Vijay Gugnani, Karunesh Arora, V N Shukla
Speech and Natural Language Processing Lab, C-DAC, Noida, India
[{vijaykumar, karunesharora, vnshukla}@cdacnoida.in]

Abstract

This paper describes experiences in designing and developing the software applications Indic languages targeted at mobile devices at CDAC Noida. CDAC has vast experience in multilingual computing and has also implemented several projects using browser based architecture. Now as we go further there is increasing trend towards browsing web through mobile devices in the remote places and while on the go, the Indic languages need to be tested out on various browsers for the performance. As a start a bilingual dictionary developed for different target platforms was ported on different mobile devices and experience was good as it is not easy for a developer to take off and start developing immediately. The environment requires lots of settings to successfully port and run application which have some language specific module. A lot of work needs to be carried out to make successful migration of useful applications and Indic languages websites accessible for masses on mobile devices.

1. Introduction

India is the best example of multilingual country in the world. There are 22 constitutionally recognized languages in India. In the last few years, there has been a sea change in the technology with global economy and strong presence of Indian companies in IT market. As the economy progressed the change has been seen in the lifestyle of people and information and communication has become an important aspect. This fact has evidence in reports published recently in the national newspaper Hindustan Times that "India, the world's fastest growing wireless services market, added 6.09 million new mobile customers in September 2006, boosting its user base to 129.53 million. Of the new additions, 4.39 million opted for services that run on the widely prevalent GSM platform, while 1.7 million signed up for CDMA-based services. The total number of new users in September was higher than the 5.9 million subscribers, who joined the booming sector in August, 2006 and the 5.39 million that entered the market in July 2006.

With the increasing use of mobile phones, people are expecting more utility applications. This kind of user base will naturally attract companies across the world to expand their market base and establish their products.

2. Localized mobile applications

Since the user base is spreading fast, a number of companies who have been working in this field may consider developing localized applications to be used by large number of people. Some of these applications may be device based and others will interact through mobile web browser.

1. **Text to Speech**: Many organizations have developed text to speech system for Indic languages which can be optimized for use with mobile devices for reading out the SMS and Contact list to help people with poor vision. It can also be used to read website content while browsing.

2. **OCR**: Most of the medium to high end mobile handsets available in Indian market today have the facility of digital camera. This feature can be extended for Indic languages by integrating the OCR in the mobile phones.

3. **Email & SMS**: Now it is possible to read email on mobile by sending the SMS to the portal offering service using a particular number without every time being connected to Internet. If the mobile phone browsers support Unicode or other standard encoding for Indic languages, this feature can be extended for Indic languages.

4. **Dictionaries**: Indic languages dictionaries can be very useful as mobile applications. These are already there for many non-Indic languages.

5. **Predictive Input** : Since mobile phones are used with single hand for inputting, predictive text inputting for Indic languages will help to input faster for SMS, email composition and searching the web using search engines where bandwidth and connectivity cost are important.

6. **Citizen Services:** People in India now use their mobiles to book railway tickets. Few companies are offering citizen services and travel information through Push based models or other implementations. This application combined with GPS can help to provide information about the nearby Railway Stations, Hotels, Hospitals, Medicine Shops and many other important places through web.

7. **GPS**: Maps and other applications are one area where lots of companies are coming up to provide help while traveling and or otherwise.

8. **Games**: Gaming is another area catching up. Games can be localized and crossword etc. can be developed in local languages to educate the children. For addressing few issues in such kind of applications requiring vertical arrangement of text, related work has gone in

this direction in internationalization working group of W3C. There is need to follow some standards which are being worked upon. The reference can be seen at

http://www.w3.org/International/notes/firstletter.html

9. C-DAC has developed speech corpus for the Hindi language under a project with ELDA France. In this project, speech of over 2000 speakers from different demographic profiles, environments and dialects have been recorded over mobile (GSM / CDMA networks). The speech data is being annotated and a lexicon is being developed. A wide range of utterances from isolated words, digit sequences, phonetically rich words and sentences to spontaneous responses have been recorded. The total vocabulary is about 11,600 items. The speech database consists of:
   - coverage of various dialectal variations in ratio of the populations speaking those dialects
   - coverage of phonetically rich words and sentences
   - coverage of speaking styles (commands, carefully pronounced and spontaneous speech)
   - coverage of environmental influences (through mobile in various environments)

This kind of database can help in producing TTS and ASR systems to work in the environment specific to mobile phones and Indic languages.

3. Application Development: Requirements

For all these applications to achieve success it is required to have access to the tools and environment for building these applications in Indic languages. There are a number of different programming languages and device types for application developers. As the memory size available with the mobile handsets are on increase one can think of the applications which take memory space in mobile phones.

4. Localization of Mobile applications

Mobile applications market is rapidly growing across the globe. The organizations that have already developed and localized their websites need to focus once again to revamp it for mobile phones. The websites/ applications development for mobile is little different from the desktop. The space is limited, bandwidth is a constraint, scrolling is cumbersome and image size etc has to be taken care of. There are lots of websites using the scripts which do not run on the mobile browsers very well. Localization experts often face the challenge of constrained memory, CPUs, power reserves, user interfaces, input devices, bandwidth, and connectivity.

Different countries have different locale requirements in terms of user preferences, technical requirements to support localized information, and the user interface requirements to present information appropriately. For Indian market and Indic languages, culture specific information needs to be presented as per local requirements.

Translation for Mobile devices

For correct translation of applications for mobile devices efforts needed from linguistic experts are huge along with domain knowledge and experience. This domain has challenges for linguists with the complexities and constraints of user interfaces, specifically with line and word-wrapping, text truncation, and the use of international icons. Text expansion or reduction may normally occur during translation because while translating the length of text may increase or decrease sometimes, e.g Tool means **उपकरण** in Hindi and **ग्राहक** stands for customer in English. Hyphenation and automatic line break issues may vary from language to language.

Linguistic experts should have the language specialization in targeted languages so that text could be rephrased for user-friendly translations in case of mobile devices. Linguists should not only with be familiar with localization process but also the cultural differences and language complexities.

Developers engaged in localization should be comfortable with vastly used technologies as well as proprietary software to successfully develop and deploy the solutions for language technologies. Therefore experienced and seasoned developers are required to handle the mix of complex scripting and technologies to develop localized applications as it is not merely replacing the content with translation.

5. Issues and challenges

Testing of Applications

Testing is a critical part of any software development life cycle and in case of localized applications it becomes even more important and complicated when application requires different combinations of hardware and testing scenarios. It requires extensive quality assurance testing, both functionally and linguistically.

- Functional testing of localized applications involves mix of hardware and multilingual environments. Functional testing can be completed with the actual devices or emulators. If the application runs on a simulator, it should be tested using various target locales in mind. Linguistic testing is the performance in functionality and usability. The testing and quality assurance people should have knowledge of targeted language to report problems related to rendering and display of characters, meanings conveyed through translations, handling composite messages and usability.

- There are number of options for mobile devices in India but it has been observed that there is huge market for low-end devices. Handsets manufacturers generally tie up with the Service providers to popularize their

phones for wide reach using promotional schemes. The applications therefore need to be tested for different handsets having different display sizes and screen resolution apart from different operating environment.

- Developers of websites and tool vendors for mobile devices sometimes forget to take into account the constraints such as rigid layout, small size of screen, resolution, navigation links, large pages to be loaded over cellular network, application in need of large memory and processing power.

- The mobile browser environment is highly fragmented. There are a number of browsers available for mobile devices. This situation has resulted in a number of challenges:
  - Tools vendors are yet to address mobile browsing in their tools;
  - Content developers have not really focused on mobile browsing as a target segment from Indic languages point of view. How the text should be encoded for mobiles UNICODE (UTF-8) or some other encoding.
  - Availability of generic and language specific guidelines for content developers in Indic languages.

6. Few experiences during development

During the course of development of few applications, the following observations were made

- There are not many resources available in terms of fonts for Indic languages for use with different mobile phones. Some available tools are not able to handle Indic language fonts as their conversion for device needs to handle the glyphs which are Indic languages specific. Many mistakes are observed for vowel modifiers while converting these fonts that means these have not been tested for Indic languages.

- There are few Unicode fonts available for specific devices and platforms which can be tested but again the application remains handset specific.

- Another issue of concern is that when a SMS is sent across devices of same make in Indian languages it works but when the same message is sent one device to other of different make, it fails to display. That means there is no common encoding being followed by devices manufacturers to transmit and receive the SMS message for Indic languages.

- Even if the fonts are made available to browse websites developed in Indian languages, a lot of work still needs to be carried out such as writing the installation steps for different platforms. Some of these phones especially the Windows based phones perhaps are able to use TTF fonts so this will be bit easy but needs to be tested out first. An example emulator image available is shown in the figure (source: http://www3.dw-world.de/mobile_emulator/emulator.php?lang=hindi

- There are a couple of mobile aware websites available which show the demonstration but limited to specific device or platform to display Indian languages content.

The big task

It has been observed that customizing a website and localizing it for use with mobile browsers is a big task especially when the world is yet resolving the basic issues such as handling long URIs, caching, going back, entry of queries by users, consistencies of UI across devices etc. However at the same time it is of utmost importance that the developers in India should test it out on different devices, platforms, and browsers, emulators to display content in Indic languages and provide their feedback to go further towards one web.

The first thing required for testing by a developer is emulator and therefore it becomes important to develop emulator images for Indic languages for different platforms.

The next important step should be to make available Indic languages fonts available for use in mobile devices across all platforms.

7. Current Status

Organizations across the world have been developing the products and devices to make the user experience better. There are number of browsers coming up for browsing the web in a better way using the mobile devices.

W3C has been doing lot of work in this area and has released few drafts and recommendations for developers and vendors under Mobile Web Initiative.

More details about Mobile Web Initiative of W3C can be had from http://www.w3.org/Mobile/. Details about SVG work on SVG can be seen at http://www.w3.org/Graphics/SVG/

Therefore, it becomes even more important to participate and provide feedback enriched with use cases for Indic languages content to take this work forward to provide Web for everyone, anywhere through any device.

8. Future directions

The future of the mobile interface is the Internet. As of now, mobile browsing is still in infancy in India, but few years down the line, most users will connect to Internet and use a mobile browser who are currently accessing it through desktop. This emphasizes bringing improvements in mobile browsing experience for users by offering fast speed, useful services and good usability. It is true that mobile phones are outscoring fixed lines in terms of number of users but in India we have long way to go to make use of mobile devices for utilities of mass use, as people are still using low-end phones and are getting acquainted with the technological advancements. The mobile even today is being used widely as a medium of communication. However as the technology has been advancing rapidly and smart devices are becoming

cheaper day by day, soon ,people will start taking the advantage of these devices for uses other than the vocal conversation. We have to really make ourselves ready to take these advancements ahead by testing the development tools for applications for mobile. Fonts and encoding issues should be resolved by choosing uniform encoding for exchange of information across all the devices and making available Indic language fonts for all devices and platforms. Guidelines for developing the content for browsing through mobile devices and usability standards should be evolved.

9. Acknowledgements

We would like to thank Executive Director, Centre for Development of Advanced Computing, Noida for constant support and conducive environment for this work. CDAC has been actively participating in W3C working groups under Web Internationalization Initiative for Indian languages project supported by Ministry of Communications & IT Govt. of India. Thanks are extended to TDIL team at DIT, Ministry of Communications & Information Technology, Govt. of India for supporting the activities. We also express our gratitude to Mr. C K Joseph Controller (Admin) and Sh. R.M. Goel Controller (Finance) for overall support. Also thanks are extended to Mr. Gour Mohan and Mr. Maneesh Kumar for setting up the environment and helping in testing the applications and browsing the web on mobile.

10. References

- Forum Nokia- resources for mobile application developers http://www.forum.nokia.com/
- W3C Internationalization (I18n) Activity http://w3.org/international
- W3C Mobile Web Initiative http://www.w3.org/Mobile/

- http://www3.dw-world.de/mobile_emulator/emulator.php?lang=hindi

- Experiences in Development of Hindi Speech Corpora based on ELDA standards Karunesh Arora, Sunita Arora, S S Agrawal, Niklas Paulsson, Khalid Choukri, O-COCOSDA Dec 2006 Malaysia