Dilemmas of Localization in Asia- A Case Study on Localization in Pakistan

SANA GUL (sana.gul@nu.edu.pk)
Regional Research Officer, PAN Localization Project (www.PANL10n.net)

Center for Research in Urdu Language Processing (CRULP)
National University of Computer and Emerging Sciences (NUCES),
B Block, Faisal Town, Lahore, Pakistan
Phone: 9242-111128128 x 268, Fax: 9242-5165323

Abstract:

Increasing efforts are being made to "Bridge the Digit Divide" by developing technical resources to develop local language enabled applications. However, investing into developing localized ICTs alone would not bridge this gap. This paper presents an overview of problems that restrict mass utilization of localized content in Pakistan. It details the major socio-economic factors attached with the slow pace of local language content flow in an era when technological infrastructure is available to support local language computing. The paper concludes by advancing recommendations for public and private sectors for effective and sustainable framework for localization in Pakistan.

Introduction

Asia houses majority of poor economies of the world. According to an estimate 32 out of 50 Asian countries are below the poverty line having a per capita income less than $500 [1]. A similar trend is observed in Pakistan where 32% of the total population is below poverty line [2].

Information and Communication Technologies (ICTs), specifically the internet has proved to be a diverse technological tool used to communicate, create, store, disseminate and manage information [3]. ICTs have elicited far reaching effects in the economic, social and cultural uplift of the once under developed populations [4]. In the developing countries, ICTs can transform old challenges and create unprecedented possibilities for sustainable economic development, just as it has done for business in the industrial world [5]. Realizing the beneficial returns from deploying ICTs, public sector is investing heavily in strengthening the telecommunication infrastructure. Government is taking measures to connect the remote areas with internet facility as an attempt to provide them access to the information world. A study of economic indicators of Pakistan's infrastructure reveal that about 1350 cities/towns/villages had been provided with internet connectivity up to March 2003. Total Telephone lines installed by March 2003 were 4.6 million and the mobile phone connectivity reached approximately 2 million by the end of the same month [6, 7]

Unfortunately even with ICT infrastructure in place, the number of total internet users is merely 5 million which totals to about 3.4% of the total Pakistani population [8,]. This suggests that only a small fraction of the total population is actually benefiting from the information available on the internet. [9].

The statistical debate presented above therefore absolutely rejects the stance that the low rate of internet usage within the country is due to the in availability of infrastructure.

Majority of this local population cannot access the ICT devices, since English and Spanish are the lingua franca for ICTs. Language barrier hampers this access, as local population is barely literate (adequately read and write) in their mother tongue. 20% of the total literates in Pakistan can read and write in English [9]. This 20% comprises entirely of urban population who form only 31% of the population. The dominance of English content on the internet therefore leaves the remaining 69% rural population "digitally divided" from the rest of the world.

Researchers are now well aware of the need to enable ICTs in the local languages of the consumer [10]. Movement for developing local language solution for legacy software systems is now underway [11]. However, even with present availability of localized tools like

Urdu fonts, proprietary software for Urdu on Windows XP, commercially developed Urdu word processors etc., only 3.4% of 149 million population in Pakistan is using ICTs.

Hence the urgency to un-cover and effectively address those issues that are major causes of depleted ICT usage within the country.

Problems:

Illiteracy

Issues pertaining to low internet usage are induced due to multiple socio-economic reasons. One obvious reason is the very low literacy level in ICT languages (English and Spanish specifically) due to which majority of the population remain aloof of the computer usage . At present only 20% of the total literate population of Pakistan understands English- the language of the internet. This language barrier could potentially be estranged through provision of localized solutions to the masses.

Lack of Computer Literacy

Even if localized solutions are built, the appallingly low computer literacy of the Pakistani population would pose another hurdle to access of information on the web. A very small fraction of the total population is literate enough to perform the basic computer operations.

Lack of Local Content

Even if the preliminary hurdle to enable ICTs in local language is overcome, the next obstacle to the access of information is the in-availability of "relevant" local language content [12]. This issue brings to the fore front the utmost need to produce local language content of direct relevance to the common man.

Lack of Standards

The exercise of content creation in local language would demand certain technological formalities to be fulfilled. Among the inevitable is the standardization of local language character sets and key board layouts both on linguistic and technological levels for developing localized applications. Software produced following these standards would ensure consistency across applications. [12].

High Cost of Localized Solutions

Still if we endeavor to overcome the issues stated above, the high cost of localized applications would keep these solutions out of the reach of the common man [13]. A few of the contemporary Urdu language packages for example, Inpage, Liwal and Urdu98 applications costing US$ 350 [14], 725 [15], 99 [16] respectively are evidently out of reach of the majority of the local populations.

Software Piracy

Even on the commercial level, these solutions are not popular due to their steep prices. The CEO of Systems Pvt. Ltd, (the developers of Raakim software, www.raakim.com) acknowledges that lack of commercial incentives and declining market for local language solutions has retarded their efforts enormously. Developing local language word processors was among the first initiatives in the area of local language technology. The clients of which were publishing houses and printing presses but they would not acquire the localized software for high cost constraints in the presence of a cheaply available pirated copies of the same solution. The clients would not mind to compromise over the output quality of the software due to their financial constraints.

Minimal ICT usage by Women

Apart from the above listed economic and technological constraints there are certain social problems that also play as an inhibitor for ICT utilization.

As an example, women visiting internet cafes is a taboo in Pakistani societal setup [17]. Such social restrictions would thus apply on the entire women population which forms about 48% of the total [6]. Similar gender biases thus decrease the ratio of women using the internet. There are further such constraints on the rural women of Pakistan because in rural

areas women are completely dependent on the male members of the family for the provision such opportunities.

No strong public policy on localization
All localization initiatives should be guarded by a back end localization policy. Unless the public agencies do not create a demand for local language applications no sound initiative for ICT development would come up.

Recommendations for Resolution
The problems quoted above, need to be addressed rigorously in order to ensure smooth transition of the digitally divided population towards its inclusion in the informed society. Various steps have to be implemented in the area of standardization for technology, human as well as technical resource development and certain level of change in the over all thinking pattern of the society regarding the uses and abuses of the ICTs.

Issue: Illiteracy
Solution: Speech Technologies for illiterate and handicapped

Foremost for the social, economic and technological uplift of a region is through the increase in its literacy rate. Increasing literacy rate to initiate ICT utilization is not a practical solution however it is realistic to build speech technologies for the illiterate and physically impaired populations. These speech tools may include text to speech conversion systems, speech to text solution that can facilitate desktop publishing and use of the internet to a considerable extent.
In order to address these issues one such initiative has been taken in Pakistan by the Electronic Government Directorate through collaboration with National University of Computer and Emerging Sciences. The project is known as the Urdu Localization Project. One part of this project is to develop a plug-in into the user browser and will be capable of translating the information from English to Urdu in real time [18].

Issue: Scarcity of Localized Solutions
Solution: Linguistic and Natural Language Processing (NLP) Studies
In order to develop sophisticated natural language processing technologies a dire need of trained human resource would therefore be felt. Universities across the country should offer under graduate and graduate level degree programs in various disciplines of natural language processing and local language technologies to catalyze the development of localized solutions which counts to be none at present [12] These universities should also establish research and development centers dedicated to research in localization. This research however has to be institutionalized within the universities in order to ensure sustainability in efforts. Pakistan at present has only one local language computing research center, Center for Research in Urdu Language Processing (CRULP), (www.crulp.org) which is involved in research in linguistics and computational aspects of the regional languages of Pakistan.

Issue: Scarcity of Localized Solutions
Solution: HR Development through Training/Seminars/Workshops
Human resource development programs through localization training can also catalyze the production of these localized solutions [19]. One such training initiative was taken by the PAN Localization project [20,21]. This training was conducted on the "Fundamentals of local language computing." The training aimed to equip the participants with education on the basic tools and technologies required to build intermediate to advanced local language applications. It was a five day long workshop that included presentations, hands on training sessions, font development exercises and lecture presentations by localization experts [22].
CRULP through the grant funds of the Ministry of IT conducted a one day long free seminar on "Font development" in Arid Agriculture University, Pakistan. Therefore, similar initiatives need to be taken up in order to stimulate the production of localized solutions.

Issue: Low Computer Literacy
Solution: End-User Training

End user training is essential to ensure the success of the developed product. Such training would further reduce the computer literacy ration within a country. To this affect presently the Pakistan Computer Bureau (PCB), successfully conducted the end user training for 6000 federal and 6800 provincial Govt. employees and the respective headquarters. PCB has also plans to dispense advisory services to various other public and private sector organization as a first step to commence the end user training practice in the country [7].

For the Urdu Localization Project, the Electronic Government Directorate of Pakistan also aims to conduct the end user training of the product after its commercial release. In a recent meeting with the Director Projects, EGD, Ministry of IT & T, Pakistan, with the author, the director explicitly specified that the organization visions to provide end user training facilities for all applications that are being developed through the initiatives of the EGD [23egd projects].

E-government Directorate of Pakistan recently performed an end user training of the Senator of the Pakistan government after deploying local language enable software in the Senate of Pakistan.

Issue: Lack of published Local Content
Solution: Local Content Creation Cells

Every possible effort to access information would be ineffective, unless the user get the information what he wants. Localized application would not prove successful unless users have access to content that is relevant to their purpose of study. For effective execution of a similar ideal content creation cells should be developed. These cells can be funded by the government to facilitate the relevant content creation in the country. These bodies should comprise of a mix of linguists, typists, social scientist, web developers and programmers to work in collaboration to produce the relevant content. Private sector organizations can be instrumental in developing such bodies.

Apart from manually translation of content into local language, sophisticated localization tools could be built to expedite the local content creation process. For instance, machine translators from English to local language can be helpful in this context. Other advanced localization tools like Optical Character Recognition Systems can be also be developed. OCR applications can be very beneficial in converting dictionaries, newspaper content, and other heavy printed material of local language into the computer formats. In addition, speech recognition tools can also be useful for local content creation which can facilitate local content creation by reading out local language data.

Issue: Lack of Local Standards
Solution: Linguist and Technological Standards development

Once the local content creation movement is in place next would be the urgency to develop technology standards for the local language. Government through the Language Authority\ Ministry of Culture should ensure that the technological and linguistic pre-requisites for enabling regional language into the ICT are provided to the technology developers. Language character set standards, key board layouts, collation sequence standards, and similar other standards necessary for localization should be maintained. [24].

Issue: High cost of Localized solutions
Solution: Investment in R n D organizations

Extremely steep costs of the proprietary local language solution would bring the entire movement of localization to an instant halt. Government through its Information technology division should invest in the organizations doing R n D in localization, facilitating them to develop local language software to be disseminated for free.

Issue: Software Piracy
Resolution Strict Public Policies

Software piracy should be completed eradicated, but at the same time options for availability of cheaper software should be introduced as explained in the preceding section. The government should ensure the development and practice of strict security of Intellectual Property Rights and copyrights of a vendor that has invested in the development of a software. Due to the negative implication of the open availability of pirated software, the trend

of development of Urdu software came to a drastic end during the middle of 90's. As a result of the copyrights violation, vendors ceased to develop localized software. [24].

Issue: Minimal Women Participation in ICT Usage
Resolution: Gender Parity in ICT Usage

Even if all of the above mentioned techno-economic issues are effectively met, one cannot adequately predict the maximal usage of the content available on the ICTs. This is because if the society itself condemns access through the ICTs, no effort for installing localization would be successful. A lady visiting an internet cafe alone is an absolute undesirable practice. To counter for such a situation, separate time slots should be reserved for women and men in the net cafes. In more conservative environments, separate internet cafes especially for women may be formed [17].

This situation is still worst in the rural areas. Reserving time slots for women in the villages would not help a lot, because women there are a dependency on their family men for majority practical decisions. Once if a negative content over the internet is observed, internet usage would be stigmatized. For this purpose, intense awareness programs must be conducted throughout the rural area explicitly highlighting the potential of ICTs as a tool for information exchange and its utilization in solving problems [25,26]. These awareness campaigns may for example be like the lady health worker initiative in the rural areas of Pakistan [lady health], Similar exercise could be duplicated to impart knowledge about ICT usage on door to door basis. Once a societal change is induced through the door to door awareness campaign, this movement can be taken to a larger scale by conducting information seminars within villages. This exercise can be complimented by distributing free demo Cds or information storage gadgets that present demonstrations on how the ICT usage can benefit even a common man [27].

There can also be target training programs especially for women, as although they constitute the 48% of the total population but are far more illiterate than the men in the rural areas. Women literacy in rural areas is only 6.4% in rural areas of Baluchistan, a largest province of Pakistan, according to one estimate [literacy ratios of population by sex, region and urban/rural area, 1981,1998 census].

Issue: No strong Public Policy for localization
Solution: Development of policies to promote Localization

The backbone for promoting localization initiatives should be available via strong public policy for endorsement of localized solution.

The public policy should explicitly address issues pertaining to securing Intellectual Property Rights. Copy right laws should be enforce but as said previously, it should also be assured that local language software is within the purchase power of its user.

To implement this, it should be policy driven that the government invests in its own IT sector and provides free to very cheap software applications.

The government should follow a focused approach towards proprietary software usage or towards the usage of open source applications for implementing local language solutions. The Government should have a targeted vision towards the uplift of it rural population through ICT utilization.

Targets for the development of Information centers within the rural areas should be introduced and separate task forces formed to address it. Government should take steps explicitly for the uplift of women as a target population by introducing awareness programs about the use and benefits of localized applications.

Computer education must be initiated at the very in initial levels of study so that students at the primary schooling receive education to use localized ICT.

Pakistan presently has a Urdu and regional Language software development clause in its Information technology policy [ ]. Owing to this policy, the government has recently started investing in the development of local language solutions.

Conclusion

In spite of the availability of the necessary infrastructure for localization, present ratio of ICT usage in the country in not encouraging at all. This would remain at lower levels unless the end users are provided an environment where ICT usage is encouraged from

within the societal set-up. Such a situation would thus adequately attend the call of the vast digitally divided Asian population, to diminish as much as possible their segregation from the information world. Government being the major player to diminish this segregation has to develop equal opportunities for the urban as well as the rural populations to participate in the information world. In efforts to minimize the digital divide the government has to eradicate social constraints for the usage of ICTs based on gender differences. Information usage is as essential to the male population as is to the other half of the population. This awareness reform has to be introduced by public as well as regional organization for the general benefit of the masses.

References

[1] http://www.worldbank.org/data/countryclass/classgroups.htm#top.

[2] **Income Distribution and Poverty,** Economic Survey 2002-03, Government of Pakistan, Finance Division, Economic Advisor's Wing Islamabad, 2003

[3] Blurton. C., "**New Directions of ICT-Use in Education**",
http://www.unesco.org/education/educprog/lwf/dl/edict.pdf

[4] Earth Report, **Communicating for Change**, UNDP, 2002-2003

[5] Talero E. and Gaudette P., **Harnessing Information for Development, A Proposal for a World Bank Group Strategy, The World Bank** Finance and Private Sector Development Vice Presidency Industry and Energy Department, Telecommunications and Informatics Division, March 1996

[6] **An overview of Economy,** Economic Survey 2002-03, Government of Pakistan, Finance Division, Economic Advisor's Wing Islamabad, 2003

[7,PCB] **Transports and Communications,** Economic Survey of Pakistan 2002-03, Government of Pakistan, Finance Division, Economic Advisor's Wing Islamabad, 2003

[8] **ICT at a glance, Pakistan**, Data and Statistics, Development Data Group, World Bank, www.worldbank.org , 2002

 [9] Masood, J. **Pakistan,** in *Digital Review of Asia Pacific 2003/2004*, Orbicom International Secretariat, Montreal, Canada, 2003.

[10] Soriano E.S., **Nets, Webs and the Information Infrastructure,** e-ASEAN Task Force, UNDP-APDIP, May 2003.

[11] **Microsoft Local Language Program,** Quote Sheet, March 2004**,**

[12] Hussain, S. and Gul, S. **LICT4D.pk**, accepted for publication in the *Proceedings of Frontiers of IT*, COMSATS Institute of Information Technology, Islamabad, Pakistan, December 2003.

[13] Sarker P. P., **ICTs for Development Initiatives in South Asia**, Bytes for All, (http://www.bytesforall.org)

[14] **Inpage Urdu Software**: Inpage personal, http://www.lingualearn.co.uk/learners/adpages/inpage.htm

[15] **Asiasoft™ Pashto, Farsi & Urdu Support, For Microsoft Windows XP**, http://www.liwal.com/windows/xp.htm

[16] **Urdu 98,** http://pakdata.com/urdu98.php

[17] Batool S., **Womenterpreneur @ Work**, in the proceedings of Internet Community Centers for and By Women, Sept 2002

[18] **Urdu Localization Project,** www.crulp.org/ulp

[19] M S Swaminathan, **Harnessing Local Knowledge Via Interactive Media**, in the proceedings of Policymakers Workshop, Village Information Centres, Chennai, India, 8-9 Oct, 2003

[20] **PAN Localization Project**, A regional initiative to develop local language computing capacity in Asia, www.panl10n.net

[21] Hussain S., **PAN Localization Project- A regional initiative to develop local language computing capacity in Asia,** I4D online magazine, June 2004

[22] Training on, **Fundamentals of Local Language Computing**, PAN Localization project, www.panl10n.net/training , Jan 2004

[23] Electronic Government Directorate, **Projects under Implementation 2003-04** www.pakistan.gov.pk

[24] Afzal, M, **Urdu Software Industry: Prospects, Problems and Need for Standards**, *Science Vision* 5(2). Islamabad, Pakistan, 1999.

[25] Kenny C., **The Costs and Benefits of ICTs for Direct Poverty Alleviation,** http://inet2002.org/CD-ROM/lu65rw2n/papers/u02-a.pdf

[26] John A, **The Information for Development Program: Encouraging the Use of ICTs in Developing Countries,** http://www.infodev.org/library/WorkingPapers/dalywp.pdf

[27] Okello D. , **Information Access for Rural Women, Experience of Women of Uganda Network (WOUGNET),** in the proceedings of Women crossing the Digital Divide: Empowerment through Information, Vienna, Austria, Oct 2003

[28] **Literacy Ratios of population by sex, region and urban/rural area, 1981,1998 census,** Economic Survey 2002-03, Government of Pakistan, Finance Division, Economic Advisor's Wing Islamabad, 2003

[29] **IT Policy,** www.pakistan.gov.pk/egd