

PAN LOCALIZATION

NIDA E-DOCUMENT PROJECT REPORT

Thursday, May 14, 2009

Taing Leang NGUON

Table of Content

PAN LOCALIZATION	1
NIDA E-DOCUMENT PROJECT REPORT	1
1 Content developed	2
1.1 Research and Interview	2
1.2 Project realization	3
1.3 Content developed by our content partners	5
2 Pan Localization Office Add on	6
3 Library.....	6
4 Issues.....	6
4.1 Writing skill	6
4.2 Team behavior	7
4.2.1 Cheating behavior	7
4.2.2 Irresponsible behavior.....	7
4.2.3 Inappropriate use of ITC.....	7
4.3 Technology issue	7
4.3.1 Khmer OCR not available.....	7
4.3.2 Office add on bug.....	8
4.3.3 Joomla Content Management System.....	8
5 Conclusion	8

Livelihood subject is chosen for content development for e-document in Cambodia. We want it to be really very helpful for most of Cambodian people.

1 Content developed

Cambodian content is insufficient for the need of the country. On the other hand, Khmer people are not really interested in reading. In big cities, people read more newspapers, get more informed.

Writing document is somehow creating new project. We use three ways to make content:

- Research and interview
- Running ourselves the project
- Content developed by our content partners.

1.1 Research and Interview

On these subjects, “pig breeding”, “duck breeding”, “chili plantation”, etc..., we go the see the people who have good experience on these subjects. They are farmers who get good benefit from these works. We interview the best way how to do. Interviews are recorded; photos and video of the place are taken.



Information obtained is checked with the scientific experts on these matters. We use their knowledge to strengthen our documents. In case we do not have any expert available, we use reference books or website as resource of our development.

Interview with Mr. Im Pheap, farmer who has successful practice on pig breeding.

1.2 Project realization

According to the survey in the rural area, some subjects are very interesting to develop in several versions. See the good impact of the “fish breeding in plastic bag” to the daily life of the population, we undertake ourselves the project of fish breeding. The Kraing Thom village, Tonle Baty district. Takoa province is chosen to do the fish breeding. Mr. Long Ratha, our content writer is in charge of this task because the breeding place belongs to his parent. It is located about 50Km from Phnom Penh and take 1 hour to go by car or motorbike.



Members of Fisf breeding in plastic bag, Kraing Thom village, Takoa province March 30 2009

This program allows us to have a document in several versions to distribute to the population: video, webpage and book.

This work allows us to make a video from the beginning to the end. Each step is shown chronologically in the documentation. The strategy in this program is to bring the document on the video because the video, Karaoke is very popular in the rural area. Cambodian people like to watch video than read book.



Ditch of 2 meter wide, 5 meter long and 1.5 meter deep.

The plastic bag is useful to prevent water to infiltrate into the earth. It preserves water for fishes.



Fishes came from the “fish breeding in plastic bag” of our neighbor, our own is not available yet.

1.3 Content developed by our content partners

Our target is to scatter knowledge to Cambodian people, as wide as possible. Content development by our own team is not simple and slow. Most of the writers and journalists enjoy to let us put their texts on our website. 70% of our content partners provide their text already typed in computer files. The remaining of our partners has books or documents in paper form that we cannot copy and paste directly to our website. In this case, we need the Khmer OCR that Pan Localization is developing.

We are working with 5 newspapers, 5 magazines and 7 organizations or writers.

List of our content partners

3	កាសែត ថ្ងៃ៧មករា	យល់ព្រម	File
4	កាសែត កម្ពុជបុត្រ	យល់ព្រម	File
5	ទស្សនាវដ្តី អាថកំបាំង	យល់ព្រម	File
6	ទស្សនាវដ្តី ឃឹស	យល់ព្រម	Send by mail
7	សហគមន៍ អ្នកនិពន្ធខ្មែរ	យល់ព្រម	Books or document
8	ព្រឹត្តាចារ្យ យូបូ	យល់ព្រម	Books
9	ទស្សនាវដ្តីពេញនិយម	យល់ព្រម	File
10	កាសែត តាព្រហ្ម	wait	
11	អ៊ុក សំអុល	យល់ព្រម	File
12	ទស្សនាវដ្តី ខេមរា	យល់ព្រម	File
13	ទស្សនាវដ្តីមរកត	ទើបជួប	
14	ទស្សនាវដ្តី ធម្មជាតិ	យល់ព្រម	File
15	លោក គឹម ចាន់ណា	យល់ព្រម	
16	ភិក្ខុញាណចារី ឆៃម ពេជ	យល់ព្រម	File
17	ព្រះមហា យុល គឹមហៃ	យល់ព្រម	Books
18	គុណ រតនា	យល់ព្រម	Books
19	លោក លន់ សុខឡែង	យល់ព្រម	File

2 Pan Localization Office Add on

We actively use the Microsoft Office add-on of Pan Localization. We really appreciate the work that Pan Localization team has done so far. NIDA E-DOCUMENT project make good use of this tool for producing Khmer content.

We get files from our content partners; the following steps perform for every text:

1. Convert Limon text to Khmer Unicode
2. Spelling check
3. Line breaking
4. Build the content of the webpage
5. Convert to pdf format file that can be download from the website.

3 Library

All collected texts, photos, videos and scanned documents are stored in the server computer in the office. All members have access to these resources that they can use to develop their content.

Paper documents are scanned waiting availability of Khmer OCR for use. We collected books that we get authorization from their author to reproduce on our website.

4 Issues

During the development of content, several issues arise. We are new in this domain that we thought that our team from ICT area is fine. Several parameters need to be considered to fulfill the task.

4.1 Writing skill

When writing document, our staff make text so difficult to be understood. They use sentences with bad use of preposition.

Despite of the spelling check tool provided by our colleagues of Pan Localization, some staff do not write words properly. The content must be checked by a second person to get it validated.

Khmer writing is influenced by several period of crisis in Cambodia. After the war, many bad spelling words get approval by Cambodian schools. Wrong spelling words that I personally taught become right for the young Cambodian today. The difficulty comes from the instability of the Khmer language spelling. Samdach Chhoun Nat dictionary seems to be reference for most of the people, but wrong spelling taught during their life at school still have bad influence for most of Khmer population.

4.2 Team behavior

Several attempts to hire technical people to write content are totally failed. They think that content is easy because they learned at school how to do. But in fact and in general, they do poor work if they do not work carefully.

4.2.1 Cheating behavior

More than other area, I think that content development needs to be done by motivated and creative people for writing. Our writers, researchers go out of office to meet with other people to get informed about the subject that they are developing.

With this kind of freedom, one person in our team uses it as pretext to be absent from working place and work for his own interest. When time comes to provide the result, he copied text from magazine and gives 10 pages to me. The cheating is discovered and he get fired.

4.2.2 Irresponsible behavior

When undertake a task, people work on it without thinking about good result. They all think that other people correct and control again and again their work. They act in the passive manner: wait and see. One writer leaves always his text with misspelling words counted 10 mistakes average per page. We corrected together the mistakes in one of his document, one month later; his other documents that contain spelling mistakes are still untouched without any correction.

Irresponsibility comes also with negligence behavior. When one word is misspelled, some staff corrected only this precise word, but the same mistaken works in other place of the document are still untouched.

4.2.3 Inappropriate use of ITC

Cambodian writers have limited knowledge on Information Technology, they use badly computer. The consequence is file lost, slow typing, and contaminating viruses.

On the other hand, for security and avoid lost, they copy and past files every where (as reproducing files is easy and without cost). It takes only one month to fill up 160 GBytes of hard drive.

4.3 Technology issue

4.3.1 Khmer OCR not available

A part from our own content, 30% of them come from our content partners who provide books or paper documents to us. In paper form, these documents are scanned and waiting the tool like OCR to convert to text. We cannot afford to re-type them manually.

4.3.2 Office add on bug

Office 2003 gets stuck when running these add on. It runs quite well on MS Office 2007.

Some tools do not run when we click on. They are just silence for a while. They seem to be very sensitive to viruses attack.

Slow processes on spelling check, line breaker. We just watch the cursor of the editor scan slowly from the beginning to the end of the text.

The end of paragraph in Khmer “្ក” is not converted correctly to pdf format. The pdf format goes well with the “Khmer OS” font.

4.3.3 Joomla Content Management System

Joomla is great open source software that we use to make our website. Expertise is Joomla is still difficult to find. We have trouble to find good expert in Joomla to build our website. We hired already 3 people so far for this CMS.

Open source software is great, but do not fit to Cambodia because lack of skill. We do not have experience programmers; we cannot support problems by our own capacity.

This issue makes our website launch behind schedule.

5 Conclusion

Despite of issues described above, EDOCUMENT project is progressing greatly. We get support from many content partners. We are very proud to provide to Cambodian people useful information totally in Khmer language. We gather people in the country; promote Khmer writing, archiving documents. We have search box that performs search in Khmer language.

Our effort is concentrate on collaborating with content partners who provide to us ton of text to publish. With newspaper and magazine partners, lot of articles is daily available for publication. We actually train 10 volunteer students who help us to publish everyday articles on the website when the time arrives.