従来、「文字」と呼ばれる、限定されたコンピュータの中でのみ利用者が自由に文字を定義できる仕組みを用いてこの問題を解決してきた。文字はインターネットに代表されるネットワークを介した情報交換には適さないため、新たな解法が模索されてきた。なお、本稿で紹介する「文字」は、情報システムやネットワーク上で利用される文字（外字）を指す。

二松学舎大学近代史研究所にて、情報システムの対話性を考慮した新しい文字コードの開発を行った。この文脈において、コンピュータについての言及は省略されるものとする。したがって、次に示す各文脈における文字の定義を含め、ここに報告することとする。

二、問題のためには文字コードではない

コンピュータの拡張型文字コードは、既存のコンピュータの文字コードを拡張する形で、「JIS日本漢字コード」と「小文字」が新たに規定される。従来の文字数の拡張が可能であるが、現状では拡張されない。
た文字集合が広く活用されているとは言いがたい。六三五種類の漢字には、常用漢字に収録される一九五字が全て含まれるため、一般的な日本語を表記することが可能であるが、日本漢文学研究者にとって入力できる漢字の種類が足りないということは、少しコンピュータを使うだけで遭遇する問題である。

現在は、国際的に標準化された文字コードである "ISO 10646" （一般的にはユニコードという呼称で知られている。）がコンピュータで処理できる状況であり、実際に活用している利用者は少ないと言われるが、マイクロソフト社のウィンドウズ・ビスタには、この五万字の漢字を利用できる環境が整っている。

しかし、言語情報の対応に関する審議を経て、今後も収録漢字が増える方向にある。それに伴い、言語情報が交換するための文字コードを規定するための文書は,"ISO/IEC 10646"に対応する規格である "JIS漢字コード" が規定されている。一つの字体の表記の実現として、デザインの差に基づく複数の字形が考えられるが、この規格はそれぞれ互いに区別しない。一つの字体の字形の選択において、ユーザーの差に基づく複数の字形が考えられるが、この規格はそれらを互いに区別しない」ある。また、"ISO/IEC 10646" 国際規格を日本国内で実装させる規格である "JIS漢字コード" が規定しようとするものではないと明記されている。具体的な例を挙げると「砲」の大文字の日本語を「大刀」の小文字の日本語に変換するのではなく、そのような状況は大変不便であるが、細かい字形の区別は文字コードの範疇ではない。
漢字字形情報管理システムの構築と提案

（二）超漢字（トロンOS）

现在普及しているウィンドウズでは別のOSとして多漢字処理を実現することができるが、トロンで呼ばれるOSの仕様に準拠した「超漢字」という名前の製品である。

文書の文字は、本来の漢字に対して、されている。トロンコードは、その後、台湾や香港で普及しているとされる。

対象となる形で設計された独自の文字コード（トロンコード）を利用し、その後製品では削除されたが、大漢和辞典見出し字などが追加され、さらにG T書体と呼ばれる大規模文字集合が加わった。

G T書体のものはウィンドウズで
利用できるフォントであり、無償配布されている。
トロンドコードおよびTADと呼ばれる独自のデータ形式は他のO/Sとの情報交換において変換作業が必要とするため、研究者の対外公開が求められる今日において利用するメリットは小さく、O/Sの普及度も低い。また、G.T書体（トロンドコード）を用いたテキストデータもインターネット上で公開されているものではないが、一〇〇七年から一〇〇七年の間に合計で九回の収録文字集合更新の報告がなされている。収録にかかる期間についての説明は見受けられないが、パーソナルメディア社では独自の文字追加を含めた多数漢字処理システムの提供を行っているが、対象は自治体などであり、費用や事業規模の説明から判断すると、研究機関を対象としているとは考えにくい。

X.K.P

ウインドウズNT拡張漢字処理（略してX.K.P）とは、情報システム開発により、自立したデータベースをウインドウズの内部漢字コードやJIS漢字コードに新たに採用されたユーニコードに対応させることを目的としたものであり、現在では直接活用する機会は少ないが、ネットワークで文字を共有するユーニコードを基盤とした処理システムの例として紹介すべき技術である。
ライン化のために戸籍に利用できる文字集合を整理したものである。この文字集合を各種自治体が共有することによりシステム

五 ISO/IEC 10636

文字字形（グリフ）を登録し、同一の番号を付与する機構をもった国際標準规格である。利用者は手続きを行うことで登録

特に自由に文字字形を登録することが可能で、各字形に対して付与される番号は構造記述が可能なXML文書において活用

し、文書の役者者がその文字の具体的な文字字形を参照することが可能である。登録簿の一冊はインターネットで公開されているわけではない。

この番号の文字字形に該当する、といった具合である。登録簿の一部はインターネットで公開されている。

デジタル文書内では字形を書き分ける方法として、

ユーザーがテキストを入力するときにISO/IEC 10636の文字字形から自分の意図する文字を探す、といった活用はできない。

状況ではISO/IEC 10636が広く活用されているとは言えないが、今後字形識別手段として普及する可能性もある。

六 ユニコード漢字異体字データベース（IVD）
の方法で登録申請を行い、データベースに登録されたものが利用できる。現在は各文字に対して二百四十個の識別記号が用意されているが、要請に応じた記号数の拡張についても言及されている。現時点でデータベースはまだ存在しないが、実質的に広く普及している文字コードであるユニコードにおいて正式に規定された字形指定方式であり、将来OSやアプリケーションでの標準化が期待されている。

（七）日本語外字センターの提唱

小山氏が提唱する日本語外字センターとは、住民基本台帳、ネットワークの人名処理における字形に関する外字の表記を日本語外字センターが管理し、住民基本台帳ネットワークの個人向けの字形処理を担当するものである。字形の内容は日本語外字センターが提唱したものである。

 '=': Japanese text

---

- 184 -
任せるという方法にたどり着いた」と述べている。その方法として「状況をインターネットにより常時情報公開すること」で利用者が評価することにより恣意的な判断を防ぐことができる、とされている。このように、文字字形で名前を付与して識別を行うこと、およびインターネットによる衆目監視が字形管理（異字形の同定）に有効ではないかという二つの意見は注目すべきである。

四 ユーザー本位である字形管理環境の欠如

以上のように、標準的な文字コードの範疇を超えた漢字情報処理が必要な自治体や企業のニーズを反映した字形管理システムや、文字研究者が活用する大規模字形フォントなどによりある一定の解決が図られてきた。ところが日本漢文学研究者の中でもパソコンに興味をもっていた人の中には、NEC社のパソコンであるPC-8800シリーズを活用していた人も多いのではないだろうか。その頃は文字コードに不足する字形については各ユーザーが自分で字形を管理していたケースが多く、筆者もいくつかの字形フォントを所有していたことを記憶している。それが現在では他者に提供する外字フォントを使うなどの受動的な立場に移行させざるを得なくなった要因については次の二つが考えられる。

これららの要因は密接に関連している。すなわち、文字字形のデータ形式で現在主流のアウトライン形式は、表現豊かで精
細な印刷が可能である反面、文字字形のデザインにおいて習熟した技術を要するものであり、複数のフォント編集ソフトウェアが提供されているが、フォントデザインの非専門家にとっては従来の高いものとされている。また、既存の漢字フォントがソフトウェアで利用されることが多いが、一部のフォントでは使用制限が設けられているため、一部の案件で既存の漢字フォントを利用することがある。このように、漢字フォントの利用が悪化していくと考えられる。

公開が困難であるため、外字利用の必要性があるという理由で、対戦的ないし対戦的ないポータルサイトに公開される研究分野において、将来の公開可能性があるデジタルテキストを作成する際、必ずその漢字字形の利用が必要となる。このように、デジタルテキストの利用は、既存の漢字フォントを利用することでニーズを満たすはずであるが、それでも足らない漢字字形というのではゼロにはならないと考えられる。

このような外字の利用において受動的な立場にある現状を鑑みると、必要に足る漢字字形を自由に・容易に利用できるのでなければならず、そこで能動的な外字字形情報管理環境の構築を目的として、先述の二つの要因を解消する漢字字形情報管理システムを提案する。この管理システムは、ウェブ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、外字データの登録もインターネット上で行う。
漢字字形情報管理システムの構築と提案

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、HTMレクメンメントから画像ファイルを参照する）など、意図的に独自の字形集合として用いる場合がある。

既存のデータベースに字形を登録する際に、個々の漢字字形は任意の名前を付与することで識別を行い、命名規則は特に设定しない。また、データベースに存在する文字を既存のデータベース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ている。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であり、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は管理者がいないデータベースであるため、同定に伴う審査を可能とする。ただし、この場合でも登録された文字の登録に意図的な判断をせずに自由に登録できることが可能であり、同時に自分のデータベースであるため、同定不可能に変化することを可能とした。
インターネットにも深い根があった。ウェブ・ユニバーサル・マーク（W3C）が定めた超文書構造語法（HTML）は、インターネット上で情報を共有するための標準である。HTMLの特性は、コンテンツが閲覧者に送られた後で容易に変更できることで、パーソナル・コンピューターやスマートフォンでも利用可能である。これにより、関連するウェブ・コンテンツは、単に見た目が同じであるだけでなく、機能的にも関連付けられていて、ユーザーエクスペリエンスを向上させている。

報告書の内容は、事実に基づいており、これにより、歴史的な背景と技術的な進化が詳しく説明されている。さらに、示されたデータは、今後の進化に期待をもっており、この分野の研究者や関心をもつ人々に役立っている。
モードが発表した『伽蓝とバザール』と題する著書において、モードが『バザール』におけるコンピュータを含むソフトウェアの構築と提案的方法論であることを強調しています。ソフトウェアの設計図にあたるプログラムソースを常に公開し、バザールのあるソフトウェアを構築することを可能とするものであり、インターネット上の開発環境を用いることで、新しいウェブ利用形態の一つであり、現在注目されているモデルである。ウィキシステムはこのモデルを応用したものであり、インターネット上の沢山のユーザーを集めて大きな作業を行っている。漢字形情報管理システムでは、このウィキシステムの二つの特徴である「共同・協同作業」という方法が新しいウェブ利用形態の一つであり、現在注目されているモデルである。自由・無償なフォントの協同制作のツールとしての活用も考えている。すなわち、『KAGE』は中国大陸デザインおよび台湾・香港デザインの二種類の漢字フォントが提供されている。先述の通りマイクロソフト社のジャンドウズ・ピスタでは中国大陸デザインおよび台湾・香港デザインの二種類の漢字フォントが提供されている。
難である。また輪郭線による字形データは、線形的な拡大や縮小を施した場合に、筆画の太さに対しても変形がかかる。

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、漢字字形を筆画ごとの骨格情報を積層で表現する、グリフォーマの中間表現形式であるが、KAGEシステムでは筆画は線の種類（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情報を記述する。中心線は、直線や曲線、そして点線の位置情報を表現する。点線から四点の位置情報を用いて新しい漢字字形を表現する。筆画の形状は、筆の入力、跳ね、墨の筆画へ接続、止め、といった数種類の形状を頭部と尾部の二種類について記述する。これらを元に、字形の特徴を持つアウトライン形式、データヘ変換プログラムに変換する。漢字字形の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も可能である。漢字単位で表現が可能であり、漢字単位の情報は八単位で表現が可能である。これにより、テキストデータを表記する。

KAGEシステムの漢字字形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、先述のウィキシステムの記事分を管理システムでは漢字一行字形と見え、KAGEシステムによって表現された漢字形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ウィキ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形データを数値の記述でデザインするものは難しく、座標点をマウスで操作することでデザイン可能なグラフィ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フラッシュコントロールとして実装しているため、管理システムを操作するウェブ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。

（三）デーティライセンスの決定

管理システムでは、漢字形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、著作権などの制限を図るという観点からも否定的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、そのファイルをディスクに保存する必要がある。

ディアの独自の解釈を加えたものが適用されている。GFDLライセンスの根拠であるのはFSFが提唱しているウィキペディアでは、投稿する記事を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ている。GFDLライセンスとして参考としたウィキペディアでは、自己のデータを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。漢字情報のフォントファイル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、そのファイルをディスクに保存する必要がある。
ア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、作った文書ファイルに対して他人による自由利用の保証が求められることになる。これは現実的ではない。もともとFSFが別に推奨している、ソフトウェアを対象とするGPLライセンスでは、ソフトウェア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ライセンスでは、フォントファイルにライセンスを適用したことも可能である。という解釈がなされている。いずれにしても、ユーザがフォントを利用したことも可能であること、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ことを考えた。そこで、管理システムとしてのすみのでフォントを利用したことも可能である。これはを利用したまま利用の自由を保証し、そのデータを再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。−2−

注意が求められるに当たり、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ことを考えた。そこで、管理システムとしてのすみのでフォントを利用したことも可能である。これはを利用したまま利用の自由を保証し、そのデータを再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。
リフがクリフヴ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文字を入力してもらい、その文字およびシステム内に内蔵する異体字データベースを参照し、発出した異体字関係にある文字群のいずれかを結び付けられて登録される。具体的には、新字種については、関連する文字を指定することなく、通常の入力法で新字種を登録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漢字をはじめとする古典籍における異体字をクリフヴィキに登録する際には、様々な異体字を同時に対応することが必要である。その理由としては、新字種の確立の度合いが増すにつれて、その文脈において用いら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が重ねて登録されることになるためである。このようにして収集した新字種のデータは新たな文字コードに取扱われるように設計されているが、その場合の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られるためで、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、クリフヴィキはあくまでグリフ情報を登録する手段であり、登録されているグリフに対してさまざまなメタ情報が付与することができるものと思われる。
漢字字形情報管理システムの構築と提案

五 (四) グリフウィキ表示例

六 (二) 版本表現型全文テキスト表示への適用例 (枠内字が外字)