



PAN Localization

Working Papers

2004 - 2007

PAN Localization Project Teams



www.nu.edu.pk

IDRC  CRDI

Canada

www.idrc.ca

Published by

Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences
Lahore, Pakistan

Copyrights © International Development Research Center, Canada

Printed by PrintWorks, Pakistan

ISBN: 978-969-8961-05-3

This work was carried out with the aid of a grant from the International Development Research Centre (IDRC), Ottawa, Canada, administered through the Centre for Research in Urdu Language Processing (CRULP), National University of Computer and Emerging Sciences (NUCES), Pakistan.

*Dedicated to the memory of Prof. V. K. Samaranayak
A PAN Localization team member
and a pioneer of localization
in developing Asia*

May his soul rest in peace

Preface

PAN Localization project has been working since early 2004 to develop language computing capacity across Asia. The project initially started working in six countries, including Bangladesh, Bhutan, Cambodia, Laos, Nepal and Sri Lanka. Though this field was already mature in Sri Lanka, this area of research was novel to many of the other partner countries in this project at that time. The project had thus kept a significant focus on the development of local human capacity to work on computing solutions for local languages. Due to a varied degree of exposure to this and related fields, all participating countries took up language computing from a different level and were able to take it forward to a varying extent. Thus, the outputs of the project were never planned to be cohesive across the partners. This is also reflected in the papers compiled in this volume.

This volume accumulates the research reports and papers written by these teams over the past three years, on the variety of topics these country teams worked on. These research reports either were the initial literature review and design of applications which were eventually developed by these teams or were lessons learnt after these developments. In either case, the relevant software is also available through the project from its project website, www.PANL10n.net. As this has been a learning process for the participating teams, one would find a variety of experimentation, maturity and innovation in these research reports. These reports make a valuable contribution to the field of language computing, a fact which is even more inspiring as they are a product of young and nascent researchers. Most of the work produced by the country teams of PAN Localization project has been included for the purpose of documentation. For this reason, the work has not been peer reviewed. Some of this work has also been presented in conferences and published in journals. The quality and diversity of work included indicates that these teams are sure to play a leading role in the development of language computing in their respective countries in the future.

The papers have been divided by languages, not countries, as the countries may have many languages, but the research through the project has been more focused on a particular language within the country. The papers within a language are a product of a single team from the relevant country. These papers by a team are arranged along the concepts which are functionally sequential in the development of language computing. However, the sequence may not correctly represent the chronology in which the research was conducted and these papers were written. Therefore, one may find papers which may refer to later papers in the volume.

Each paper lists the people and organization(s) involved in the research and development. These authors deserve complete credit of the work. Furthermore, I would like to acknowledge the effort of the PAN Localization team at the regional secretariat for compiling and formatting this work. In addition, the support of International Development Research Center, Canada, is also acknowledged for funding PAN Localization project and of National University of Computer and Emerging Sciences, Pakistan, for executing and coordinating this project.

Sarmad Hussain
Regional Project Leader
PAN Localization Project

PAN Localization Project

Enabling local language computing is essential for access and generation of information, and also urgently required for development of Asian countries. PAN Localization project is a regional initiative to develop local language computing capacity in Asia. It is a partnership, sampling eight countries from South and South-East Asia, to research into the challenges and solutions for local language computing development. One of the basic principles of the project is to develop and enhance capacity of local institutions and resources to develop their own language solutions.

The PAN Localization Project has three broad objectives:

To raise sustainable human resource capacity in the Asian region for R&D in local language computing

To develop local language computing support for Asian languages

To advance policy for local language content creation and access across Asia for development

Human resource development is being addressed through national and regional trainings and through a regional support network being established. The trainings are both short and long term to address the needs of relevant Asian community. In partner countries, resource and organizational development is also carried out by their involvement in development of local language computing solutions. This also caters to the second objective. The research being carried out by the partner countries is strategically located at different research entry points along the technology spectrum, with each country conducting research that is critical in terms of the applications that need to be delivered to the country's user market. Moreover, PAN Localization project is playing an active role in raising awareness of the potential of local language computing for the development of Asian population. This will help focus the required attention and urgency to this important aspect of ICTs, and create the appropriate policy framework for its sustainable growth across Asia.

The scope of the PAN Localization project encompasses language computing in a broader sense, including linguistic standardization, computing applications, development platforms, content publishing and access, effective marketing and dissemination strategies and intellectual property rights issues. As the PAN Localization project researches into problems and solutions for local language computing across Asia, it is designed to sample the cultural and linguistic diversity in the whole region. The project also builds an Asian network of researchers to share learning and knowledge and publishes research outputs, including a comprehensive review at the end of the project, documenting effective processes, results and recommendations.

Countries (and languages) directly involved in the project include Afghanistan (Pashto and Dari), Bangladesh (Bangla), Bhutan (Dzongkha), Cambodia (Khmer), Laos (Lao), Nepal (Nepali), Sri Lanka (Sinhala and Tamil) and Pakistan, which is the regional secretariat. The project started in January 2004 and continued for three years in its first phase. It supported a team of seventy five resources across these eight countries to research and develop local language computing solutions. Further details of the project, its partner organizations, activities and outputs are available from its website, www.PANL10n.net.