

December 13, 2005

Press Release

Graduate School of Interdisciplinary Information Studies,
The University of Tokyo
Ken Sakamura, Professor
7-3-1 Hongo, Bunkyo-ku, Tokyo
e-mai:t-font@sakamura-lab.org

Sakamura Laboratory at the University of Tokyo will make available the world's largest set of Chinese character fonts with 360,000 characters for free

120,000 characters x 3 font types (T Font Set) will be made available from December 14, 2005

Graduate School of Interdisciplinary Information Studies at the University of Tokyo (Bunkyo, Tokyo) has completed the world's largest set of Chinese character fonts (T Font Set) with 360,000 characters and will make it available for free from December 14. This set of fonts with 360,000 characters includes three types of TrueType fonts, namely *Mincho*, Gothic and *Kaisho*, and a Chinese character set with 120,000 characters (See Appendix 1 and 2). The set of 120,000 Chinese characters includes the GT character set with approximately 80,000 characters, most of which are included in the "Dai Kanwa Dictionary" and the "Kouki Jiten Chinese Character Dictionary", etc (see Appendix 5), the Song-Ming different-style character set with approximately 35,000 characters which are different-style characters selected based on literature of the Song-Ming Period, 661 characters selected from a re-written version of the *Kimbun* that is used in classic literature classes (see Appendix 3), and obsolete *Kana* characters included in the Basic Resident Register character set and *Kana* characters with voiced sound symbols used in comics (See Appendix 6).

As a result, this will be a major solution toward the problem of the shortage of fonts for different-style characters that has been a concern in document information processing in East Asia since this set includes almost all the different-style characters used on a daily basis in modern Japan except for rare exceptions. Furthermore, this set can be used for the computerization of Chinese classical literature, such as *Kanseki* (Chinese classics), thus it will serve as an academic infrastructure that supports the Oriental cultural research.

[Information 12]

For instance, in modern Japan, some characters used in many names of people and places are not included in the normal character sets on computers, such as the JIS character set, and each local government creates an original external character set to be used in its normal database systems for family registers, etc. Such ad-hoc handling may cause problems, for example, when a person whose name includes an external character moves his or her family register or resident card to another city, then a new external character will have to be created, or when multiple family register databases are integrated due to the recent municipal mergers, it will take a lot of work to integrate the external character sets with a large number of characters into a single database. The use of the character sets provided in the T Font Set with TRON character codes will solve almost all the problems of handling external character sets in Japan.

Additionally, from an academic point of view, it is expected that the T Font Set will bring innovative development to the digital archives of classic literary treasures in East Asia. To begin with, in Western countries, since documents are written in the Roman alphabet, which has a small number of characters set, it is easy to build digital archives of both classical literature and modern documents. Therefore, a large number of digital archives of classical literature already exist. However, in East Asia, one reason why classical documents have not been digitally archived is that there are many characters existing computers cannot handle correctly. Even in Japan, literary work from the Meiji Era, only 100 years ago or so, cannot be correctly expressed in characters on computers. It is expected that the T Font Set will enable literary works and historical materials, no matter if they are classical or modern literature, to be digitally archived and will significantly contribute to the establishment of new research methods that utilize digital technologies for East Asian cultural research.

By providing three types of fonts, namely, T *Mincho*, T Gothic and T *Kaisho* in this T Font Set, it will be possible to use them in many different types of documents. The T Font Set can be used on a BTRON3-OS, “*Cho-Kanji 4* (Super Chinese Character)”, and the T-Kernel / T-Shell which operate on the T-Engine, and Windows. The installation of this T Font Set into “*Cho-Kanji 4*” makes it possible to use a large character set of 540,000 character fonts for 220,000 character sets. Additionally, a character search tool that supports the input of characters in the Song-Ming different-style character set with approximately 35,000 characters will be made available for free (See Appendix 4). Furthermore, in order to expand and promote the use of this T Font Set, the “T Font Consortium (tentative name)” will be formed.

[Information 12]

The T Font Set and the character search tool will be made open to public from December 14 at the TRONSHOW2006 (12/14 – 16,2006 TOKYO INTERNATIONAL FORUM Co., Ltd. in Yurakucho, Tokyo) in a format that can be used on a BTRON3 OS, “*Cho-Kanji 4*”, the T-Kernel / T-Shell which operates on the T-Engine, and Windows. Additionally, their release on the following web site is also scheduled from the next spring. The research results of the “Building of Systems with TRON, using many Chinese characters” project conducted with a scientific research grant (specially-promoted research) is included in the T Font Set (See Appendix 7).

Notes

1. TRONSHOW2006 “Ubiquitous Computing is here to stay!”

Period: Wednesday, December 14th through Friday to December 16th, 2006

Venue: TOKYO INTERNATIONAL FORUM CO., Ltd.

Organizers: T-Engine Forum/TRON ASSOCIATION

2. URL for the T Font Set release

The T Font Set will be made available on the following site from the next spring.

<http://www.sakamura-lab.org/FONT/>

3. Inquiries

The Sakamura Laboratory at the University of Tokyo

Contact: Mr. Noboru Koshizuka, Associate Professor

(Email: t-font@sakamura-lab.org)

Appendix

Appendix 1 : List of Released Chinese Character Fonts

		Character sets						
		GT 78,765 characters	Non- Chinese characters 1,814 characters	Song-Ming different style characters			Re-writ ten version of the Kimbu n 661 charac ters	
				Ryugan shukan 15,117 characters	Henkai 12,837 charact ers	Hibets uji 4,190 charac ters		Record of Simplified Characters in the Song/Yuan period and later 2,355 characters
Font	GT Font (Mincho-type)	Available (Released by the Research Group on Multi-langua ge Processing at the University of Tokyo)						
	T Font Set	T- Mincho	Current release 115,739 characters x 3 types of fonts					
		T-Gothic						
		T-Kaisho						

[Information 12]

Appendix 2 : T Font Set Detailed Data

Character set	No. of characters	Font	Data size (MByte)	TRON Character code
GT	78,675	T-Mincho T-Gothic T-Kaisho	43.2 23.4 112.6	
Non-Chinese characters	1,814	T-Mincho T-Gothic T-Kaisho	3.5 3.5 2.7	Side 1/A Zone 0 x 2121 to 0 x 2D70 Side 1/B Zone 0 x A221 to 0 x AB77 Side9/B Zone 0 x 8021 to 0 x 8231 Side9/B Zone 0 x 8321 to 0 x 846A
Ryuganshukan	15,117	T-Mincho T-Gothic T-Kaisho	6.5 4.4 21.4	Side10/ A zone 0 x 3021 to 0 x 7E7E Side10 / B zone 0 x 8021 to 0 x D16D
Henkai	12,837	T-Mincho T-Gothic T-Kaisho	5.8 4.0 18.7	Side10/ B zone 0 x D221 to 0 x FD7E Side10 / C zone 0 x 2180 to 0 x 6686
Hibetsuji	4,190	T-Mincho T-Gothic T-Kaisho	1.7 1.1 5.3	Side10 / C zone 0 x 6780 to 0 x 7EFD Side10 / D zone 0 x 8080 to 0 x 899F
Record of simplified characters in the Song/Yuan period and later	2,355	T-Mincho T-Gothic T-Kaisho	0.9 0.7 3.0	Side 10 /D zone 0 x 8A80 to 0 x 9CD6
Re-written version of the Kimbun	661	T-Mincho T-Gothic T-Kaisho	0.4 0.3 1.3	Side 10 /D zone 0 x 9D80 to 0 x A29E

[Information 12]

Appendix 3: Song-Ming Different-Type Character Set/ Re-written version of the Kimbun Character Set (35,160 characters)

The Song-Ming different-type character set is a set of Chinese characters that includes a large number of different-style characters collected from text data of literature from the Edo Period in Japan and the Song, Ming and Qing periods in China. Characters that are in 4 Chinese character dictionaries of different-style characters, i.e. “*Ryuganshukan*”, “*Henkai*”, “*Hibetsuji*” and “Record of simplified characters in the Song/Yuan period and later” are the basis, and that are not included in the GT character set are included in the Song-Ming different-type character set. This set is divided into four different character sets based on the above four (Chinese character) dictionaries that were used to sort out the different-style characters. In the T Font Set, three different font types, i.e. *Mincho*, Gothic and *Kaisho*, are prepared for each character set.

The Re-written version of the *Kimbun* character set is a set of characters collected from a re-written version of the *Kinmon* in the *Kinmon* research. The selection of characters and character shapes when making fonts conform to the Song-Ming different-type character set and the “*Kimbun Tsushaku*” (by Shizuka Shirakawa) was used to sort out the characters. *Mincho*, Gothic and *Kaisho* fonts are available also for the Re-written version of the *Kimbun* character set.

Table: Song-Ming Different-Type Character Set/ Re-written version of the *Kimbun* Character Set

Source	Number of characters
<i>Ryuganshukan</i>	15,117
<i>Henkai</i>	12,837
<i>Hibetsuji</i>	4,190
Record of simplified characters in the Song/Yuan period and later	2,355
Re-written version of the <i>Kimbun</i>	661
Total	35,160

[Information 12]

Appendix 5 : GT Character Set (78,675 characters)

The GT character set was established by the “Research on Multi-language Processing in Multimedia Communication Systems” which is a project of the Research for the Future Program of the Japan Society for the Promotion of Science, and by its successor, the Research Group on Multi-language Processing at the University of Tokyo. The GT Font Set is available for free from the Research Group on Multi-language Processing at the University of Tokyo (<http://www.l.u-tokyo.ac.jp/GT/>) website. Additionally, the GT Font Set is pre-installed in the BTRON OS, *Cho-Kanji4*.

The T Font Set includes the latest GT Character Set (78,675 characters) that was updated on July 29, 2005 in the *Mincho*, Gothic and *Kaisho* fonts (it was found that some characters were included twice, therefore, the total number of characters which appears in the Table is slightly bigger than 78,675).

Appendix 6: Non-Chinese Character Set (1,814 characters)

The non-Chinese character set of fonts is a set of 1,814 characters, consists of the following; 168 obsolete *Kana* characters included in the Basic Resident Register character set, 204 *Kana* characters with voiced sound symbols (characters seen in comics, such as “あゝ” and “めゝ”), and 1,442 characters that are already included in the JIS mapping. In the T Font Set, the *Mincho*, Gothic and *Kaisho* fonts are prepared for these character sets and a fixed-width font and proportional font are available for vertical and horizontal writing.

Table: Characters Included in the Non-Chinese Character Set

Source	Number of characters
JIS X 0213	1,442
JIS X 0212	
<i>Kana</i> characters with voiced sound symbols	204
Obsolete <i>Kana</i> characters included in the Basic Resident Register character set	168
(Total)	1,814

[Information 12]

Appendix 7: “Building of Systems with TRON, using many Chinese characters” project conducted with a scientific research grant (specially-promoted research)

The Sakamura Laboratory at the Graduate School of Interdisciplinary Information Studies at the University of Tokyo has been conducting research on the establishment of digital platforms that support research on the Oriental culture through the “Building of Systems with TRON,using many Chinese characters” project with a scientific research grant (specially-promoted research) (From 2002 to 2005). In this research, the Laboratory has been working toward the building of digital platforms that support the field of study on *Kanseki* (Chinese classics), etc. focusing on the character systems of the Oriental cultures, especially “Chinese Characters”.

The Font set that will be made available this time was created as a result of this research and will serve as a basis of digital platforms that support research activities on the Oriental Culture.

At the Release of the T-Chinese Character Set

The Institute of Oriental Culture at the University of Tokyo
Takao Hirase, Professor
(Supervising Editor of the Chinese Character Set)

In the world that creates Chinese characters, both diversity and uniformity exist. Chinese characters that originated from one city have spread explosively and created a large area in which Chinese characters are used. Since then, the diversity of Chinese characters in different parts of the area that use Chinese characters and standardization for unification have co-existed until now. The *Koki Jiten*, a Chinese Character Dictionary was written to show the standard of character styles and modern dictionaries have inherited this spirit. The concept of “font”, in the sense that subtle differences, such as in dots, circles and straight lines, must be taken care of, adopted this spirit thoroughly and precisely.

A character that was considered to be a so-called different-style character, when looking at it through the eyes of the *Koki Jiten*, a Chinese Character Dictionary, was not a “different-style character” for the users. They just thought that it was the style they should use. It is us who consider it as a “different-style character”. We also use the term “simplified character”.

In this project, “different-style characters” and “simplified characters” in the above sense were collected and fonts were created in order to find a way to get into the hearts of the people who used them. Precision is required when making different fonts. Therefore, characteristics, such as differences between hand-written characters, have to be somewhat controlled. To deal with this unavoidable reality of restrictions, it was decided to take over the know-how of the GT Font set based on past experience. As a result, it was decided to include characters that had never been adopted in the GT Font Set.

To sort out different-style characters, “*Ryuganshukan*” and “*Henkai*” from the Song period, “*Hibetsuji*” from the Qing period and “Record of simplified characters in the Song/Yuan period and later” from the modern era were used. Different-style characters up to now were sorted out and most of these characters can be included.

The Re-written version of the *Kimbun* character set was made based on these and we selected “*Kimbun Tsushaku*” (by Shizuka Shirakawa) to start with.