Heading: Localisation in Pakistan

Lead to by line: Dr. Sarmad Hussain & Sana Gul take a look at the localisation industry in Pakistan

Side-Heading: Country Focus

With 150 million people speaking more than 60 languages, localization of computing has been a natural requirement for Pakistan and has become an important industry in 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 the early 1980s, when DOS based word processors first 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 (e.g., because it is much more cursive and context sensitive), 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.

At first, most of the word processors developed were not based on any standards. However, the advent of the internet forced encoding standardisation. Most of the existing software now is either based on Unicode or provides export and import facility for it. The 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. Although Unicode has minor problems (e.g., letter Hamza joins in Urdu, but is defined as a non-joiner in Unicode), Unicode 4.0 supports these languages fairly well. Work is also underway to develop other computing standards in joint efforts by the National Language Authority and Ministry of IT of Government of Pakistan. Work is already in progress on locale standardisation including definition of keyboard and collation. Work is also underway for the translation and standardisation of terminology used for interface of Information and Communication Technologies (ICTs) like computers, mobile phones, handheld devices, etc.

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

One of the major public sector localisation initiatives has been the National ID cards project. Earlier hand-written manually made National ID cards were issued in Pakistan. The National Database and Registration Authority (NADRA, www.nadra.gov.pk) was specifically commissioned to automate this work in 2001. Their task was to develop a complete Urdu language database of Pakistani citizens and issue them computerised local language ID cards. To date, over 18.3 million Pakistanis have been issued these identity cards. Most recently, the Pakistani Government has also commissioned NADRA to develop computerised Pakistani passports as well as 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 localisation companies, for example, NCR's TeraData is being used by NADRA for development of multilingual (English and Urdu) data storage systems.

Another large-scale initiative of Punjab's state government includes the automation of land revenue records. Similar projects are also being started in other states of Pakistan. Smaller projects include recently developed software that records Senate and Government proceedings in Urdu. Work has also stared to convert parts of the official Pakistani Government website (www.pakistan.gov.pk) into Urdu. Most of this work is being sub-contracted to private localization/web-development companies by Ministry of IT. Many of these companies are also involved in developing localised websites for other languages, for example, Spanish and German. This work is being brought in from and developed for European businesses requiring e-commerce portals.

The Pakistani Government is also eagerly supporting Research and Development projects to help develop local language support. Urdu Localisation Project (project link from www.e-government.gov.pk) is aiming to develop an English to Urdu Machine Translation system to help 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 the increased usage of computers and the internet, localised operating systems are also under development. Work is in progress to localise both Microsoft Windows, Microsoft Office suite and Linux. In addition, the localisation of hand-held devices is also being completed by the private sector. However, except for Urdu, little work is being done for other local languages.

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

The last few years have seen an immense realisation and focus shift from earlier English-centric work to localised tools and technologies. With increased customer demand and

technological possibilities, the localisation industry in Pakistan looks more promising and dynamic than ever.

Dr. Sarmad Hussain heads the Center for Research in Urdu Language Processing (CRULP, www.crulp.org) at the National University of Computer and Emerging Sciences (NUCES) in Lahore, Pakistan. His interests include local language computing, including script, speech and language processing. He can be reached at sarmad.hussain@nu.edu.pk

Sana Gul is Regional Research Officer at the Center for Research in Urdu Language Processing (CRULP, NUCES) for PAN Localisation project. She is actively involved in documenting Localisation for Asian languages. She can be reached at sana.gul@nu.edu.pk