www.panl10n.net

Research Capacity Building Through PANL10n

2004 – 2010

Sana Shams

Center for Language Engineering (CLE)

Al-Khawarizimi Institute of Computer Science (KICS)

University of Engineering & Technology (UET)

www.cle.org.pk

Objectives of PAN Localization Project

❑ 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

Evaluation Question

❑ To what extent PAN Localization project contributed in building Research capacity for local language computing in the country partner institutions (CPI's)?

Research Capacity Building

❑ Capacity Building within the context of Research is enhancing the abilities of individuals, organizations and systems to undertake and disseminate high quality research efficiently and effectively (Department for International Development, 2010).

❑ "Capacity building is a process whereby people are enabled to better perform defined functions either as individuals, through improved technical skills and or professional understanding, or as groups aligning their activities to achieve common purpose" (Breen. et.al., 2004).

Research Capacity Building (RCB) Framework

From literature (Cooke 2005, Neilson and Lusthaus 2007, Wignaraja 2009) RCB frameworks are organized into
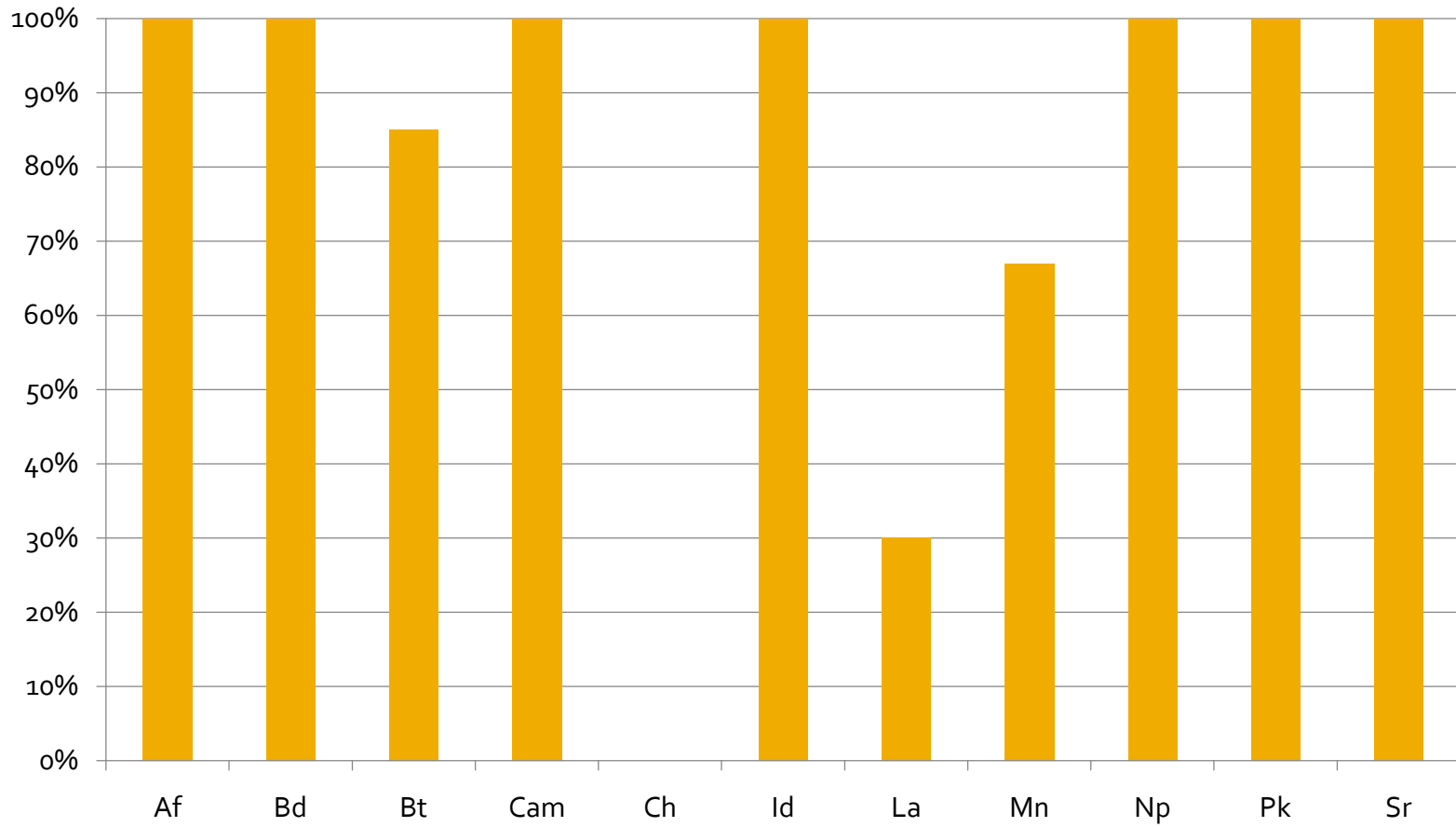
- Structural Levels
  - Individual
  - Organizational
  - System

- Principles of Capacity Building
  1. Skill Development
  2. Close to Practice Research
  3. Development of linkages
  4. Dissemination and impact
  5. Sustainability and Continuity
  6. Infrastructure development

1 . Skill Development

❑ Research skill development increases research activity

❑ Skills development has been focused on training researchers to conduct and publish research on local language computing

❑ *Indicators* used to assess skill development
1. Completion of project's software deliverables
2. No. of Research Publications

Completion of Project's Software Deliverables

Phase I: Completion of Localized Software

Countries	Languages	Localized software
Afghanistan	Pashto	Keyboard, Font
Bangladesh	Bangla	Bangla Pad, OCR, Lexicon, Spell checker, Collator
Bhutan	Dzongkha	Fonts, Dzonghalinux
Cambodia	Khmer	Spell Checker XP & Vista, Line breaker, Unicode Standardization, Collation, Lexicon, Word-Wrap utility, Sorting utility
Laos	Lao	Fonts, Pad, Keyboard
Nepal	Nepali	Nepalinux 1.0 & 2.0, Dictionary , Lexicon
Pakistan	Urdu	NVU Web Development Tool, SeaMonkey
Sri Lanka	Sinhala and Tamil	Unicode converter, OCR system, Text to Speech System

Phase II: Completion of Localized Software

Countries	Languages	Localized Software
Afghanistan	Pashto	SeaMonkey, Character Set for IDNs
Bangladesh	Bangla	OCR, TTS
Bhutan	Dzongkha	Dzongkha Linux
Cambodia	Khmer	Text to Speech System, OCR, Encoding Conversion, Line Breaking, Collation, Spell Checking, Find and Replace, SMS J2Me Application, Open Office Writer Plugin, PLC Typing Tutor
Indonesia	Bahasa Indonesia	SMT, Part of Speech Tagger
Laos	Lao	OCR, Line Breaking &Collation fro Open Office and Microsoft Office, Corpus Analysis Tool,
Mongolia	Mongolian	Localized Chat Client, SeaMonkey, Spell Checker, POS Tagger and Corpus Bilder, 3K Word ASR for Mongolian
Nepal	Nepali	Grammar Checker, Spell Checker, OCR, NepaLinux 3.0
Pakistan	Urdu	Email Client, Internet Browser, Website Development Tool, Online Stemmer, Machine Translation System, Part of Speech Tagger, Text Normalization Utility, Spell Checker, OpenOffice.org Suite, Psi Chat Tool
Sri Lanka	Sinhala and Tamil	TM Application, Tamil Language Learning Tool

Skill Development

- Advancing Localization Research Capacity

	Basic		Inter-mediate		Advanced	
	Ph 1	Ph 2	Ph 1	Ph 2	Ph 1	Ph 2
Afghanistan	*	*	*	*		
Bangladesh	*		*	*	*	*
Bhutan	*	*	*	*		*
China		*		*		*
Cambodia	*	*	*	*		*
Indonesia						
Laos	*		*	*		*
Mongolia		*		*		*
Nepal			*			*
Pakistan		*		*		*
Sri Lanka	*		*	*	*	*

Publication of Research Papers

Country Team	No. of Papers	Focus of the Publication
Bangladesh	8	MT, Script & Speech Processing
Bhutan	1	TTS
Indonesia	2	SLP, MT, POS
Mongolia	6	POS, Corpus, Speech
Nepal	1	NLP
Pakistan	5	IDN, POS, M&E
Sri Lanka	10	MT, Lexicon, Speech, IDN

Skill Development

Capacity Building Target Area	Year 2007	Year 2009
Project development	2	5
Project design	3	5
Problem identification	3	5
Project implementation	3	5
Analysis Ability	3	4
Communication Ability	2	4
Multi disciplinary research Ability	2	4
Quantitative analytical skills	2	4
Qualitative analytical skills	2	4

Self assessment of county project leaders, regarding their team's capacity over the years

Skill Development – Strategies Used

Training Strategy	Phase 1 (2004-07)	Phase 2 (2007-10)
Short Term Training	6	-
Mentor Placement (I)	3	2
Mentor Placement (II)	1	5
Summer School	1	-
Support for Higher Studies	5	9
Conference Participation	12	40

Training on localization and Khmer Language Processing, Cambodia, 2004

Training on Phonetics, Sri Lanka ,2004

Training on Computing for Localization, Laos, 2005

Workshop on IDNs for Pakistan Languages, 2008

2. Close to Practice Research

❏ A foremost principle of RCB is in directing researcher's ability to produce research that is useful for practice

❏ As defined, the 'ultimate goal' of research capacity development as the generation and application of new knowledge

❏ There is strong support that 'useful' research is that which is conducted 'close' to practice by generating research knowledge that is relevant to service user and practice concerns

Close to Practice Research

- ❑ End User Training and Content Development

- ❑ Build partnerships across technically and socially oriented organizations

- ❑ Publish research focused on dissemination and content development

Close to Practice Research

Country	Trainees	Trainer	Content
Bhutan	Govt. Officials, Private Sector	Govt./Private Sector	Govt.
Bangladesh	Rural Population	Infomediaries	Partner NGO
Cambodia	Govt. Officials, Teachers	Govt.	Govt.
China	Farmers in TAR	Govt.	Govt.
Nepal	Women, Teachers, Farmers	Partner NGO	Community
Pakistan	Students, Teachers	University	Students, Teachers
Sri Lanka	Blind Children	University	

Development of Linkages

❑ Research groups often operate in isolation, limiting the scope and success of their work. Thus in order to enhance the capacity, resources must be appropriately linked up and connected with active groups working on similar initiatives for robust and collaborative learning.

❑ It is the mechanism by which research skills, and practice knowledge is exchanged, developed and enhanced

❑ *Indicators :*
1. No. of formal organizational collaboration
2. Online research networks

Development of Linkages

- Inter-Disciplinary collaborations within teams
    - Computer Scientists
    - Linguists
    - Social Scientists

- Inter-Disciplinary Collaborations Across Teams
    - Universities
    - Non-Governmental Organizations
    - Language\Technical Standardization Authorities
    - Relevant IT and Language Ministries

- Regional and International Collaborations
    - Organization of regional training
    - Participation in regional conferences and workshops

- Online Research Networks (11 researchers at the beginning of Phase I and 110 researchers by the end of Phase II)

Dissemination

❑ Dissemination of research, through peer reviewed publications and presentations at academic conferences, is essential for sharing knowledge (Harris 2004, Breen et al 2004).
❑ Capacity building for wider research dissemination incorporates instruments of publicity through factsheets, the media and the Internet (Cooke 2005) for a variety of stakeholders, including public, policy makers and the relevant research community

❑ **Indicators**
1. Development of a local project website
2. Organization of awareness seminars
3. Creation of promotional materials
4. Participation in workshops and conferences

Development of a Local Project Website

Multilingual website www.panl1on.net

Afghanistan

Bangladesh

Bhutan

Cambodia

Indonesia

Laos

Mongolia

Nepal

Pakistan

Sri Lanka

Awareness Seminars

Country Component	Number of Total Events/Seminars
Afghanistan	1
Bangladesh	2
Bhutan	2
Cambodia	1
Indonesia	1
Laos	1
Mongolia	1
Nepal	1
Pakistan	4

Awareness seminar on Localization,
Afghanistan, 2006

TTS Launching Seminar at BRAC University,
Bangladesh, 2009

Promotional Material

❑ Distribution of CDs/DVDs containing project outputs like NepaLinux, Dzongkha Linux, LaoPad, BanglaPad, etc.

❑ Video of the project for global audience

Participation in Events and Competitions

❑ Presentation of the project at national and International forums

❑ NepaLinux-Prestigious international APC Chris Nicol FOSS Prize 2007

❑ Sinhala Text-to-Speech System-"Most Innovative Product" award at the Biennial Infotel Trade Exhibition 2008

Mr. Rafiqullah Kakar is receiving Manthan Award South Asia Pashto SeaMonkey in 2008

Professor Mumit Khan is receiving BASIS IT Innovation Search award for Bangla TTS in 2010

Professor Mumit Khan is receiving e-Content & ICT
for Development award for software Katha ,2010

Infrastructure Development

❑ Infrastructure as a set of structures and process that are set up to effective running of research project

❑ These include availability of technical resources including equipment, books, connectivity, etc. as well as sound academic and managerial leadership and support for developing and sustaining research capacity

❑ *Indicators*
1. Acquisition of academic resources
2. Procurement of equipments
3. Provision for Operating expenses

Infrastructure Development

- Specialized Research Centers
- Equipment
- Software
- Linguistic Resources
- Books & Journals
- Recurring Admin Expenses

Sustainability and continuity

❑ Long term sustainable capacity development requires consolidation of local system and process

❑ *Indicators:*
1. Organizational skill development
2. No of trained recourses in the different domains of localization

Sustainability and continuity

- Indigenous Localization Research Capacity development

  - *Management*
  - *Technical*
  - *Linguistics*
  - *Social Sciences*

	Mgmt	Tech	Lang	Social Sc.	
Af	2	2	3	0	7
Bd	3	10	3	6	22
Bt	1	8	4	0	13
Kh	7	39	13	3	62
Cn	4	11	3	0	18
Id	3	3	12	0	18
La	5	9	6	3	23
Mn	3	15	9	0	27
Np	6	12	4	9	31
Pk	6	27	7	4	44
Sl	3	7	3	1	14
	43	143	67	26	279

Sustainability & Continuity

■ Advancing
Localization
Research Capacity

	Basic		Inter-mediate		Advanced	
	Ph 1	Ph 2	Ph 1	Ph 2	Ph 1	Ph 2
Afghanistan	*	*		*		
Bangladesh	*		*	*	*	*
Bhutan	*	*	*	*		*
China		*		*		*
Cambodia	*	*	*	*		*
Indonesia						
Laos	*		*	*		*
Mongolia		*		*		*
Nepal			*			*
Pakistan		*		*		*
Sri Lanka	*		*	*	*	*

Sustainability & Continuity – Research Centers

- Center for Research in Bangla Language Processing, BRAC Univ., Bangladesh
- PAN Cambodia, Cambodia
- Language Technology Research Center, UCSC, Sri Lanka
- R&D Division, Department of IT, Bhutan
- Language Research Group, National Agency for Science and Technology, Laos
- Language Technology Research Lab, National University of Mongolia
- Center for Language Engineering, University of Engineering and Technology, Pakistan
- Language Technology Kendra, Nepal

Sustainability & Continuity

PAN L10n Network

PAN L10n Multilingual Chair in Local Language Computing

❑ Permanent research chair for multilingual computing in Pakistan

❑ Funds by International development research center (IDRC) Canada

❑ Nurture and grow the network of localization researchers

Achievements

- ❑ Seven of eleven CPIs have successfully submitted the requisite localized software as per the contract

- ❑ During phase II 10 out of 11 countries very researching on advanced local language software

- ❑ Release of NepaLinux 1.0 took place within 11 months in December 2005

- ❑ Enrolment of researchers on the Support network rose from 11-110 at the end of Project's phase 1

- ❑ Each Country partner institute developed its local language website initiating the development to local language digital content

- ❑ Total team of researchers involved in Local language computing during phase 2 were **279**

- ❑ Sri Lanka produced the maximum research publications during the phase2

Roadmap for Maximal Impact

- Step 1:
    - Skill Development
    - Infrastructure Development
- Step 2:
    - Development of Linkages
    - Close to Practice Research
- Step 3:
    - Dissemination & Impact
    - Sustainability & Continuity

Conclusion

❑ UN-APCICT/ESCAP (2010) note a steep demand for localization skills in ICT professional with the increasing ICT diffusion in the Asia Pacific

❑ Capacity building to conduct localization research must be a national and regional priority to bridge the gap

Thank You