

2005年12月13日

報道関係者各位

東京大学大学院情報学環

教授 坂村 健

東京都文京区本郷7-3-1

e-mail: t-font@sakamura-lab.org

東京大学坂村研究室が36万文字からなる 世界最大の漢字フォントセットを無料公開

12万文字×3書体の文字フォント（T書体フォント）を2005年12月14日より公開

東京大学大学院情報学環坂村研究室（東京都文京区）では、36万文字からなる世界最大の漢字フォントセット（T書体フォント）が完成し、このたび12月14日より無償で公開することをお知らせします。36万文字のフォントは、12万文字の漢字セットに対して、明朝体・ゴシック体・楷書体の3書体のTrueTypeフォントから構成されています（付録1、2参照）。12万文字の漢字セットの中には、『大漢和辞典』や『康熙字典』などに収録されている文字を中心とした約8万文字のGT文字セット（付録5参照）、宋明時代の文献を基に抽出された異体字からなる約3万5千文字の宋明異体字文字セットや古典で用いられる金文の釈文から抽出された661字（付録3参照）、そのほかにも住民基本台帳文字セットに収録された変体仮名やコミックなどで用いられる濁点かな文字なども含まれています（付録6参照）。

これによって、従来より東アジア地域における文書情報処理における懸案であった、異体字フォントの不足の問題に向けて大きく踏み込み、現代の日本において日常的に用いられる異体字問題はかなり希少な例外を除きほぼ解消するだけでなく、更に漢籍といった中国の古典文献の電子化にも耐えうる文字セットとなっており、東洋文化研究を支援する学術基盤ともなるものです。

例えば、現在日本では人名や地名がJIS等によるコンピュータ上の通常の文字セットに含まれていないものが多くあり、戸籍等を扱う通常の行政データベースシステムでは、行政単位毎に独自の外字セットを作成して対応しています。従って、外字が必要となる名前の人が戸籍や住民票を移動する時に新たに外字を作成する作業が生じることや、更に現在進められている市町村の合併などで複数の戸籍データベースを統合する際に、膨大な外字セットを統合する作業が大きな問題となっています。TRON文字コードによるT書体の文字セットを利用することで、国内においてはほぼ外字を利用する際の諸問題を解消することができます。

また、学術的な面においては、T書体フォントによって、東アジア地区の貴重な古典文献のデジタルアーカイブに画期的な進展をもたらすことが期待されています。そもそも、西洋においては、文書は小さい文字セットであるアルファベットによって表現されるため、古典でも現代文でもデジタルアーカイブを構築することが容易です。従って、既に現在でも膨大な古典文学のデジタルアーカイブが存在します。ところが、東アジアでは現在のコンピュータでは正確に扱えない

【資料12】

文字が多く存在することが原因の一つとなり、古典文献のデジタルアーカイブ化が疎外されてきました。日本でも、たかだか100年程度前の明治時代の文学ですらコンピュータ上で正確に文字表現できないのが現状です。T書体は、こうした古典か現代文学かを問わず、文学作品や歴史資料をデジタルアーカイブすることを可能にし、東アジア文化研究にデジタル技術を用いた新しい研究手法を確立することに大きく寄与するものと考えています。

今回、T明朝体・Tゴシック体・T楷書体の3書体を提供することで、様々な文書に利用することができます。なお、このT書体フォントは、BTRON3仕様OS「超漢字4」、T-Engine上で動作するT-Kernel/T-Shell、及びWindowsで利用することが可能です。このT書体フォントを、「超漢字4」にインストールして利用することで、22万文字分の文字セットに対する、54万文字フォントといった大規模文字セットを利用することが可能になります。また、約3万5千文字の宋明異体字文字セットの入力を支援する、文字検索ツールを同時に無料公開します(付録4参照)。更に、本T書体フォントの利用の拡大と普及をはかるために、「T書体コンソーシアム(仮称)」を結成します。

このT書体フォントと漢字検索ツールは、BTRON3仕様OS「超漢字4」、T-Engine上で動作するT-Kernel/T-Shell、及びWindowsで利用可能な形式で、12月14日よりTRONSHOW2006(12/14~16、東京有楽町東京国際フォーラム)で公開いたします。また、来春より下記のWebページからもリリースする予定です。なお、T書体フォントは、科学研究費補助金(特別推進研究)「トロンによる多漢字利用システムの構築」(付録7参照)の研究成果が含まれています。

記

1. TRONSHOW2006「ユビキタス始動！」

会期：2005年12月14日(水)~16日(金)

会場：東京国際フォーラム

主催：T-Engineフォーラム/社団法人トロン協会

2. T書体フォント公開URL

来春より、下記のwebサイトにて公開いたします。

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

3. お問い合わせ先

東京大学坂村研究室 担当：越塚 登 助教授

(E-mail: t-font@sakamura-lab.org)

付録

付録1：公開漢字フォント一覧

		文字セット						金文 釈文 661字
		GT 78,765字	非漢字 1,814字	宋明異体字				
				龍龕 手鑑 15,117字	篇海 12,837字	碑別字 4,190字	宋元以来 俗字譜 2,355字	
フォ ント	GT書体(明朝系)	既存(東京大学多言語処理研究会よりリリース)						
	T明朝体	今回リリースするフォント 115,739文字×3書体						
	Tゴシック体							
	T楷書体							

付録2：T書体フォント詳細データ

文字セット	文字数	フォントデータサイズ(MByte)	TRON文字コード
GT	78,675	T明朝体 43.2 MByte Tゴシック体 23.4 MByte T楷書体 112.6 MByte	
非漢字	1,814	T明朝体 3.5 MByte Tゴシック体 3.5 MByte T楷書体 2.7 MByte	1面Aゾーン 0x2121 ~ 0x2D70 1面Bゾーン 0xA221 ~ 0xAB77 9面Bゾーン 0x8021 ~ 0x8231 9面Bゾーン 0x8321 ~ 0x846A
龍龕手鑑	15,117	T明朝体 6.5 MByte Tゴシック体 4.4 MByte T楷書体 21.4 MByte	10面Aゾーン 0x3021 ~ 0x7E7E 10面Bゾーン 0x8021 ~ 0xD16D
篇海	12,837	T明朝体 5.8 MByte Tゴシック体 4.0 MByte T楷書体 18.7 MByte	10面Bゾーン 0xD221 ~ 0xFD7E 10面Cゾーン 0x2180 ~ 0x6686
碑別字	4,190	T明朝体 1.7 MByte Tゴシック体 1.1 MByte T楷書体 5.3 MByte	10面Cゾーン 0x6780 ~ 0x7EFD 10面Dゾーン 0x8080 ~ 0x899F
宋元以来俗字譜	2,355	T明朝体 0.9 MByte Tゴシック体 0.7 MByte T楷書体 3.0 MByte	10面Dゾーン 0x8A80 ~ 0x9CD6
金文釈文	661	T明朝体 0.4 MByte Tゴシック体 0.3 MByte T楷書体 1.3 MByte	10面Dゾーン 0x9D80 ~ 0xA29E

付録5：GT文字セット（78,675字）

GT文字セットは、日本学術振興会の未来開拓学術研究推進事業「マルチメディア通信システムにおける多国語処理の研究」プロジェクト、およびそれを引き継いだ、東京大学多国語処理研究会によって制定された文字セットであり、GT書体フォントが東京大学多国語処理研究会のWebページから無料配付されています（<http://www.l.u-tokyo.ac.jp/GT/>）。また、GT書体フォントはBTRON仕様OS超漢字4に標準装備されています。

T書体フォントは、2005年7月29日付けで更新された、最新のGT文字セット78,675字に対応する、明朝体・ゴシック体・楷書体のフォントを備えています。（一部、文字の重複が判明した文字などが欠番になっているため、表の文字数合計は78,675よりも若干多くなります）

付録6：非漢字文字セット（1,814字）

非漢字文字セットフォントは、住基収録の変体仮名168字、濁点かな文字（マンガなどに見られる「あゝ」「めゝ」などの文字）204字、JISマッピングの既収録文字1,442字からなる、合計1,814字の文字セットです。T書体ではこれらの文字セットに対して、明朝体・ゴシック体・楷書体のフォントを提供し、更にそれぞれ縦書き・横書き用の、固定ピッチ・プロポーションアルフォントが利用可能です。

表：非漢字文字セット収録文字

出典など	文字数
JIS X 0213【第1面 非漢字】	1,442
JIS X 0212	
濁点かな文字	204
住基変体仮名	168
（合計）	1,814

付録7：科学研究費補助金（特別推進研究）「トロンによる多漢字利用システムの構築」

東京大学大学院情報学環坂村研究室では、科学研究費補助金（特別推進研究）「トロンによる多漢字利用システムの構築」（平成14年度～平成17年度）により、東洋文化研究を支援するデジタルプラットフォームの確立に関して研究を行ってきました。本研究では、東洋文化圏の文字体系の中でも特に「漢字」に焦点を絞り、漢籍研究などの分野を支援するデジタルプラットフォームの構築を目指しています。

今回公開されるフォントは、この研究の成果として作成されたもので、東洋文化研究を支援するデジタルプラットフォームの基盤となるものです。

T 漢字文字セット公開にあたって

東京大学東洋文化研究所

教授 平勢隆郎

(漢字文字セット監修)

漢字が作り出す世界は、多様性と統一性が混在する。漢字は一つの都市の文字として始まり、爆発的に広がって広域的漢字圏をつくりあげた。そのときから、各地でひきおこされる多様化の芽と、統一を指向する基準づくりが、併行して存在し、今にいたった。字体としての規範を示したのが、康熙字典であり、現代の辞書はその精神を継承する。フォントという概念は、点マル、直線など微細な差異に気を配るという意味で、康熙字典の精神を徹底し厳密化した。

康熙字典の目から見て、異体字とされるものは、実はその異体字を用いていた人々からすると「異体字」という意識をもって使ったのではない。それが使うべき字体だと考えて使ったものである。それを「異体字」だと見なすのは、我々である。「俗字」という言葉もわれわれが使う。

われわれは、その意味における「異体字」・「俗字」を集めてフォント化した。それをを用いていた人々の心に触れる方法を模索するためである。フォント化は厳密化を伴う。だから、手書きの字の違いなど個性は多少規制される。この避け得ない規制という現実に対し、われわれは、これまでの経験から GT フォントのノウハウを継承することにした。その結果として、われわれは GT にこれまで採用されていない字を作ることになった。

異体字の整理には、宋代の『龍龕手鑑』・『篇海』、清代の『碑別字』、近代の『宋元以来俗字譜』を使った。これまでの異体字の整理が主としてこれらにより、またこれらでほとんどのものがカバーできるからである。

金文釈文は、これらに準じて作ったもので、材料は、手始めとして『金文通釈』(白川静)を選んだ。