

NLP in Developing Asia

PAN Localization Project

www.PANL10n.net

Sarmad Hussain

PAN Localization Project

- A regional initiative to develop local language computing capacity in developing Asia 2004-2007
 - Objectives
 - Human Resources, Technology, Policy for Local Language Computing
 - Funded by IDRC, Canada
 - Coordinated by Center for Research in Urdu Language Processing, Pakistan
 - Afghanistan, Bangladesh, Bhutan, Cambodia, Laos, Nepal, and Sri Lanka
 - Myanmar, Iran, China, Thailand, India, Japan, Korea

PAN Localization Project

- Working on
 - Standards (encoding, KB, Collation)
 - Fonts and rendering, Collation, Terminology, Encoding Converters
 - Spell Checkers, Lexica, Line Breaking
 - Corpus Development, TTS, OCR
- Trainings
 - Two regional Trainings
 - Fundamentals of Local Language Computing (15 countries, 60 participants)
 - From Localization to Language Processing (12 countries, 45 participants)
 - National Trainings
 - Afghanistan, Bhutan, Cambodia, Laos, Nepal and Sri Lanka (1 week to 6 months)

AFNLP's Role in Promotion of NLP in Developing Asia

- Gap between AFNLP and Developing Asia
 - Very limited representation of developing Asian countries at AFNLP
 - Research presented is far too advanced compared to work going on in developing countries
 - Only token participation of researchers in Developing Asia at IJCNLP

AFNLP Membership

- Most developing countries do not have NLP associations. At best they have one group of people working in NLP. What could be the mechanism for these countries to have representation at AFNLP?

AFNLP in Promoting NLP in Developing Asia

- Training
 - Asian Summer School in NLP
 - Asian Master's Program
- Scholarships for students from developing countries to these programs

Accessibility of AFNLP

- SIGs or sub-regional Chapters of AFNLP?
 - SIG on South Asian Languages?
 - Easier for more researchers to travel within a sub-region
 - More localized on language specific issues in addition to algorithms

Broader base for AFNLP in Asia