

Learning from Example and By Doing Approach to Train Data Entry Operators at Universal Digital Library

Lavanya Prahallad, Vamsi Garimella and Praveen Garimella

MSIT Division, International Institute of Information Technology, Gachibowli,
Hyderabad 500032, India

Email: lavanyap@cmu.edu; stpvamsi@gmail.com; praveeng@cmu.edu

Abstract - One of the major works in the Universal digital library (UDL) is scanning and entering the metadata for the books. This metadata can be entered with the help of transliteration editors for Indian languages. Transliteration editors are essential for keying in the Indian languages with the help of QWERTY keyboard and the notation used is IT3. Learning the IT3 codes and entering the data is time consuming and error prone. In this paper we propose a new methodology to train the data entry operators at UDL, namely, *learning from example and by doing*. The paper discusses how this approach of training is faster, better and cheaper.

Index Terms — Learning from Example, Learning by Doing, IT3, Transliteration Editors, Training

I. INTRODUCTION

Several processes exist in realizing the concept of universal digital library (UDL). These processes include scanning of the books, improving the quality of scanned images, entry of meta-data of the scanned books, storage of the data, retrieval and access to the data as and when required. These processes could be standardized so that they could be replicated at several scanning centers throughout India and world. At the same time, to implement these processes in different centers, it requires training of the personnel to ensure quality implementation of these processes. However, the existing training procedures used in several scanning centers seems to be ad-hoc and ineffective. Often traditional approaches of demonstration and lecturing are followed to train the data entry operators at digital libraries. Some of the limitations of these paradigms including availability of trained personnel and motivation of trainees for self-discovery etc cripple the

process of effective training. To address these and other relevant issues in training like quality and cost of training, we propose applying the learning from examples and by doing approach to impart skills based training to the workers of digital libraries. Lot of research has been done to show how people can learn from worked-out examples and problem solving. Learning by doing offers an effective and faster ways of training the personnel. Development of such training modules is critical and acts as supplementary to the several standard UDL processes implemented at several of its scanning centers. In this paper, we explore the application of learning from examples and by doing paradigm to develop training module for workers using transliteration editors. The paper is organized as follows: Section 2 discusses the transliteration editors for Indian languages. Section 3 describes the paradigm of learning from examples and by doing (LFED). Section 4 applies the LFED paradigm to the transliteration editors. Section 5 presents the concluding remarks of the authors.

II. TRANSLITERATION EDITORS

Transliteration editors are essential for keying-in Indian language scripts into the computer using QWERTY keyboard. Applications of transliteration editors in the context of Universal Digital Library (UDL) include entry of meta-data and dictionaries for Indian languages. The issues in building transliteration editors include design of a user-friendly and readable transliteration scheme, user interface to key-in the text and have the text rendered in native script, provide quick and easier help to the user to find the transliteration code for a character in Indian languages. A transliteration scheme referred to as

IT3 developed by IISc Bangalore and Carnegie Mellon University is as shown in Fig 1.

| | Consonants | | | | |
|-----------|------------|-----|----|-----|--|
| IT3 | ka | kha | ga | gha | |
| Sanskrit | क | ख | ग | घ | |
| Hindi | क | ख | ग | घ | |
| Marathi | क | ख | ग | घ | |
| Telugu | క | ఖ | గ | ఘ | |
| Kannada | ಕ | ಖ | ಗ | ಘ | |
| Malayalam | ക | ഖ | ഗ | ഘ | |
| Tamil | க | க | க | க | |
| Assamese | ক | খ | গ | ঘ | |
| Bengali | ক | খ | গ | ঘ | |

Fig.1 A section of the transliteration scheme

This transliteration scheme considers user readability aspect and is used to build 'OM' transliteration editors and its updated versions (Madhavi, 2005; Lavanya, 2005;). These editors are put to use to key-in the Indian languages books, short stories and meta-data in the context of digital libraries.

However, a user needs to learn the mapping of a character in Indian language to its corresponding IT3 transliteration code to use the transliteration editor effectively. In the present training workshops the data entry operators try to memorize the mapping table or lookup the help-pages of the editor. This is time consuming and it typically takes about 2-3 days for a person to memorize the IT3 codes. During the learning process the user also introduces several errors in the keyed-in data thus generating a poor quality output. This further leads to lot of time being invested in proof-reading the text.

It is essential to reduce the learning time and minimize the errors for faster adaptability of such transliteration editors to improve quality of transliterated text. The paper discusses the framework of LFED and illustrates how it can be used to create effective training modules for the workers at digital libraries.

III. LEARNING FROM EXAMPLES AND BY DOING (LFED) PARADIGM

For effective training using LFED, there are two prerequisites:

1. Examples and problems should be sufficiently rich in number and variety

to encourage highly redundant learning of interrelated skills.

2. Students should be given practice in explaining their problem solving processes after they have found the problem solution. [1]

The workflow (as seen by a user of the LFED training module) of LFED includes:

- o Skills imparted by demonstrating examples/scenarios
- o Trainees internalize skills by practicing similar examples/scenarios
- o Instant feedback and just in time lessons are provided while trainees practice examples

To author LFED based training modules, a one time set up of LFED based learning environment is sufficient. This environment includes:

- o LFED authoring tool
 - o Screen capture tool
 - o Examples authoring tool
 - o Workflow authoring tool

Once the environment is setup the tools can be used to create multimedia based interactive lessons that can be hosted on a training website or in a learning management system. The trainees can access the LFED based modules anytime from a central server. The need for training workshop itself is eliminated and the LFED based module can be used to train the workers irrespective of where they are. The requirements/tasks can be communicated to the workers via the LFED environment and the output is uploaded by the workers to a designated server. This process is hassle free giving the users liberty to work at their convenience from anywhere, anytime and seek help from the just in time lessons (JITL). The training is delivered via example based JITLs and workers learn by doing the practice sheets (sample text to be transliterated) and referring back to the corresponding help files/videos in case of doubts. An instant feedback indicates the number of errors and tips on improvement. The workers can also communicate via voip/Instant message with a human instructor if needs to be.

IV. LFED FOR TRANSLITERATION EDITORS

The workers at digital library center, India, use a tool called 'UniTrans' to manually enter the text from the scanned books of different languages. They use UniTrans to key in the book in English and the tool transliterates the text to the selected

native language. The input text (in English) is saved in ANSI format as a text file using UniTrans while the transliterated text is copy-pasted into a text file and is saved in Unicode format. These text files are later proof-read and the defect free files are used to generate the HTML version of the scanned books.

Screen shot of the “Unitrans” is shown in Fig 2.

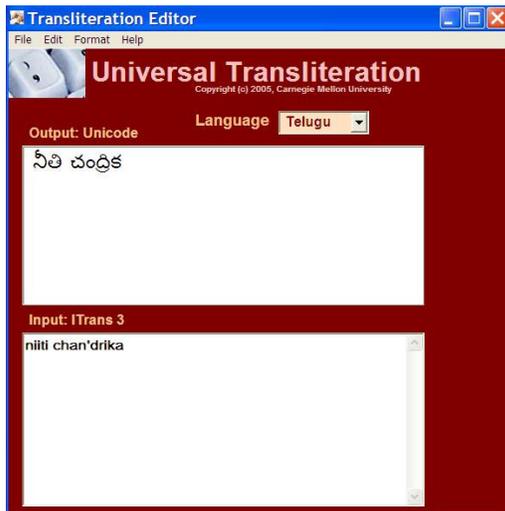


Fig 2 Screen shot of Unitrans

Once the software is installed and the prerequisites to run the software are set up, the Data Entry Operators (DEOs) are trained in a one day workshop on data entry in IT3 using Unitrans. In the first half of the workshop DEOs get familiar with the complex mapping table of the language that needs to be transliterated with English and a human instructor demonstrates the tool usage. Most often the human instructor ends up giving multiple demos to the individuals as they have doubts in using the system or could not memorize the sequence of steps. It's been found that most of the questions are related to saving the files in multiple formats and starting with a new page. Typically after having keyed in 4-5 pages the workers are familiarized with the process and the instructor has minimal or no questions from them. Typically in this workshop all the participants key in the same book and based on the number of error free pages transliterated, selected people are given the job of transliterating other books. This turns out to be an inefficient way to train the workers as it involves overhead of instruction time, instructor's time, Trainer availability and memorizing the transliteration schemes and steps etc.

To overcome these concerns at digital library center, Hyderabad, India, LFED based training has been experimented with. The LFED based modules are created using a screen capture tool called Camtasia that captures the demonstration of UniTrans by a trainer. The trainer adds voice narration to describe the steps in using the tools and Macromedia Flash has been used to make the modules interactive. These are the Just in Time lessons (short videos describing specific concepts). The JITL, along with the requirements list, resources, instructions and other resources are hosted on a server containing an open source learning management system, Moodle. Though Moodle is intranet based, LFED based training modules can also be hosted on web. The workers found it pretty easy to access the Moodle based system and among other things, it was easy for the trainers to authenticate each trainee, monitor their progress, generate visual reports indicating the number of scanned books and pages by each worker, give feedback online and communicate effectively. The workers also were enthused by the ease of use, clarity in process and effective work and information flow.

V. CONCLUSIONS

The paper presents the limitations of current training methodologies, and elucidates how LFED provides a better alternative to train the workers at digital library centers, in usage of transliteration editors.

VI. ACKNOWLEDGEMENT

We express our profound gratitude to Prof. Raj Reddy, Carnegie Mellon University, for his motivation and guidance.

VII. REFERENCES

- [1] Zhu, Zhu, Lee, Simon “Cognitive Theory To Guide Curriculum Design for Learning from Examples and by Doing” *Jl. of Computers in Mathematics and Science Teaching*, 2003, 22(4), 285-322.
- [2] Prahallad Lavanya, Prahallad Kishore, Ganapatiraju Madhavi, “A simple approach for building transliteration editors for Indian languages”, Oct 2005
- [3] Xinming Zhu, and Herbert A. Simon “Learning Mathematics from Examples and by doing”, 16th July, 1987
- [4] Ganapathiraju Madhavi, Balakrishnan Mini, Balakrishnan N, Reddy Raj, “Om: One tool for

many (Indian) languages”, *Journal of Zhejiang University SCIENCE*, Vol 6A, No. 11, pp 1348-1353, Oct 2005

- [5] Vamshi Ambati, N.Balakrishnan, Raj Reddy, Lakshmi Pratha, C V Jawahar: The Digital Library of India Project: Process, Policies and Architecture, In the Proceedings of 2nd International Conference on Digital Libraries(ICDL), 2006
- [6] <http://www.ulib.org>