Standardization Activities and Open Source Movements in Thailand[1]

Theppitak Karoonboonyanan, Thaweesak Koanantakool

National Electronics and Computer Technology Center
National Science and Technology Development Agency
Ministry of Science Technology and Environment, Thailand.
theppitak@nectec.or.th, htk@nectec.or.th

ABSTRACT — Standardization plays an important role in settling interoperability problems. This paper presents a survey on activities relating to IT standardization during the recent years in Thailand. Character sets, internationalization, font metrics, and studies on Tai scripts are discussed. In addition, the open source movements provide the stage for the standards to be realized and are beneficial to the promotion of the standards, as summarized in the paper.

บทคัดย่อ – การกำหนดมาตรฐานมีบทบาทสำคัญต่อการทำงานเข้ากันได้ของระบบต่างๆ บทความนี้นำเสนอบทสำรวจกิจกรรมต่างๆ ที่เกี่ยวข้องกับการกำหนดมาตรฐานเทคโนโลยีสารสนเทศในประเทศไทย อันประกอบด้วย รหัสอักขระ การทำให้เป็นสากล (internationalization) สัดส่วนขนาดของฟอนต์ภาษาไทย และการศึกษาตัวอักษรตระกูลไท นอกจากนี้ ยังกล่าวถึงความเคลื่อนไหวของซอฟต์แวร์ต้นรหัสเปิด (open source software) ซึ่งสามารถเป็นเวทีสำหรับการสร้างซอฟต์แวร์ที่ตรงตามมาตรฐาน และในขณะเดียวกัน ก็มีผลต่อการรณรงค์การใช้มาตรฐานเองอีกด้วย

0. Introduction

Standardization of IT in Thailand was recognized since 1984, when there were more than 26 sets of character codes were in use [1]. Two years later, an agreed standard code for Thai language was announced as a Thai Industrial Standard, TIS 620-2529/1986. However, at that time, only the codes were standardized. The input/output systems for computer processing [2] have not yet been unified. Operating systems and applications have been localized individually, based on different conventions. The proprietary standard that gains the lion's share in the market becomes *de facto*, no matter how its enhancement makes it deviated from industrial standards. Interoperability problem is therefore inevitable, especially in the age that different systems are connected through the Internet. Hence, standardization plays an important role in moderating the plethora of practices.

Recently, the open source paradigm has been widespread, and has become another model for software development. The openness of the source code also gives the chance to control the conformance to the standards of the software, as well as the satisfaction to users' needs.

To shape consistent language support technology in the country, standardization

1 Published in the Proceedings of MLIT-4 : Fourth International Symposium on Standardization of Multilingual Information Technology, October 27-28, 1999, Yangon, Union of Myanmar. It is sponsored by CICC/MITI/MCF. Contact person of CICC is Mr. Takayuki Sato (sato@net.cicc.or.jp).

activities and responses to open source movements are thus indispensable, and will be described in this paper.

1. Character Sets

The national standard character set for use in computers is TIS 620-2533/1990, from which several character sets are derived, for example, IBM code page 874 (cp-874), Microsoft code page 874 (windows-874) and Apple Thai (MacThai) [3]. These character sets are widely adopted in proprietary software, causing conflicts in communication among different platforms in the Internet.

Ironically, it's the game of the name. Only TIS 620 common characters are exchanged in practice, with different code set labels. The response to the code set with "unknown" name depends on applications. Some ignore the code set and process the text with their default preferences, while others simply reject.

Ad hoc solutions are also ubiquitous, such as using "iso-8859-1" or "x-user-defined" code name for Thai E-mails and web sites, by which Thai message could pass through the hole to the receiver in some weak situations. But that is not always the case.

In September 1998, the "tis-620" MIME character set has been registered by Trin Tantsetthi [4] with the Internet Assigned Number Authority (IANA) of the Internet Engineering Taskforce (IETF). A campaign has been set up by a group of developers [4] [5] to promote the use of the new standard MIME character set.

In 1999, the international standard ISO/IEC 8859-11 Latin/Thai characters has been reactivated by the ISO/IEC JTC1/SC2/WG2, and is becoming another potential choice of the standard MIME character set. When applied, "tis-620" and "iso-8859-11" are likely to be aliases to each other.

For multilingual documents, "utf-8" [6] is another possible alternative encoding. Nonetheless, the lack of UTF-8 editor is still the problem.

2. Internationalization

The third edition of ISO/IEC 14651 International String Ordering [7] has included an informative annex describing Thai string ordering. And, hopefully, the ordering of Thai in the standard would be satisfactory for Thai users.

A principle for Thai string ordering in detail has been proposed by a group of developers [8], and, as a consequence, the LC_COLLATE category of POSIX locale has been defined, as well as the other categories in a later time [9].

With the cooperation with the GNU C library project, the drafted POSIX locale has been made effective with glibc 2.1.1, which is used in modern distributions of Linux operating systems, such as Red Hat 6.0. Applications known to be internationalized and reflect the Thai locale include Linux 'date' and 'cal' commands, GNOME calendar, GNOME panel clock, KDE panel clock, and Perl 5.

3. Fonts

Thai fonts currently available in the market are designed based on Roman font metrics. This is not appropriate for Thai glyphs, since Thai characters are written in 4 levels. As a result, Thai glyphs are usually compressed to accommodate space for the 4 levels, and look smaller than Roman letters with the same point size.

The National Electronics and Computer Technology Center (NECTEC) therefore set up a committee for drafting the standard metrics for Thai glyphs relative to Roman and for creating prototype fonts to be used in public domain.

Three public-domain fonts, knowned as National Fonts (NF) 1, 2 and 3, are now available to the public. They are aimed to be the default fonts available in every platform. NF1 and NF3 are serif fonts. NF2 is sans serif. NF4 is planned for a "calligraphic" model font and NF5 is planned for a "handwriting" model font. Within December 1999, the official names of these fonts will be announced as part of the celebration of the 6th cycle anniversary (72nd birthday) of His Majesty The King of Thailand.

4. Tai Scripts Studies

Thai language used in the central Thailand belongs to the *Tai* language family. The scripts belonging to the family have caught the interests from a group of standardization committees. For example, New *Tai Lue* and *Tai Dam* scripts have been proposed to be encoded in the ISO/IEC 10646-1 character set.

In Thailand, Mr. Thawee Sawangpanyangkoon has done a research on *Tai* scripts and has created TrueType fonts for 13 *Tai* scripts, through the funding of the Thailand Research Fund (TRF).

We expect that more efforts will be made in the study of unification of these scripts with Thai scripts.

5. Open Source Movements

Several developers in Thailand have adopted the philosophy of open-source software in their works and have joined the world in this movement. Linux, the free OS of Linus Torvald, has become popular in Thailand and many developers have joined together in boosting the use of Thai language in the OS, with X Window as the GUI environment.

5.1 Distributions

There are currently four local Linux distributions in Thailand: Kaiwal Linux by Kaiwal Software, Linux School Internet Server (Linux SIS) and Linux with Thai Language Extension (Linux-TLE) by the National Electronics and Computer Technology Center (NECTEC), and Burapha Linux by Burapha University. These distribution developers meet regularly and join in regular Linux/Open-Source Symposia. It is expected that some distributions may merge in the new releases.

5.2 Development Projects

Several efforts are made to enable Thai language in open-source applications. Here are some examples:

1. **NACSIS-Thai Project** [10] is probably the first effort to support Thai on various platforms that are not Thai-localized.
2. **ZzzThai** [11] is another project to enable Thai in operating systems and applications on various platforms.
3. **Thai Linux Working Group** [12] is a Thai developer community concentrating on developments on Linux.
4. **WindowMaker** [13] is a GNU window manager project based on GNUStep. The Thai XKB with language mode locking allows user to input Thai characters conveniently. A Thai developer has also been one of the development team.
5. **Mozilla** [14] is an open-source project set up by Netscape Communication Co., Ltd., by opening the source code of its browser and other components. Three Thai developers have contributed the Thai language support to the browser [15]. Mozilla now can recognize the "tis-620" MIME character set, and can wrap Thai text lines appropriately.
6. **Thai X Terminal** is a free terminal emulator on X Window. It has been modified to enable Thai input/output for natural use.
7. **GNU Emacs** [16] now becomes multilingual. Collaboration between ETL and NECTEC has been set to add complete Thai language support and dictionary companion to the editor environment [17].
8. **MySQL** [18] database server has been modified to sort Thai fields appropriately [19].
9. **Thai POSIX Locale** [9] is a set of Thai cultural conventions for standard C library. It works with GNU LibC 2.1.1.
10. **Thai LaTeX**, based on Babel package, allows Thai documents preparation on Linux.
11. **Thai Library** is an effort to define standard API for Thai support in applications and to provide some chosen solutions.

6. Conclusion

Solutions and practices are usually one step further than the standards. In such situation, interoperability problem will call for new

standards. The Internet has proven to be the main force in making new standard and interoperability adopted a lot faster than in the past. More and more developers are now joining force in the making of standards and putting these standards to work

Open source model does not only provide a means of cooperative development, but also allows the software to be standardized, and the standard conventions to be realized. Therefore we take both streams as our means to develop our information technology for the future. We have illustrated the case of Thailand, which is now gaining a tremendous trust from the open-source movement. The outcome is amazing: something real, usable and stable enough for mission-critical applications.

References

[1] Thaweesak Koanantakool and the Thai API Consortium. **Computer and Thai Language.** National Electronics and Computer Technology Center, 1987. *ISBN* 974-7570-66-1. (in Thai)

[2] Thaweesak Koanantakool and Adshariya Agsorn-intara, **Character Codes and Input/Output Method for the Thai Language.** CICC/NACSIS/National Electronics and Computer Technology Center, Tokyo, Japan, 1990. *http://thaigate.nacsis.ac.jp/refer/thaiconf/*

[3] Trin Tantsetthi. **An Annotated Reference to the Thai Implementations.** *http://www.inet.co.th/cyberclub/trin/thairef/*

[4] Trin Tantsetthi. **Campaign for Internet-Standard-Conforming Thai Usage.** (in Thai) *http://software.thai.net/tis-620/*

[5] Worawit Khangtrakool. **TIS-620 Friendly Version.** (in Thai) *http://www.thai.net/tis-620/*

[6] F. Yergeau. **UTF-8, a transformation format of ISO 10646.** RFC 2279. Alis Technologies. January 1998. (Obsoletes RFC 2044)

[7] Alain LaBonté. (Editor) **ISO/IEC FCD 14651.3 – International String Ordering – Method for comparing Character Strings and Description of the Common Template Tailorable Ordering.** June 1999.

[8] Theppitak Karoonboonyanan, Samphan Raruenrom and Pruet Boonma. **Thai-English Bilingual Sorting.** *http://www.links.nectec.or.th/~thep/blsort.html*

[9] Theppitak Karoonboonyanan. **Thai Locale.** *http://www.links.nectec.or.th/~thep/th-locale/*

[10] National Center for Science Information Systems. **NACSIS-Thai Project.** *http://thaigate.rd.nacsis.ac.jp/*

[11] uecthai@fedu.uec.ac.jp. **ZzzTh@i : How to use Thai with various computer platforms.** *http://zzzthai.fedu.uec.ac.jp/*

[12] Thai Linux Working Group. **Thai Linux Working Group.** *http://linux.thai.net/*

[13] Alfredo K. Kojima. **WindowMaker.** *http://www.windowmaker.org/*

[14] The Mozilla Organization. **Mozilla.org.** *http://www.mozilla.org/*

[15] Samphan Raruenrom. **Mozilla Thai Enabling.** *http://developer.thai.net/mozilla/*

[16] Free Software Foundation. **GNU Emacs.** *http://www.gnu.org/software/emacs/emacs.html*

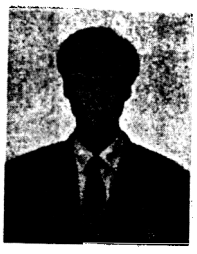
[17] Software and Language Engineering Laboratory, NECTEC. **Thai Language Support for Emacs.** *http://www.links.nectec.or.th/themacs/*

[18] T.c.X DataKonsultAB. **MySQL by T.c.X DataKonsultAB.** *http://www.mysql.net/*

[19] Pruet Boonma. **Thai Sorting Support for Free Database Server.** *http://linux.intanon.nectec.or.th/thaisortdatabase.html*

BIOGRAPHIES

Theppitak Karoonboonyanan



Theppitak Karoonboonyanan received his B.Eng. (Computer Engineering 2nd class honor) from Chulalongkorn University in 1993. He joined the Software and Language Engineering Laboratory (SLL), NECTEC since August 1995 as a research assistant. His experience includes Thai word processor and the Royal Institute Dictionary CD-ROM. This made him interested in Thai algorithms and implementations. He has also been involved in a number of Thai open-source projects, such as Thai POSIX locale. His interests include object-oriented software engineering and artificial intelligence.

Dr. Thaweesak Koanantakool

Director of NECTEC

Dr. Thaweesak "Hugh" Koanantakool received his Bachelor and Ph.D. degrees in Electrical Engineering from Imperial College of Science and Technology, London University. He had a number of industrial contracts in the UK before he came back to Thailand to start his government service career in 1981. He taught in Electrical Engineering with the Faculty of Engineering, Prince of Songkla University. In 1985, he moved to Bangkok, Thammasat University, and was appointed Associate Director of the Information Processing Institute for Education and Development. Since 1994, he became Deputy Director of NECTEC as well as leading the Network/Software Technology labs. Thaweesak introduced the Internet into Thailand and set up the largest academic and research network known as ThaiSarn under NECTEC. He later co-founded the first Internet Service Provider (ISP) owned by Thai government in 1995. The ISP, Internet Thailand Company Limited, at present is the largest ISP in Thailand, has 45% market share (by IP numbers managed). In 1996-1997, Thaweesak led Thailand's Information Superhighway test bed Project funded by NECTEC. The project was a major test bed in Thailand using ATM switches for both local area and wide-areas. In August 1998, Thaweesak was appointed the Director of NECTEC.