Localization Industry in Pakistan

By
Dr. Sarmad Hussain and Sana Gul
Center for Research n Urdu Language Processing
National University of Computer and Emerging Sciences, Lahore, Pakistan

With 150 million people speaking more than 60 languages, localization of computing has been a natural requirement for Pakistan. Urdu is the national language of Pakistan and also the lingua franca. Other main languages include Punjabi, Siraiki, Pashto, Sindhi and Balochi, all of which are written in Arabic script in Pakistan. Local language computing in Pakistan dates back to early 1980s, when first DOS based word processors started emerging. Eventually, systems were developed for Microsoft Windows 3.1. Most of these systems were based on Naskh style of writing and supported Urdu and many other Pakistani languages. However, the preferred style of writing (especially for Urdu) has been Nastaleeq. As the latter style is very complex, initial word processors could not support it. Commercial level Nastaleeq support first emerged in mid-1980s when Monotype released Inpage Urdu word processor. This software is still widely used and is now slowly being replaced with the advent of Open Type Font technology based software.

Initially, most of the word processors developed were not based on any standards. However, advent of internet forced encoding standardization. Most of the existing software now is either based on Unicode or provides export and import facility for it. Unicode standard has also been updated by a proposal from Pakistan to include the initially missing characters of Urdu and other major languages of Pakistan. With minor problems, Unicode 4.0 supports these languages fairly well. Work is also underway to develop other computing standards in joint efforts by National Language Authority and Ministry of IT. Work is already in progress on locale standardization including definition of keyboard and collation. Work is also underway for translation and standardization of terminology used for interface of ICTs.

Internet and email have widely triggered localization in Pakistan. Local language internet, email and chat programs have also developed expectations of the user to have other application in local languages as well. This increased demand from users is encouraging the localization industry, which was earlier limited to desktop publishing. Having realized the necessity, multiple large projects are also being initiated by public sector for its internal use and for providing services to citizens through the e-governance programs. This is giving further boost to the localization industry of Pakistan.

One of the major public sector localization initiatives has been the National ID cards project. Earlier manually made ID cards were issued in the country. National Database and Registration Authority. (www.nadra.gov.pk) was specifically commissioned for automating this work in 2001. Their task was to develop a complete Urdu language database of Pakistani citizens and issue them computerized local language ID cards. Up till now, over 18.3 million Pakistanis have been issued these identity cards. Very recently Government of Pakistan has also commissioned development of computerized

Pakistani passport to NADRA as well. NADRA is also managing birth certificates and other similar national projects. Though many of these projects are developed in house, some work is also out-sourced to national and international localization companies, e.g. NCR has worked with NADRA for development of multilingual (English and Urdu) data storage systems.

Another large-scale initiative of Punjab state government includes automation of land revenue records. Similar projects are also being started at other states in Pakistan. Smaller projects include recently developed software to record proceeding of Senate and National Assembly of Government of Pakistan in Urdu. Recently, work has also stared to convert parts of official website of Government of Pakistan (www.pakistan.gov.pk) into Urdu. Most of this work is being done by private localization industry for the Government. Many of these companies are also involved in developing localized solutions for other countries, e.g. Spanish, German. This work is being outsourced from the respective countries.

Government is also eagerly supporting RnD projects to develop local language support. Recently awarded Urdu Localization Project (www.e-government.gov.pk) is aiming to develop an English to Urdu Machine Translation system to enable citizens to access English centric information on the Internet in their local language. The project also aims to develop an Urdu text to speech system to target the illiterate and disabled population. This three year project will end in June 2006.

Due to increased usage of computers and internet, localized operating systems are also under development. Work is in progress to localize both Microsoft Windows and Office suite and Linux. In addition, localization of hand-held devices is also being done by private sector. However, beyond Urdu, little work is being done for other local languages.

With incredible development of mobile sector, there is large demand for localized handsets and services. Companies like Nokia and Samsung are already providing localized interfaces in Urdu, though work in other languages is still missing. There are also limited services being provided by telecommunication service companies but there is huge potential in the market at this time, from local language SMS to more advanced localized services.

Last couple of years have seen an immense realization and focus shift from earlier English-centric work to localized tool and technologies. With increased customer demand and technological possibility, the localization industry in Pakistan looks more promising and dynamic than at any time in last two decades.

About the Authors:

Dr. Sarmad Hussain heads the Center for Research in Urdu Language Processing (CRULP, www.crulp.org) at National University of Computer and Emerging Sciences (www.nu.edu.pk) in Lahore, Pakistan. His interests include local language computing, including script, speech and language processing. He is currently leading Nafees Font (www.crulp.org), Urdu Localization (www.e-government.gov.pk) and PAN Localization (www.PANL10n.net) projects. He is also member of the national body of Pakistan for standardization of Pakistani languages for computing. He may be contacted at sarmad.hussain@nu.edu.pk.

Sana Gul is Regional Research Officer at CRULP for PAN Localization project focusing on Asian localization. She is actively involved in document localization status of Asian languages.