

漢字グリフ管理 Wiki システム (GlyphWiki) の構築

上地 宏一

二松学舎大学 21 世紀 COE プログラム

本研究は漢字資料のデジタル化の際に問題となる文字コードに含まれない漢字の扱いについて、インターネット上で共有可能な漢字字形データベースの構築によって、ユーザーが必要な文字を自由に登録し、他者に公開できることを提案するものである。Wiki システムを応用したデータベースとすることで、誰もが自由に平易にシステムを利用することができ、また漢字グリフのデータ処理に KAGE システムを用いることで、従来のアウトライントフォント形式よりも平易なデザイン作業を可能とする。本論文は実際にデータベースを実装し、システムの評価を行ったところ一定の評価が得られたことを報告する。

GlyphWiki : Construction of Wiki System for Kanji Glyph Management

Koichi Kamichi

21st Century COE Program

Nishogakusha University

In this study, I propose a construction of shareable Kanji glyph database system on the Internet for a solution of using undefined Kanji characters in character coded sets in digitizing Kanji text resources. Users can register Kanji glyphs to the database if necessary and the data will become open to the public automatically. The database applied the Wiki system, that users can use it freely and easily. And also database applied the KAGE system for processing of Kanji glyphs data so users can easily design glyphs in comparison with ordinary Outline based font formats. The system has been implemented and I report that evaluation of the system was provided.

1. はじめに

今般のインターネットの普及、研究成果情報の公開、研究活動への外部資金導入の奨励に見られるように、学術研究者は研究成果の対外公開促進が望まれている状況にあり、コンピュータで研究資料や発表論文等を作成することが求められている。しかしながら漢字を扱う東洋学研究者の多くは「使いたい漢字が使えない」という問題に常に悩まされてきた。

ユーザー定義文字（いわゆる外字）は、符号化文字集合（文字コード）において定義されていない文字を使いたい、収録文字とは異なる字形を使いたい、フォントが用意されていない文字を使いたい、といったニーズに対して有効であるが、データの公開において情報交換が保証されないという欠点を持つ。また外字フォントを作成するのは面倒である。そこで現実的な解として、大規模外字フォントである「今昔文字鏡[1]」や「GT フォント[2]」といった無償公開フォントを利用し、情報提供者と閲覧者が同じ外字フォントを共有する方法が一般的となっているが、これらの大規模外字フォントにも含まれていない文字を使いたい、というニーズは無くなる。

本研究では、インターネット上で漢字字形を管理する文字データベースを提案する。ユーザーが漢字字形を登録すると字形画像や外字フォントが自動的に生成され、公開状態となることで、ユーザーの外字管理・公開の手間を省くことができるものである。本論文では、データベースの概要および評価実験について報告する。

2. ネットワーク外字共有とフォント

これまでもネットワーク上で外字を共有し、テキストデータの情報交換を保証する仕組みが考案されてきた。実用化された代表例として XKP[3]が挙げられるが、その主対象は閉じた一定のネットワーク（例えば自治体や企業内の LAN）であり、研究成果の公開のような不特定多数の閲覧者を対象とするものではなかった。

また、Unicode 技術文書として公開されている Ideographic Variation Database[4] (IVD) では、データベースに登録した異字形をテキスト内で指定できるものである。アドビ社が独自に規定した文字集合である Adobe-Japan1[5]は現在この IVD に登録申請中である。申請手順などを考慮すると誰もがすぐに異字形を登録できるとは考えにくく、また新字種には対応していない。

このように、ユーザーが主体となつて必要な文字を利用する形態として用意されているはず

の外字が、単にフォント提供者が用意した外字データをいかに共有するか、という目的に置き換わっているのが現状である。

一方、グリフ（文字字形）を登録し付与された一意の番号を XML などの構造化文書においてメタ情報として参照可能とする ISO/IEC 10036[6]の仕組みも有効であると考えられるが、実態はグリフの登録簿であり、登録するグリフのデザインやフォントの公開は登録者が行うべきものであるため、コンテンツ提供者の手間はこれまでと変わらない。

今日普及しているフォント形式である TrueType や OpenType は、文字字形の輪郭線を直線や曲線の集合として記述するデータ形式であるが、実際に文字をデザインするためには熟練した技術が求められ、東洋学の研究者が外字フォントを作成することは困難である。また輪郭線による字形のデータ化は、漢字の特徴である部品の組み合わせによるデザイン手法と親和性が低く、部品を組み合わせた後に筆画の太さの調整作業が必要となり、特に曲線の太さの調整は煩雑である。さらに、多くの既存のフォントは、使用許諾契約（ソフトウェアライセンス）によって二次使用が制限されており、フォントに含まれる部品を活用して新たにフォントを作ったとしても、公開できないケースがある。

3. グリフウィキ (GlyphWiki) の概要

そこで能動的な外字管理環境の構築を目的とした新たな漢字グリフ管理システムを提案する。管理システムはいわゆる Web データベースの形態をとり、データ管理をインターネット上で行う。つまり、外字データの登録や編集をインターネット上で行い、情報の更新がすぐに公開内容に反映されるようにする。このことで、デジタルテキストを作成する際、足りない漢字（字形）があれば、管理システムの外字データを検索し、既に用意されている場合はそれを用い、登録されていない場合はその場で登録を行うことができる。別のユーザーがデジタルテキストを閲覧するために外字フォントが必要である状況は今までと変わらないが、必要な外字フォントは、システムから自由にダウンロードできることとする。このため、データ作成者は外字フォントを公開・配布する作業から解放される。もしくは、登録されている字形を画像ファイルとして取得しテキストに埋め込む（または、HTML ドキュメントから画像ファイルを参照する）利用方法も考えられる。

データベースに外字データを登録する際に、個々の漢字グリフは一意の名前を付与することで識別を行い、命名規則は特に設けない。また登録する文字が既存のデータに含まれているかどうかの判断は各ユーザーが行うこととし、同定判断の審査は行わない。その代わりに漢和字

典や文字コードにおいて規定されている文字、出典が明確な文字などについては（例として「諸橋大漢和 2751 番」「ISO/IEC 10646 の U-00020000」など）、強制力のない命名ガイドラインを設けることとするが、ユーザーが判断できない、または意図的に独自の外字集合としたい場合は、ユーザーが自由に命名することになる。さらに、ユーザーごとに操作できるデータベースを区別せず、全てのユーザーが全てのデータにアクセスできる特徴を持つ。これにより他のユーザーが作成したデータをそのまま、あるいは部分的に活用することが可能となり、同時に自分のデータを他人に提供することにもなる。このようにデータの内容についての管理者がいないデータベースであるため、同名による漢字グリフが登録された場合は、新しいデータによって上書きすることとなる。ただし、過去の全履歴を参照可能で、版の指定により個々の版のデータを一樣に利用できることで、登録された全データの利用を保証する。また、各ユーザーが作成・編集する権利を占有するデータを登録するための命名規則を特別に用意することで、ユーザー独自の外字集合を登録することを可能とする。ただし、この場合でも登録されたデータの参照や二次利用の権利は一般に開放される。

データの管理者を置かず、文字の登録に恣意的な判断をせず自由に登録できることを定義した結果、漢字グリフ管理システムを実際に設置する上で、設置者の選定（私企業が行ってもよいのか、中立的な国家機関が行うべきか、そもそも国が文字字形の判断を行ってよいのか）に迷うことも無く、また大掛りな予算が不要で、ビジネス面での需要と供給を満たさなくても設置できるため、小さな外字集合の利用を目的とした研究者などの個人ニーズをも満たすことが可能となる。

3. 1 Wiki システムの利用

このような Web データベースを実現するために、ワード・カニングハムが提唱した WikiWikiWeb[7]（本稿では Wiki システムと記す）をシステムのベースとした。その理由は二つある。一つは Wiki システムを管理コストの小さい、操作の平易な Web システムとして利用するためである。これはブログと呼ばれる Web 日記システムが、日記本文を打ち込むだけで自動的に Web コンテンツとして加工され、インターネット上にすぐに公開できる便利さと類似する。もう一つは Wiki システムの中でも特に一般に知られている「フリー百科事典」であるウィキペディア [8] (Wikipedia) を代表例とするパズールモデルを活用し、共同・協同作業としての漢字グリフ管理ツールを構築するためである。

特に後者の理由はまた副次的な狙いとしての、フリー（自由・無償）なフォントの協同制作の

ツールとしての活用を考えている。すなわち、ISO/IEC 10646 において定義されている七万字のうち、五万数千字については日本デザインのフォントとしては提供されていない（マイクロソフト社の Windows Vista では中国大陸デザインおよび台湾・香港デザインの 2 種類の 7 万字フォントが用意されているが、制作コストと実際のニーズを比べると今後日本デザインでの 7 万字フォントが提供されることは考えにくい）。この空白部分について、ネットワーク上の一般ユーザーが分担して漢字グリフを埋めていけば、少ない制作コストで、しかもフリーなフォントができるのではないだろうか。これはバザールモデルを用いた大規模漢字フォントデザイン、つまり Wikipedia 百科事典ならぬ漢字グリフ百科辞典の実現である。当然のことながらデザイン品質や字体の統一性など、考えられる問題点は複数あるが、従来ないフォントデザインのモデルとして思案中である。

また管理システムでは Wikipedia の運用ソフトウェアである MediaWiki の「ノート」機能を取り込んだ。もともと Wiki システムは不特定多数のユーザーによるデータの編集がなされるため、データの統一性保持や運用に一定の権限を持った管理者の存在が望まれる。しかし Wikipedia のように参加するユーザーが多く、特定の管理者に頼っては運用できない場合は、データ細部の調整自体もユーザーが行うことが要求される。このため MediaWiki 独自の機能として「ノート」と呼ばれる、個別の記事に対してユーザーがコメントを記述できる場が用意されている。例えば主観の異なるユーザー同士が一つの記事に対する差し戻し編集を繰り返すような場合に、意見調整の場としてノートが活用される。ここでは、当事者の議論だけでなく傍観者からの賛否の表明も行われ、一定のルール（意見表明期間の設定と多数決による決議）に則り解決が図られる。必ずしも全ての議論に対してノートが有効であるとは言えないが、Wiki システムの中でも独特の機能である。漢字グリフ管理 Wiki システムの場合、想定される利用ユーザー数はそれほど多くないが、データ内容についての管理者は不在であり、また文字のデザインは見慣れているかどうかなどの主観が大きい分野であるため、このような意見の調整機能が有効である

と考えた。

以上のようにインターフェースをはじめとする Wiki システムの設計において MediaWiki を直接の参考とした。このことにより、システム設計や画面デザインに費やす時間を減らすことができた。さらに、ユーザーにとっても見慣れている Wikipedia と同じシステムであれば親和性が高いというメリットも考えられる。

3. 2 KAGE システムの利用

漢字グリフ管理 Wiki システムの構築において、大きな問題となるのが「平易な漢字グリフデザイン手段」の確保である。文字字形の輪郭線を直線とベジエまたはスプライン曲線の集合で表現する既存のアウトラインデータは、デザイン技術の習得が困難である。そこで、管理システムでは、既に開発していた KAGE システムを利用することとした。KAGE システムとは、漢字グリフを筆画ごとの骨格情報の集積で表現する、グリフデータの中間表現形式である。現在フォント処理で主流のアウトライン形式によるデータの最小単位は、輪郭を構成する 1 つの座標点であるが、KAGE システムでは漢字字形を筆画に分解したものをデータの 1 単位とする。各筆画は線の種類（直線、曲線など）を指定し、筆画の骨格に相当する中心線の位置情報および筆画の形状を情報として記述する。中心線は、直線の筆画の場合には始点と終点の座標情報を、曲線の筆画はさらに 1, 2 点の制御点の座標情報を加えた、2 点から 4 点の座標情報で表現する。筆画の形状は、筆の入り、跳ね、他の筆画への接続、止め、といった数種類の形状を頭部と尾部の 2 種類について記述する。これら数項目の情報で一つの筆画をあらわし、筆画種数分のデータを組み合わせると一つの漢字グリフデータを構築する（表 1、図 1）。このデータを元に、明朝体の特徴を持つアウトライン形式のデータへの変換プログラムを用意することで、漢字グリフ画像ファイルや一般的なフォントファイルに変換することを可能とする（図 2）。また筆画による表現だけでなく、漢字部品を位置と大きさの二つの情報の記述により引用することも可能であり、漢字部品を組み合わせると新しい漢字グリフを容易にデザインすることが可能となっている。例として「永」という漢字グリフを表現するためにアウトライン形式では 62 単位の

線種:2	頭形状:7	尾形状:8	座標:(73, 13)	(98, 18)	(110, 36)
線種:1	頭形状:0	尾形状:2	座標:(48, 50)	(103, 50)	
線種:1	頭形状:22	尾形状:4	座標:(103, 50)	(103, 184)	
線種:1	頭形状:0	尾形状:2	座標:(22, 80)	(76, 80)	
線種:2	頭形状:22	尾形状:7	座標:(76, 80)	(70, 143)	(14, 174)
線種:2	頭形状:0	尾形状:7	座標:(173, 71)	(154, 87)	(124, 108)
線種:2	頭形状:7	尾形状:0	座標:(108, 57)	(122, 139)	(177, 168)

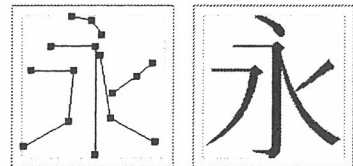


表 1 KAGE システムによる「永」字形データ 7 筆画 39 単位

図 1 骨格による筆画の表現 図 2 最終字形

情報が必要であるのに対し、KAGE システムでは 39 単位で表現が可能であり、「議」では 187 単位に対して 8 単位での表現が可能となっている（これほど少ない情報量であるのは「言」と「義」の部品引用を行っているためであり、筆画ごとに分解した場合は 112 単位の情報量となる）。

KAGE システムの漢字グリフデータは数値の集合で表され、これを一行一筆画に相当するテキストデータとして表記することができる。このため、先述の Wiki システムの一記事分を当管理システムでは漢字一字形と見なし、KAGE システムによって表現された漢字グリフデータを一つの記事として登録する方式を取ることで、データベースの設計は Wiki システムの仕様をそのまま利用することが可能となった。実際には KAGE システムによる漢字グリフデータを数値の記述でデザインすることは難しいため、座標点をマウスで操作することでデザイン可能なグリフエディタを用意している。このエディタは一般に普及しているアドビ社の Flash コンテンツとして実装しているため、管理システムを操作する Web ブラウザ上でそのまま利用し、またデザイン編集した結果を同じウィンドウ内で管理システムに登録できる。

3. 3 データライセンスの決定

管理システムでは、漢字グリフデータを自由に登録し、自由に活用することを想定している。そのため、著作権などに制限されないデータのライセンスを決定する必要がある。Wiki システムとして参考とした Wikipedia では、投稿する記事のライセンスは GFDL [9] というフリーソフトウェア財団 (FSF) が提唱している文書を対象としたライセンスに Wikipedia の独自の解釈を加えたものが適用されている。GFDL ライセンスの根底にあるのは FSF が提唱する「コピーレフト」という概念で、フリーとは、「無償」ではなく「自由」を指し、将来にわたってそのデータおよびその全ての派生物が自由に利用できる状態を維持できることを保証するための制限が設けられる。このためコピーレフトを主張するデータを、別のデータと結合して新しいデータを作成した場合、新しいデータ全体がコピーレフトとなる必要がある。漢字グリフ情報の場合、グリフそのものに著作権が発生するかどうかは、少なくとも日本の現行法においてフォントに著作権が存在しないという見方が強いことから否定的に解釈するべきであるが、仮に著作権が存在すると見なしたときに、その漢字グリフを集めてフォントファイルに変換し、そのフォントの漢字グリフを埋め込んだ文書ファイル (PDF 形式などが想定される) を作成した場合には、他の著作物と合体することになる。その際に漢字グリフに GFDL ライセンスを適用しているとフォントファイルとして合体した著作

物 (文書ファイル) 全体に対して GFDL ライセンスの適用が求められる可能性が生じる。つまり、作成したプライベートな内容の文書ファイルに対しても他者による自由利用の保証が求められることになるが、これは現実的ではない。もっとも FSF が別に提唱している、ソフトウェアを対象とする GPL ライセンスでは、フォントファイルにライセンスを適用した際の文書へのフォントの埋め込みについて、埋め込んだ文書に対しては GPL ライセンスの適用を行わない旨の特記事項を付記することで GPL ライセンスをフォントに適用することも可能である、という解釈がなされている。いずれにしても「コピーレフト」の思想は、将来にわたる自由利用を担保するために、著作物の利用者においてライセンスに対する一定の注意が求められることになり、管理システムの求める自由利用の方向性とは一致しないと考えた。そこで、管理システムとして一次的にのみ利用の自由を保証し、そのデータを二次利用して作成した派生著作物に対しては自由の保証を求めない最も緩い形式のライセンスとすることにした。その方法としては、ユーザーがデータを管理システムに登録する際、そのデータはシステムを運用する主体者に権利を移譲することとし、さらに日本の著作権法では放棄できないとされている人格権については、今後権利を行使しないことを了承してもらった上でデータの登録を受け付けることとした。そのデータを管理システムでは条件をつけずに自由な利用を認める形で公開する。このことによりシステムに登録するデータは将来的にいかなる形態の利用も妨げないことを保証できる。権利の移譲については、他のブログサイト等にも見られる方式であり、著作権権利の完全放棄が法的に有効かどうか、移譲契約として有効かどうかの厳密な判断は不明であるが、運用ポリシーの明示という点においてユーザーに注意を促すことで自由利用の保証に関する理解を得ることができるものと考えられる。

4. Wiki システムの構築

ユーザーが必要な漢字グリフを自由に登録し、自由に利用できる漢字グリフ管理 Wiki システムを提案し、実際にシステムを構築した (図 3)。先述の通り、MediaWiki (Wikipedia) をシステム設計の直接の参考としたが、登録データの対象が漢字グリフでありソフトウェアの機能改造が必要な工数が多いと考え、ソースが公開されている MediaWiki そのものは用いずに、仕様だけを参考に独自にソフトウェアの実装を行った。管理システムの名称は「グリフウィキ」とした。システムの実装は Web サーバ上で動く CGI プログラムとして Perl 言語を利用した。KAGE データをアウトライン形式に変換するプログラムとして JavaScript (Mozilla Project による

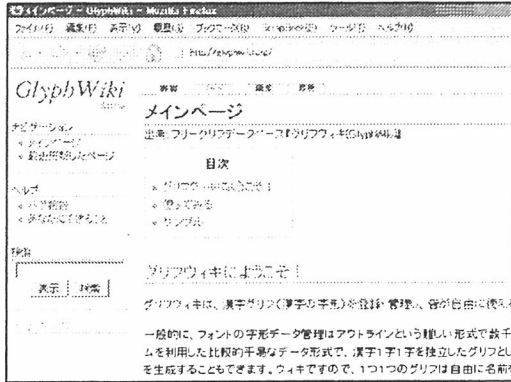


図3 グリフウィキ トップページ

SpiderMonkey エンジン[10]) を用いているほか、画像形式の変換、TrueType フォント生成のための補助プログラムとして ImageMagick[11]および FontForge[12]を援用している。データベースは現時点では SQLite[13] (Perl モジュール組み込み版) を利用している。Wiki 書式記述データの HTML 形式への変換部分については Wiki クローンである YukiWiki[14]を参考とした。

また、デザイン時に利用することを想定した基本漢字部品のために、漢字データベース計画[15]で公開している JIS X 0208:1997 に含まれる 6,355 漢字+1 字の漢字字形データを登録している。これにより多くの漢字字形について、部品を組み合わせた平易なデザインが可能となっている。

現在、システムの実装が完了し、システム評価版としての運用を行っている段階である。

5. システムの評価

実際に現在のコンピュータでは利用できない、もしくはフォントの実装例が少ない文字種をグリフウィキに登録したところ、問題なく利用できることがわかった。このほか、Web システムとしてのユーザビリティを調査するための実験を行った。

5. 1 グリフの登録

フォントの実装例が少ないサンプルとして、ISO/IEC 10646 の Ext. B や、2005 年の Amd.1[16]

煇 曙 匪 蕪
 藪 戠 伝 ヂ
 廡 輓 睿 蝨

図4 実際に登録したコード外字の一部

で追加された U+9FA6 以降の文字の一部についてグリフの登録を行った。また、笹原宏之著『日本の漢字』、エツコ・オバタ・ライマン著『日本人の作った漢字』に見られる漢字の一部を登録した。このほか現在、二松学舎大学 21 世紀 COE プログラムで公開中の日本漢文文献目録データベースに登録されているコード外字の登録を行っている。これらの漢字について、いずれも問題なく登録可能であり、グリフ画像を参照できることが確認できた(図4)。

5. 2 初期利用に関する評価実験

グリフウィキは、フォントデザインの非専門化が平易に外字管理を可能とすることを目標としている。このことを検証するため、被験者 3 名による実験を行った。被験者はいずれもメディア系学部的大学生である。

実験は、被験者に見知らぬ漢字を提示し、グリフウィキに登録した上で、漢字字形画像をマイクロソフトワード上で利用する、というタスクを与えた。

提示した漢字は、いずれも ISO/IEC 10646 Ext. B 集合の半偏 (U-00024614~U-00024719) の漢字から、タスク 1 : 平易な漢字 1 つと「牛」を組み合わせた漢字を 5 字、タスク 2 : 平易な漢字に含まれている部品 1 つと「牛」を組み合わせた漢字を 5 字、タスク 3 : 平易ではない漢字または部品 1 つ以上と「牛」を組み合わせた漢字を 3 字、の計 13 漢字である。3 名の被験者にはそれぞれ別の漢字を提示した。

実験の流れとしては、まずグリフウィキの URL を提示し、一通り Wiki システムやグリフエディタに触れてもらい、作字登録できる状況になったと自分で判断した上でタスクに取り組んでもらうという手順を取ることで、一般的な利用形態に近づけた。最後に Wiki システムのユーザビリティについてのアンケートを行った。このアンケートは、ウェブユーザービリティ評価スケール[17]をそのまま利用した。

課題漢字ごとのデザイン所要時間をまとめたものが表 2 である。各項目は課題として提示した漢字字形 (右側)、および実際に被験者がデザインした漢字字形 (左側) と、所要時間 (下側) である。時間の単位は秒である。タスク 1 は、難易度の低い漢字であるが、初めてのデザイン作業になるため、タスク 2 の課題との所要時間の大きな差は見られなかった反面、同じタスクであっても対象となる文字によって編集時間に大きなぶれがある。また、被験者 2 はほとんどの文字が未完成で終わっている (理由は後述) 一方、被験者 3 は及第点に達する品質の漢字グリフデザインができていた。被験者 2 の一部の課題を除くと、作業時間はおおむね 200~400 秒となり、初めてのユーザーでも 5 分程度で 1 文字を作ることができることがわかった。なお、Wiki システムの URL を提示してからタ

スク1の作業に入るまでの時間はそれぞれ27分、25分、26分であった。

また、被験者1と3は、「牛」と「牛」の違いを自分で認識して正しい漢字部品を用いていたが、被験者2については、途中で違いに気がついたものの、最後まで「牛」を利用する結果となった。

WikiシステムをMediaWikiに近づけることで、より平易な操作が期待できると考えたが、実際にはいくつかの点で操作に戸惑っている様子が見られた。実験後のヒアリングで判明したことは、Wikipediaの記事を利用する場合は多くはGoogleなどの検索エンジンの検索結果からのリンクを直接たどってくる事が多く、Wikipedia自体の検索機能を使うことは少ないとのことである。またWikipediaの利用はあくまで閲覧であって、記事を投稿した経験が無いため、新規ページを作る方法(ページを検索し、「検索されませんでした」というメッセージに付随するリンクをクリックすることで新規ページの編集画面に移動する)が分かりにくい、という意見があった。

グリフウィキの使い方については、一通りの説明ページを用意していて、それを各人が好きな時間配分で読むこととしたが、被験者2はこれらのマニュアルを読むことが苦手であり、中途半端な理解のままデザイン作業に入ったため、結果として一部の文字についてデザインを完成させられず、途中であきらめている様子が伺えた。またWikiシステムの利用法について文書に

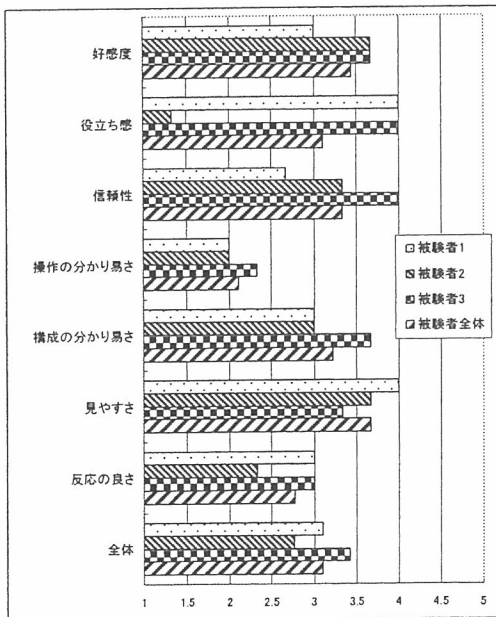


図5 ユーザビリティアンケート結果 (5:大変そう思う~1:全くそう思わない)の5段階評価。7評価軸21問

		被験者 1	被験者 2	被験者 3
タスク1	1	打打 197	扣扣 495	牧牧 766
	2	牻牻 178	犊犊 259	犊犊 332
	3	辅辅 205	犊犊 92	犊犊 116
	4	犊犊 460	犊犊 267	犊犊 85
	5	犊犊 189	犊犊 79	犊犊 252
	平均	245.8	238.4	310.2
タスク2	1	犊犊 162	犊犊 160	犊犊 110
	2	犊犊 399	犊犊 72	犊犊 645
	3	犊犊 302	犊犊 93	犊犊 148
	4	犊犊 297	犊犊 283	犊犊 354
	5	犊犊 656	犊犊 273	犊犊 315
	平均	363.2	176.2	314.4
タスク3	1	犊犊 155	犊犊 136	犊犊 126
	2	犊犊 427	犊犊 185	犊犊 470
	3	犊犊 548	犊犊 823	犊犊 332
	平均	376.7	381.3	309.3
タスク全体	平均	321.2	247.5	311.6

全体	平均	279.5
----	----	-------

表2 課題漢字ごとのデザイン所要時間 (単位:秒)

よる説明ではなく、実演による説明が望ましいといった意見が得られた。

ユーザビリティアンケート調査の結果は、図5の通りである。実験所要時間が2時間30分と長かったこともあり、評価は厳しいものとなっている。先述の通り「操作の分かり易さ」については標準以下の2.1ポイントと最も悪いが、全体平均としては3.1ポイントとなり一応の及第点が得られたものと考えられる。また、被験者2については「役立ち感」が1.3ポイントと飛びぬけて厳しい評価となった。このことから初めて利用するユーザーに対する導入の準備が非常に重要であることがわかった。

この実験の結論としては、およそ30分の準備時間を経ることにより、自分の目的とする漢字字形を5分程度で登録し、作成した漢字グリフを活用することができる、という一定の評価が得られた。

5. 3 考察

当初の目的である、フォントデザインの専門家ではないユーザーが初めてグリフウィキを使って、必要とする漢字グリフを登録し活用できる点については、一定の評価が得られた。その一方で、初めて使うユーザーに対する利用方法の説明およびデータベースフロントエンドのインターフェースには、さらなる工夫が必要であることも分かった。現在、グリフウィキの説明方法は全て文書にて用意している。これはWikipediaの手法を踏襲しているものである。しかし漢字グリフのデザインなど、視覚的要素の大きい対象については、説明と画面操作が同期するムービーを使うことで自身での自習が可能となると考えられる。

これらのことを踏まえると、集合知を利用して大規模漢字フォントを不特定多数のユーザーで制作するためにはWebシステムとしての敷居が高くなる。現時点では漢字を使いたい、という意味を持ったユーザーが主対象といえるだろう。

6 将来の構想

本研究構想を何度か発表した折、複数の研究者から、登録されているグリフに対してそのグリフが確かに利用されている典拠情報を自由に登録できると文字データベースの性格を持たせることができより学術的に価値が出るのではないかというコメントをいただいた。現在のグリフウィキの構想では、登録されている各グリフが持つメタ情報は、「関連字」と呼ばれるISO/IEC 10646に収録される文字との関連付け情報のみである。関連字は、ユーザーの目的とするグリフがグリフウィキに登録されているかどうかを判断するときに、ユーザーが関連すると考える一字を入力してもらい、その字およびシステムに内蔵する異体字データベースを参照し

導出した異体字関係にある文字群のいずれかと結び付けられている登録グリフを一覧できるようにすることで、重複したグリフを登録させにくくする目的がある。しかしながら異体字ではなく新字種については、関連する一字を指定することは難しく現状ではゲタ文字（関連字なし、の意）を指定することを許容している。

特に今後発見の可能性が高い日本の古典籍における新字種をグリフウィキに登録する際に典拠情報を同時に入力し、また既に登録されているグリフが別の出典において用いられている場合にその情報を重ねて登録することにより、その文字の新字種としての確立の度合いが増すことになる。このようにして収集した新字種のデータは新たに文字コードに収録要求を出す際の基礎資料として有益であると考えられる。

このような典拠情報などのメタ情報をグリフウィキでどのように扱うかについては、さらに議論が必要である。その理由としては、典拠情報等を固定した書式で記録する方式が運用可能であるかどうか、逆に自由書式での情報記述を許容した場合に検索や集計で問題が出るかどうかの検討が不足していると考えためであり、文字データベースの研究者からのアドバイスを求めるべきである。

例えばグリフウィキはあくまでグリフ情報を登録する手段と考え、典拠情報や他のメタ情報については、CHISEプロジェクト[18]が提唱・公開しているCHISE文字情報データベースとの連携で実現することも考えられる。CHISE文字情報データベースは、文字1つ1つをオブジェクトと見なし、そのオブジェクトに対してさまざまな情報（素性）を付与することでその文字を表現する概念モデルを用いている。一つの文字オブジェクトとグリフウィキのグリフを結びつけることで、登録されているグリフに対してさまざまなメタ情報を付与することができるものと思われる。

7 おわりに

現在グリフウィキはシステムを公開し、試験運用を実施している段階にある。不特定多数のユーザーによる共同管理やフォント制作が実用的となるかどうか、実際に漢字を扱う研究者をはじめとするユーザーに受け入れられるかどうかは今後の検証を待つ必要がある。グリフウィキが将来的に漢字字形処理問題の一助となれば幸いであり、そうなるべく今後も研究を進める所存である。

参考文献等

- [1] <http://www.mojikyo.org/>
- [2] 東京大学多国語処理研究会 <http://www.l.u-tokyo.ac.jp/GT/>

- [3] Windows NT 漢字処理技術協議会, Windows NT 拡張漢字処理仕様書, 第 2.1 版, 1998.
- [4] Hideki Hiura, Eric Muller 「 Ideographic Variation Database 」 , Unicode Technical Standard #37, 2006.
- [5] Adobe-Japan1-6 Character Collection for CIDKeyed Fonts, Technical Note #5078, 2004.
- [6] ISO/IEC 10036:1996, Information technology - Font information interchange -- Procedures for registration of font-related identifiers, 1996.
- [7] <http://c2.com/cgi/vwiki>
- [8] <http://wikipedia.org/>
- [9] GNU Free Documentation License <http://www.gnu.org/licenses/fdl.html>, 第 1.2 版, 2002.
- [10] <http://www.mozilla-japan.org/js/spidermonkey/>
- [11] <http://www.imagemagick.org/>
- [12] <http://fontforge.sourceforge.net/>
- [13] <http://www.sqlite.org/>
- [14] <http://www.hyuki.com/yukiwiki/>
- [15] <http://kanji-database.sourceforge.net/>
- [16] ISO/IEC 10646:2003 / Amd. 1:2005, Information technology - Universal Multiple-Octet Coded Character Set (UCS) AMENDMENT 1: Glagolitic, Coptic, Georgian and other characters
- [17] 仲川ほか, ウェブサイトユーザビリティアンケート評価手法の開発, 第 10 回ヒューマンインターフェース学会紀要, pp.421-424, 2001.
- [18] CHISE プロジェクト : 文字に関するさまざまな知識のデータベース化 <http://kanji.zinbun.kyoto-u.ac.jp/projects/chise/char-data/>