

PAN Localization Project Phase II

I. Background

PAN Localization: A Regional Initiative to Develop Local Language Computing Capacity in Asia in partnership with National University of Computer and Emerging Sciences (NUCES), Pakistan. This project is funded by the International Development Research Center (IDRC), Canada. The first phase of the PAN Localization was initiated in 2004. The second phase was initiated in 2007

II. Objectives

The overall objectives of the research project are as follows:

- a) Examine effective means to develop digital literacy through the use of local language computing and content
- b) Explore development of sustainable human resource capacity for R&D in local language computing as a means to raise current levels of technological support for Asian languages
- c) Advance policy for development and use of local language computing and content
- d) Study and develop coherent instruments to gauge the effectiveness of multi-disciplinary research concerning the adoption of local language technology by rural communities

The specific objectives of the Project are as follows:

1. Investigate challenges and solutions to promote Use of multilingual ICTs
 - a. Research into issues related to content development and deployment
 - I. Investigate methods to determine end-user local language content requirements
 - II. Investigate challenges associated with development of local language content
 - III. Experiment with technological and social solutions for deployment of local language content
 - IV. Define conventions used for publishing content in local languages and assess and improve related standards
 - b. Research into development and conduct of training to promote use
 - I. Develop and trial reuse of training materials in local languages for different end-user groups in different countries
 - II. Conduct end-user training to determine effective strategies to retrieve content

III. Conduct end-user training to determine effective strategies to generate culturally-relevant content in local languages

2. Research into effective solutions to mature multilingual **Access** to ICTs

- a. Continue to develop standards to enable local language computing
- b. Continue to develop language resources, computing and frameworks for reuse, and research into localization of emerging mobile and open source platforms
- c. Research into localized and multi-modal easy to use interfaces to ICTs for a variety of end-user groups
- d. Continue to develop human resources to conduct research and development in local language computing and its impact on ICT access in Asia

3. Promote culture and policy for local language computing across Asia

- a. Support and expand the existing regional network of researchers, practitioners and policy-makers for collaborative learning in local language computing
- b. Survey and research into IPR issues related to software development, software localization and content
- c. Explore gender issues related to software development, software localization, training and content development
- d. Evolve a policy framework for local language computing and local language content
- e. Contribute to the state-of-practice in local language computing through a rigorous research publication program

4. Use and evaluate of OM and GEM II methods for project appraisal for this multidimensional and multi-stakeholder initiative

- a. Develop capacity in regional and country components to use OM and GEM II methodologies for project planning, monitoring and evaluation
- b. Use and adapt Outcome Mapping framework for localization related projects
- c. Apply Gender Evaluation Methodology within the framework of OM within the localization domain to determine (i) the effectiveness of GEM II in localization domain, and (ii) find out how GEM II and OM can jointly be used within a project, adding to the state of practice and the body of knowledge in evaluative methods

III. Administering Institutions

The administering institutions for the project are the Department of Budget, Finance and Department of Information Technology, Ministry of Information and Communications, Thimphu.

IV. Project Duration

The second PAN Localization project is for three years starting from 2007 to 2010.

V. Deliverables

| Phases | Milestones/Deliverables |
|---|---|
| Phase 1.1 May 2007 till Sept 2007 | <ol style="list-style-type: none"> 1. Initial Translations of gTLDs and ccTLDs in Dzongkha 2. Training on Linux Installation in Dzongkha 3. Training on Linux Operating System in Dzongkha 4. Report on Meeting of Project Country Partners with Policy Makers 5. OM and GEM II Analysis for the Country |
| Phase 1.2 October 2007 till March 2008 | <ol style="list-style-type: none"> 1. Dzongkha Natural Language Processor for TTS 2. Research Report on Dzongkha NLP Engine for TTS 3. Research Report on Character Set and Encoding Constraints for IDN in Dzongkha 4. Reviewed List of gTLDs and ccTLDs in Dzongkha 5. Training on Open Office 6. Training on Mozilla FF and TB 7. Training Conduction and Survey Report on Linux Installation and Operations 8. Initial Survey to Baseline Dzongkha Linux Usage Capacity of End User Groups 9. Report on Meeting of Project Country Partners with Policy Makers 10. OM and GEM II Report |
| Phase 2.1 April 2008 till Sept 2008 | <ol style="list-style-type: none"> 1. Test Report on IDNs for Dzongkha 2. Research Report on Dzongkha Lexicon Design 3. Training on GIMP, Inskape and GNU Cash 4. Training Conduction and Survey Report on Open Office 5. Training Conduction and Survey Report on Internet Applications 6. Report on Meeting of Project Country Partners with Policy Makers 7. OM and GEM II Report |
| Phase 2.2 October 2008 till March 2009 | <ol style="list-style-type: none"> 1. Phonetic Module for Dzongkha TTS 2. Dzongkha Speech Corpus 3. Training Conduction and Survey Report on Other Applications |

| | |
|---|---|
| | <ul style="list-style-type: none"> 4. Comparative Survey Analysis on ICT Diffusion in Different User Groups 5. Report on Meeting of Project Country Partners with Policy Makers 6. OM and GEM II Report |
| <p>Phase 3.1 April 2009 till Sept 2009</p> | <ul style="list-style-type: none"> 1. Dzongkha TTS 2. Dzongkha Diphone Database 3. Research Report on Dzongkha TTS 4. Dzongkha Corpus 5. Updated Dzongkha Linux 3.0 with TTS 6. Report on Meeting of Project Country Partners with Policy Makers 7. OM and GEM II Case |
| <p>Phase 3.2 October 2009 till March 2010</p> | <ul style="list-style-type: none"> 1. Dzongkha Lexicon 2. Research Report on Dzongkha Lexicon 3. Research Report on Dzongkha Corpus Development 4. Report on Meeting of Project Country Partners with Policy Makers 5. Final Report |

VI. Work Plan




| ID | Task Name | Duration | Start | Finish | H1 | | | | | H2 | | | | | H3 | | | | | | |
|----|----------------------------------|---------------------|-------------------|--------------------|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|--|
| | | | | | Q-1 | Q1 | Q2 | Q3 | Q4 | Q5 | Q1 | Q2 | Q3 | Q4 | Q5 | Q1 | Q2 | Q3 | Q4 | Q5 | |
| 1 | Dzongkha POS Collection | 206.76 days? | Mon 8/4/08 | Tue 5/19/09 | | | | | | | | | | | | | | | | | |
| 2 | Study on Dzongkha grammar | 49.88 days? | Mon 8/4/08 | Mon 10/13/08 | | | | | | | | | | | | | | | | | |
| 3 | Study on Dzongkha Morphology | 39.88 days? | Wed 1/28/09 | Wed 3/25/09 | | | | | | | | | | | | | | | | | |
| 4 | POS Tagging | 14.88 days? | Wed 3/25/09 | Wed 4/15/09 | | | | | | | | | | | | | | | | | |
| 5 | Review POS with DDC | 24.88 days? | Wed 4/15/09 | Tue 5/19/09 | | | | | | | | | | | | | | | | | |
| 6 | Research report | 0 days | Tue 5/19/09 | Tue 5/19/09 | | | | | | | | | | | | | | | | | |
| 7 | Building Dzongkha Lexicon | 174.88 days? | Mon 8/4/08 | Mon 4/6/09 | | | | | | | | | | | | | | | | | |
| 8 | Translation | 46.13 days? | Mon 8/4/08 | Tue 10/7/08 | | | | | | | | | | | | | | | | | |
| 9 | Design | 49.88 days? | Mon 8/4/08 | Mon 10/13/08 | | | | | | | | | | | | | | | | | |
| 10 | Implement | 50 days? | Mon 10/13/08 | Mon 12/22/08 | | | | | | | | | | | | | | | | | |
| 11 | Testing | 60 days? | Mon 12/22/08 | Mon 3/16/09 | | | | | | | | | | | | | | | | | |
| 12 | Draft research Report | 15 days? | Mon 3/16/09 | Mon 4/6/09 | | | | | | | | | | | | | | | | | |
| 13 | Building Dzongkha Corpus | 114.64 days? | Mon 8/4/08 | Fri 1/9/09 | | | | | | | | | | | | | | | | | |
| 14 | Digital text collection | 79.13 days? | Mon 8/4/08 | Fri 11/21/08 | | | | | | | | | | | | | | | | | |
| 15 | Design corpus database | 44.88 days? | Mon 8/4/08 | Mon 10/6/08 | | | | | | | | | | | | | | | | | |
| 16 | Implement | 34.88 days? | Mon 10/6/08 | Fri 11/21/08 | | | | | | | | | | | | | | | | | |
| 17 | Integrate with applications | 24.88 days? | Fri 11/21/08 | Fri 12/26/08 | | | | | | | | | | | | | | | | | |
| 18 | Research report | 10 days? | Fri 12/26/08 | Fri 1/9/09 | | | | | | | | | | | | | | | | | |
| 19 | Dzongkha TTS | 206.76 days? | Mon 8/4/08 | Tue 5/19/09 | | | | | | | | | | | | | | | | | |
| 20 | Ex-country training | 64.88 days? | Mon 8/4/08 | Mon 11/3/08 | | | | | | | | | | | | | | | | | |
| 21 | Study on Festival | 42 days? | Mon 8/4/08 | Wed 10/1/08 | | | | | | | | | | | | | | | | | |
| 22 | Phone set definition | 25 days? | Wed 10/1/08 | Wed 11/5/08 | | | | | | | | | | | | | | | | | |
| 23 | Tokenization and Normalization | 25 days? | Wed 11/5/08 | Wed 12/10/08 | | | | | | | | | | | | | | | | | |
| 24 | Letter-to-sound conversion (G2P) | 19.88 days? | Wed 12/10/08 | Wed 1/7/09 | | | | | | | | | | | | | | | | | |
| 25 | Syllabification Algorithm | 20 days? | Wed 1/7/09 | Wed 2/4/09 | | | | | | | | | | | | | | | | | |
| 26 | Stress Assignment Algorithm | 15 days? | Wed 2/4/09 | Wed 2/25/09 | | | | | | | | | | | | | | | | | |
| 27 | Diphone Generation | 14.88 days? | Wed 2/25/09 | Tue 3/17/09 | | | | | | | | | | | | | | | | | |
| 28 | Integration | 15 days? | Tue 3/17/09 | Tue 4/7/09 | | | | | | | | | | | | | | | | | |
| 29 | Testing | 15 days? | Tue 4/7/09 | Tue 4/28/09 | | | | | | | | | | | | | | | | | |
| 30 | Research report | 15 days? | Tue 4/28/09 | Tue 5/19/09 | | | | | | | | | | | | | | | | | |










Project: PAN Project II - 2007-2010
Date: Wed 7/23/08

| | | | | | |
|----------|--|-----------------|--|--------------------|--|
| Task | | Milestone | | External Tasks | |
| Split | | Summary | | External Milestone | |
| Progress | | Project Summary | | Deadline | |

| ID | Task Name | Duration | Start | Finish | H1 | | | | | H2 | | | | | H3 | | | | | | |
|----|-----------------------------------|---------------------|--------------------|---------------------|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|--|
| | | | | | Q-1 | Q1 | Q2 | Q3 | Q4 | Q5 | Q1 | Q2 | Q3 | Q4 | Q5 | Q1 | Q2 | Q3 | Q4 | Q5 | |
| 31 | Dzongkha OCR | 146.88 days? | Mon 8/4/08 | Wed 2/25/09 | | | | | | | | | | | | | | | | | |
| 32 | Ex-country training | 32 days? | Mon 8/4/08 | Wed 9/17/08 | | | | | | | | | | | | | | | | | |
| 33 | Data collection | 19.88 days? | Mon 8/4/08 | Mon 9/1/08 | | | | | | | | | | | | | | | | | |
| 34 | Design OCR Engine | 17 days? | Mon 9/1/08 | Wed 9/24/08 | | | | | | | | | | | | | | | | | |
| 35 | Pre-processing methods | 24.88 days? | Mon 9/1/08 | Mon 10/6/08 | | | | | | | | | | | | | | | | | |
| 36 | BW image | 14.88 days? | Mon 9/1/08 | Mon 9/22/08 | | | | | | | | | | | | | | | | | |
| 37 | Grayscale image | 10 days? | Mon 9/22/08 | Mon 10/6/08 | | | | | | | | | | | | | | | | | |
| 38 | Feature Extraction | 63.88 days? | Wed 10/1/08 | Tue 12/30/08 | | | | | | | | | | | | | | | | | |
| 39 | thining | 9.88 days? | Wed 10/1/08 | Wed 10/15/08 | | | | | | | | | | | | | | | | | |
| 40 | edge detection | 10 days? | Wed 10/15/08 | Wed 10/29/08 | | | | | | | | | | | | | | | | | |
| 41 | skew detection | 10 days? | Wed 10/29/08 | Wed 11/12/08 | | | | | | | | | | | | | | | | | |
| 42 | chaincode | 10 days? | Wed 11/12/08 | Wed 11/26/08 | | | | | | | | | | | | | | | | | |
| 43 | Testing pre-processor | 12 days? | Wed 11/26/08 | Fri 12/12/08 | | | | | | | | | | | | | | | | | |
| 44 | pixel mapping, etc | 12 days? | Fri 12/12/08 | Tue 12/30/08 | | | | | | | | | | | | | | | | | |
| 45 | Classification | 69.88 days? | Wed 10/1/08 | Wed 1/7/09 | | | | | | | | | | | | | | | | | |
| 46 | Neural Network | 17 days? | Wed 10/1/08 | Fri 10/24/08 | | | | | | | | | | | | | | | | | |
| 47 | Hidden Markov Models | 17.88 days? | Fri 10/24/08 | Wed 11/19/08 | | | | | | | | | | | | | | | | | |
| 48 | K-Mean | 17 days? | Wed 11/19/08 | Fri 12/12/08 | | | | | | | | | | | | | | | | | |
| 49 | Fuzzy C-Mean | 18 days? | Fri 12/12/08 | Wed 1/7/09 | | | | | | | | | | | | | | | | | |
| 50 | Evaluation & Testing | 20 days? | Wed 1/7/09 | Wed 2/4/09 | | | | | | | | | | | | | | | | | |
| 51 | Research report | 15 days? | Wed 2/4/09 | Wed 2/25/09 | | | | | | | | | | | | | | | | | |
| 52 | Dzongkha IDN | 191.76 days? | Mon 8/4/08 | Tue 4/28/09 | | | | | | | | | | | | | | | | | |
| 53 | Character set encoding | 44.88 days? | Mon 8/4/08 | Mon 10/6/08 | | | | | | | | | | | | | | | | | |
| 54 | Constraints for IDN | 34.88 days? | Mon 10/6/08 | Fri 11/21/08 | | | | | | | | | | | | | | | | | |
| 55 | Testing IDN | 40 days? | Fri 11/21/08 | Fri 1/16/09 | | | | | | | | | | | | | | | | | |
| 56 | Implementation of IDN in ISP | 45 days? | Fri 1/16/09 | Fri 3/20/09 | | | | | | | | | | | | | | | | | |
| 57 | Test Report | 27 days? | Fri 3/20/09 | Tue 4/28/09 | | | | | | | | | | | | | | | | | |
| 58 | Training on Dzongkha Linux | 39.88 days? | Mon 8/4/08 | Mon 9/29/08 | | | | | | | | | | | | | | | | | |
| 59 | Training Open Office | 27 days? | Mon 8/4/08 | Wed 9/10/08 | | | | | | | | | | | | | | | | | |
| 60 | Training on FF and TB | 27 days? | Mon 8/4/08 | Wed 9/10/08 | | | | | | | | | | | | | | | | | |

| | | | | | | |
|--|----------|--|-----------------|--|--------------------|--|
| Project: PAN Project II - 2007-2010 Date: Wed 7/23/08 | Task | | Milestone | | External Tasks | |
| | Split | | Summary | | External Milestone | |
| | Progress | | Project Summary | | Deadline | |

| ID | Task Name | Duration | Start | Finish | H1 | | | | | H2 | | H3 |
|----|--|---------------------|-------------------|--------------------|-----|---|----|----|--|----|--|----|
| | | | | | Q-1 | Q1 | Q2 | Q3 | Q4 | Q5 | | |
| 61 | Compile feedback and draft report | 27 days? | Mon 8/4/08 | Wed 9/10/08 | |  do | | | | | | |
| 62 | Training on GIMP, Inskape and GNU C | 27 days? | Mon 8/4/08 | Wed 9/10/08 | |  do | | | | | | |
| 63 | Training report | 12.88 days? | Wed 9/10/08 | Mon 9/29/08 | |  do | | | | | | |
| 64 | Reports [to be submitted every six mont | 149.88 days? | Mon 8/4/08 | Fri 2/27/09 | |  | | | | | | |
| 65 | Report of Meeting with Project Partner | 52.88 days? | Mon 8/4/08 | Thu 10/16/08 | |  Pema/Tenzin | | | | | | |
| 66 | OM and GEM Analysis for Country | 54.75 days? | Mon 8/4/08 | Fri 2/20/09 | |  | | |  do | | | |
| 67 | Technical Progress Report | 57 days? | Mon 8/4/08 | Tue 2/24/09 | |  | | |  do | | | |
| 68 | Financial Progress Report | 50.88 days? | Mon 8/4/08 | Tue 2/24/09 | |  | | |  do | | | |
| 69 | Training Conduction and Survey Repo | 34.75 days? | Mon 8/4/08 | Fri 2/20/09 | |  | | |  do | | | |
| 70 | Training Conduction and Survey Repo | 39.75 days? | Mon 8/4/08 | Fri 2/27/09 | |  | | |  do | | | |

| | | | | | | |
|--|----------|---|-----------------|---|--------------------|---|
| Project: PAN Project II - 2007-2010 Date: Wed 7/23/08 | Task |  | Milestone |  | External Tasks |  |
| | Split |  | Summary |  | External Milestone |  |
| | Progress |  | Project Summary |  | Deadline |  |