

ISO/IEC 10036 によるグリフ識別子登録の現状とその応用

Registration of Glyph Identifiers by ISO 10036 and Its Application

上村 圭介[†] 小町 祐史[‡]

Keisuke KAMIMURA[†] Yushi Komachi[‡]

[†] ISO 10036 RA (国際大学 GLOCOM) [†] GLOCOM, International University of Japan

[‡] ISO 10036 RA (パナソニック・コミュニケーションズ) [‡] Panasonic Communications, Co., Ltd.

E-mail: [†] kmmr@glocom.ac.jp, [‡] komachi@y-adagio.com

1. はじめに

インターネットの普及・発展を通じて、電子文書による情報交換の役割はますます重要になっている。しかし、標準化された符号化文字集合に記載されていない文字や記号類の電子文書での扱いが、改めて問題となっている。

この問題を解決するための一つの手段として、利用者からの要求がある文字や記号類に対して、文字コードとはレベルの異なる識別子を付与し、その識別子によって文字や記号類についての情報交換を行なう仕組みが考えられる。

本稿ではこのような仕組みの一つとして ISO/IEC 10036 によるグリフ登録制度を取り上げ、その現状と効果、そして今後の応用の可能性について述べる。

2. ISO/IEC 10036 による登録

ISO/IEC 10036 (Information technology -- Font information interchange -- Procedures for registration of font-related identifiers)は、単体のグリフ、グリフ集合、フォント属性集合といったフォントリソース(これらを総称して「フォント関連オブジェクト」と呼ぶ)を登録し、一意の識別子を付与する手続きを定めた国際規格である。

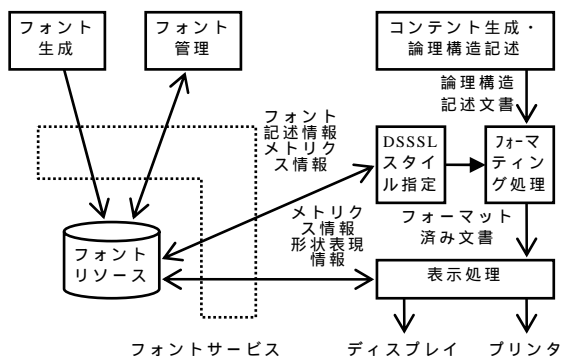


図 1: 電子化文書生成環境でのフォント情報[1]

ISO/IEC 10036 に規定される登録手続きは、図 1 に示す電子化文書の処理環境の体系において、フォントリソースをフォント作成者やフォント管理者、あるいは DSSSL スタイル指定や表示処理の処理系に対して提供するフォントサービスの一部を担うものである。

現在、ISO/IEC 10036 による登録手続きは指定登録機関である国際大学グローバル・コミュニケーション・センター (GLOCOM) によって処理されている。登録機関では、グリフに対して付与された識別子の管理を行なうとともに、登録簿の一部 (登録済みグリフの一覧など) をオンラインで公開している [2]。

グリフは、登録申込人の登録要求に応じて原則として先着順に登録されるが、多数の登録要求が同一の登録申込人から行なわれる可能性がある場合には、一定のブロックをその登録申込人に対して割り当てることもある。現在、登録は 100 グリフを単位に受け付けられ、登録に際しては登録されるグリフの数に応じた登録費用が発生する。

登録簿に記載されるのは、登録申込人情報 (氏名、所属、連絡先など)、グリフ情報 (形状データの本体、形式、容量など)、グリフ記述 (そのグリフがもつ機能)、フォント情報 (そのグリフに関連づけられる文字コードや、そのグリフをレンダリングするためのフォント等の入手先)、および補助情報である。

2004 年 5 月 1 日現在、グリフ登録簿には、151,367 件のグリフが登録されている。図 2 に、登録されたグリフの例を示す。

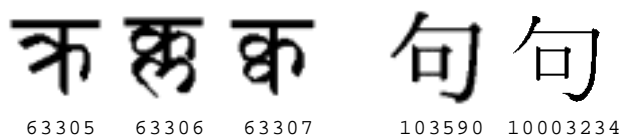


図 2: 登録されているグリフと識別子

3. ISO/IEC 10036 による登録の特徴

現在の運用では登録済みのグリフと重複するグリフについて登録要求がなされた場合でも、登録を拒否することはない。つまり、仮に同一のグリフだったとしても、それに対して複数の識別子が登録されることがある。図 2 の 103590 および 10003234 のように「一見」同一のグリフが登録され、異なる識別子を付与されているケースがある。

このような割り当て方針に対しては、情報交換の可能性を狭めるものであるという懸念も表明されている[3]。しかしながら、もともと、グリフの同一性の判断は、文字の同一性の判断に比べると主観性が高い。ある利用者にとって同じグリフと判断されるものであっても、別の利用者の要求の下では異なるグリフと判断されることがありうる。

そのような性格をもつグリフの同一性の判断を登録機関が一律に行なうことは適切でない。そこで、現在の運用では、グリフを登録する登録申込人が、登録済みのグリフと異なると判断したグリフについてはすべて登録を受け付けている。一方で、登録制度の濫用は、登録の有料化によって牽制されている。

ISO/IEC 10036 のグリフ登録簿では、グリフ登録は、識別子によってグリフを一方向的に特定するための仕組みであり、その逆ではない。これは、文字コードがコードから文字を特定し、また文字からコードが特定できるものであることとは対照的である。

そのため、登録簿に登録されたグリフは、符号化文字集合に記載された文字と比べると再利用性に劣る。しかしながら、ISO/IEC 10036 が登録の対象とするオブジェクトは、同一性や再利用性についての強い合意が得られなかったと考えるべきであり、グリフに対して文字と同等の同一性や再利用性を期待するのは正しくないと思われる。そのため、グリフの再利用性を高めるためには、マッピング・テーブルなど外部的な手段が必要となる。

4. グリフ登録の応用

現時点の登録簿の機能は、グリフに与えられた識別子を管理するだけで、レンダリング用リソースの転送を初めとした機能は提供されていない。そのため、今後、自動登録システムを含む、グリフリソースの管理システムを開発する必要がある。

また、登録されるグリフ情報は、登録申込人情報、グリフ情報およびグリフ記述といった「基本情報」に限られているが、将来的には補助情報として音声、アニメーション、動画などのメディアオブジェクトを登録できる登録簿の機能を整備する計画である。

そうすることで、グリフ登録、静止画像ではなく動

画として表現されるロゴマークや、音声を伴うロゴマークなどの登録への応用が可能になる。

グリフ登録制度が応用できるもう一つの分野として考えられるものに、OpenType フォントのグリフ置換メカニズムがある。このメカニズムは、文脈によって異なる字形を選択する合字の要求や、単語中の位置によって字形を変化させるアラビア文字などの要求に対応するために、グリフのレパートリーから文脈によって適切なグリフを選択するための機能であり、日本語の異体字にも応用され始めている。

ところが、異体字など、メーカー依存のグリフに対しては、メーカーごとに別々に番号が振られているため、グリフ番号の重複が起こる可能性がある。番号の重複は今の段階では大きな問題とはなっていないが、今後 OpenType フォントの普及とともに、グリフ置換メカニズムによるグリフや字形の切り替えが普及すると、番号の重複による混乱が生じることが予想される。

そこで、このようなフォント置換メカニズムを利用するにあたっては、グリフ番号の重複を避けるためにも、グリフ登録制度を活用することが必要である。

ところで、すでに述べたように、この意図は各ベンダー間の異体字を共通化するというのではなく、あくまで番号の重複による混乱を回避することである。

5. 結論とまとめ

非標準的な字形やグリフをどのようにして識別し、そのようなグリフの出力用データをどこに保持するかという問題は、今後、住民基本台帳ネットワークなどで非標準的なグリフや字形を含む人名・地名が電子化され、交換されるケースが増えることを考えると、早急な解決が求められている分野である。また、この登録制度は、クメール文字やモンゴル文字など、日本語以外のアジアの諸言語に存在する「文字ニーズ」の吸収にも活用できる。

登録簿は現時点では十分な実証が詰まったものではないが、今後、ISO/IEC 10036 のメカニズムを応用したグリフリソース交換環境の整備を進め、電子文書の利用性、ひいては社会基盤としての電子文書の信頼性を向上させることを目指したい。

文 献

- [1] 小町祐史, 電子出版技術入門, オーム社, 東京, 1993.
- [2] ISO/IEC Font-Related Object Registry.
<<http://media.glocom.ac.jp/kmmr/10036/>>
- [3] C. Wittern, "Embedding Glyph Identifiers in XML Documents," 漢字文献情報処理研究, vol. 4, pp. 1-5, October 2003.