

# Introduction & Scope of Localization

Sarmad Hussain

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

# Overview

- ▶ Introduction to Center for Research in Urdu Language Processing
- ▶ Introduction to PAN Localization Project
- ▶ Scope of Localization
- ▶ Introduction to the training on 'Fundamentals of Local Language Computing'

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

# مرکز تحقیقاتِ اردو

(Center for Research in Urdu Language Processing)

ایک تعارف

(An Introduction)



مرکز تحقیقاتِ اردو

# CRULP Objectives

- ▶ To conduct linguistic research for Urdu and regional languages
- ▶ To participate in standardization efforts in Urdu and regional languages
- ▶ To evolve computational models of Urdu and regional languages
- ▶ Promote content development in Urdu and regional languages

# CRULP Research

- ▶ Linguistics
- ▶ Script Processing
- ▶ Language Processing
- ▶ Speech Processing

# CRULP Resources

## ► Team

- 4 Full-time Faculty Members
- Adjunct Faculty
- 12 Graduate Students
- 45 Undergraduate Students
- 25 Full-time staff

# CRULP Coursework

- ▶ Phonetics and Phonology
- ▶ Morphology and Syntax
- ▶ Digital Signal Processing
- ▶ Random Variables and Stochastic Processes
- ▶ Speech Processing
- ▶ Computational Linguistics
- ▶ Image Processing
- ▶ Calligraphy and Font Development

# CRULP Research - Linguistics

## ► Areas

- Acoustic Phonetics
- Phonology
- Morphology
- Syntax

# CRULP Research - Script

- ▶ Font Development: Nafees Font Family
  - Nafees Nasta'leeq, Nafees Naskh, Nafees Pakistani Naskh (Urdu, Punjabi, Pashto, Sindhi, Balochi, Siraiki)
  - Freely downloadable from [www.crupl.org](http://www.crupl.org)
  - Supported mainly by UNDP/IDRC/APNIC Small Grants Program and partially by Microsoft, Pakistan
- ▶ Optical Character Recognition
  - Naskh (segmentation based)
  - Nasta'leeq (Ligature based)

# Nafees Nasta'leeq & Nafees Naskh

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ  
الحمد لله رب العالمین - الرحمن  
الرحیم - مالک يوم الدين - ایاک نعبد و  
ایاک نستعین - اهدنا الصراط المستقیم -  
صراط الذین انعمت علیهم  
غیر المغضوب علیهم ولا الضالین

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ  
الحمد لله رب العالمین -  
الرحمن الرحیم - مالک يوم  
الدين - ایاک نعبد و ایاک  
نستعین - اهدنا الصراط  
المستقیم - صراط الذین  
انعمت علیهم غیر المغضوب  
علیهم ولا الضالین -

وَسَخَّرَ الشَّمْسَ وَالْقَمَرَ

Nasta'leeq

وَسَخَّرَ الشَّمْسَ وَالْقَمَرَ

Kufi

وَسَخَّرَ الشَّمْسَ وَالْقَمَرَ

Sulus

وَسَخَّرَ الشَّمْسَ وَالْقَمَرَ

Diwani

وَسَخَّرَ الشَّمْسَ وَالْقَمَرَ

Riqa

وَسَخَّرَ الشَّمْسَ وَالْقَمَرَ

Naskh

# CRULP Research - Language

- ▶ Corpus Development
- ▶ Computational Linguistic Applications
  - Spell Checker
  - Grammar Checker
  - Lexicon
  - English to Urdu Machine Translation

# CRULP Research - Speech

- ▶ Text to Speech Synthesis
- ▶ Automatic Speech Recognition

# Projects

- ▶ Nafees Font Family
- ▶ Urdu Localization Project
- ▶ Microsoft Spell Checker
- ▶ PAN Localization

# PAN Localization

A regional initiative to develop local language  
capacity in Asia

[www.PANL10n.net](http://www.PANL10n.net)

# PAN Localization Project

## ▶ Partnership

- PAN program of IDRC
- CRULP at NUCES

## ▶ Objectives

- Develop localization technology for Asian languages
- Develop human resource to develop and use localized computing
- Research into policy framework to develop local language computing

## ▶ Timelines

- January 2004 till December 2006

# PAN L10n Project

## ► Collaborations

1. BRAC University, Bangladesh
2. Department of IT, Ministry of Information and Communications, Bhutan
3. Khmer Computerization Committee, National ICT Development Agency, Cambodia
4. Science Technology and Environment Agency, Laos
5. Madan Puraskar Pustakalaya, Nepal
6. University of Colombo School of Computing, Sri Lanka
7. ...

# Salient PAN L10n Project Outputs

- ▶ Localization Technology
- ▶ Asian Localization Peer Support Network
- ▶ Bibliography of Asian Localization
- ▶ Who's Who of Asian Localization
- ▶ Multi-lingual Website: [www.PANL10n.net](http://www.PANL10n.net)
- ▶ Asian Localization Handbook

# Scope of Localization



# Localization

“enabling computing experience according to linguistic culture of the user”

# Localization Requirements

- ▶ Standards
- ▶ Basic Applications
- ▶ Intermediate Applications
- ▶ Advanced Applications
- ▶ Soft Issues

# Standards

- ▶ Character Set
- ▶ Keyboard/Keypad layout
- ▶ Locale
- ▶ Collation Sequence
- ▶ Terminology Translation
- ▶ Fonts (?)
- ▶ ...

# Basic Applications

- ▶ Character set encoding(s)
- ▶ Utility for converting among various encodings
- ▶ Keyboard/Keypad drivers
- ▶ Collation algorithm
- ▶ Local language interface
- ▶ Fonts for various devices
- ▶ ...

# Intermediate Applications

- ▶ Find/Replace utility
- ▶ Natural language processor/Bidirectional processor
- ▶ Lexicon
- ▶ Spell checker
- ▶ ...

# Advanced Applications

- ▶ Grammar checker
- ▶ Automatic speech recognition
- ▶ Text to speech system
- ▶ Automatic machine translation
- ▶ Optical character recognition
- ▶ Handwriting recognition
- ▶ Speech to speech translation
- ▶ ...

# Introduction to Training

## ► Objectives

- Overview scope of localization
- Study in detail basic issues regarding localization standards and development
- Develop Asian peer support network

# Summary of Topics

- ▶ Encoding Standards
- ▶ Font Development
- ▶ Localization on Microsoft Platform
- ▶ Localization on Linux Platform
- ▶ Defining Normalization and Collation
- ▶ Overview Advanced Applications
- ▶ Overview Software Engineering

# Thank you

مرکز تحقیقات اردو