利用者認証に用いる識別子の決定方法に関する考察

のぎ田, めぐみ
九州大学情報基盤センター : テクニカルスタッフ

笠原, 義晃

伊東, 栄典
九州大学情報基盤研究開発センター : 准教授 : 情報学

鈴木, 孝彦
九州大学情報基盤研究開発センター : 准教授 : 情報学

出版情報：電子情報通信学会技術研究報告, pp.67-72, 2006-12. The Institute of Electronics, Information and Communication Engineers
バージョン：
権利関係：

社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE

利用者認証に用いる識別子の決定方法に関する考察

のぎ田 めぐみ　笠原 義晃　伊東 栄典　鈴木 孝彦

九州大学情報基盤センター

〒812-8581 福岡県福岡市東区箱崎 6-10-1
E-mail: {megumi, kasahara, itou, suzuki}@cc.kyushu-u.ac.jp

あらまし 利用者認証を必要とする情報システムが増えている．認証のためには利用者を特定する識別子が必要になる．本論文では識別子を決定する方法について考察する．決定方法を検討する際の要件として，管理性，利便性，耐性の 3 つを挙げる．識別子の作成に関して通し番号，ランダムな文字列，部分毎に何らかの意味を持たせた文字列，行政など別の枠組で設定された識別子の流用，氏名からの自動生成，利用者が希望する任意の文字列の 6 種類について検討する．

キーワード 利用者認証，識別子，身元情報

A Study of Identifier Naming Conventions Suitable for User Authentication

Megumi Nogita　Yoshiaki Kasahara　Eisuke Itoh　Takahiko Suzuki

Computing and Communications Center, Kyushu University

6-10-1 Hakozaki, Higashi-ku Fukuoka 812-8581 JAPAN
E-mail: {megumi, kasahara, itou, suzuki}@cc.kyushu-u.ac.jp

Abstract User authentication plays an important role in information services such as e-mail, PC login, and web databases. They need an identity database to realize user authentication, and identifiers to refer the database. Identifier is a special name to uniquely identify an entity (a user, in this case). We considered six possible identifier strings: serial number, random string, concatenated string, localization of global ID, derivation from real name, and self selection. We evaluated those types of identifiers with three aspects: manageability, usability, and robustness.

Keyword identifier, identity information, user authentication

1. はじめに

利用者認証を必要とする情報システムが増えている．利用者認証や認可を要する情報システムでは，各利用者の身元情報データベースが必要である．ある情報システムを利用者が使う際，身元情報と利用者を紐づけするため，身元情報データベースの情報を参照する識別子が必要となる．会社などの組織では身元情報が先にあり，それを用いて各利用者へ識別子を付与する事になる．また，一般のウェブサービスでは利用登録する際に同時に作成されることが多い．いずれにしても，利用者の識別子は利用者全体で一意でなければならないため，衝突回避など何らかのルールが必要となる．システムの利用法によっては，識別子として何らかの意味を持った文字列（利用者氏名，役職等）を採用する場合もある．

昨今，第三者による成りすましや改ざん，詐称などのネット上の犯罪が問題になっている．識別子の決定方法により，このような犯罪に対する情報システムの耐性が影響を受ける場合があるので，識別子の決定は慎重に行わなければならない．

現在，多くの大学で統合認証基盤の再構築がされている[1]．筆者らが所属する九州大学の場合，現状では各部局が独自に利用者の身元情報を管理している．その関係を図 1 に示す．

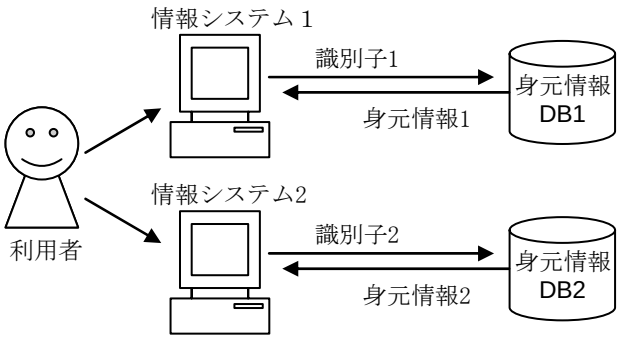

図 1 複数の識別子管理

この方法では利用者側から見ると，同じ大学内のシステムでもシステム毎に異なる識別子を用意していることになる．その為，どの識別子がどのシステムのものか混乱する場合がある．これを改善する為，大学全体での識別子の統合が求められている．

大学内で統一した情報基盤を構築するには，システムの管理性，利便性を考慮すると，識別子を学内で統一するほうが良い．その為には，識別子の決定方法が問題となる．本論文では，情報システムの利用者認証に用いる識別子の決定方法について考察する．

2. 身元情報と識別子について

この章では本論文で述べる身元情報，識別子，およびそれらを使用する情報サービスとの関連性について述べる．

2.1. 身元情報

身元情報とは利用者がサービスを受けるときに必要な個人情報の集合体を意味する．例えば，パスワードや住所,電話番号,所属などの集合体が挙げられる．どのような属性を身元情報に用いるかは，サービスによって異なる．例えば，クレジットカードのシステムでは，カード番号や氏名などはサービスを利用する上で必要となりうる基本的な情報なので，身元情報といえる．一方，カードの利用明細は身元情報として扱われない．但し，本論文では身元情報の詳細については議論しない．

2.2. 識別子

識別子は，個人の情報の集合体を参照するポインタの役割を持つ．図２に示すように，身元情報データベースは個人情報の集合体と，それを指定する識別子から構成される．識別子は一意でなければならない．

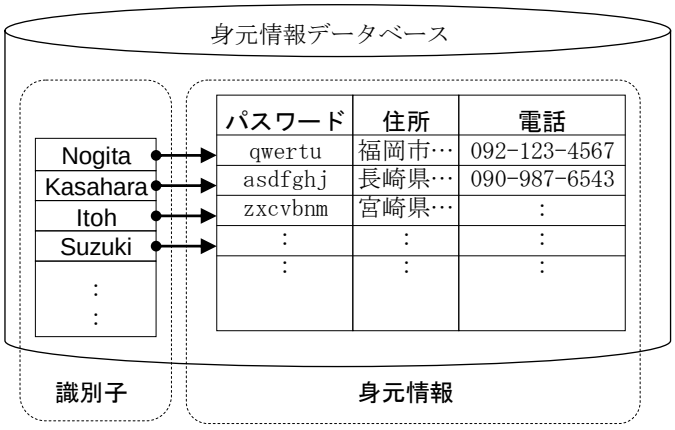

図 2　識別子と身元情報

2.3. 情報システムにおける識別子および身元情報の関係

本論文が対象とする情報システムとはネットワークなどを介して様々な情報サービスを行うシステムである．例えば，筆者らが所属する九州大学では，学生には履修登録などの学務処理が，教職員には予算管理を行う財務処理が web を介して提供されており，これらが本論文で対象とする情報システムである．

情報システムが利用者向けにサービスする際，図３に示すように，識別子を用いて利用者の身元情報をデータベースから参照する．入力された利用者識別子が適切な文字列かどうかの判断を行う情報システムもある．利用者認証を行う場合は，データベース内に格納されている（ハッシュ化された）パスワードを参照する場合もある．また，利用者に個別のサービスをするために，利用者の所属部局や学年などの．他の属性を参照する場合もある．

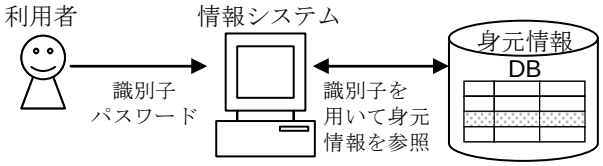

図 3　情報システムと身元情報

九州大学では現在，図１に示したように，身元情報を情報システム毎に管理している．そのため，ある利用者の身元情報が，複数の身元情報データベースに登録されている．また，情報システム毎に異なる識別子が用いられている．こうした場合，いくつかの問題がある．

利用者側から見ると，情報システム毎に異なる識別子が用いられているため，複数の識別子（とパスワード）を覚える必要があり，利便性が低下している．シングルサインオンによる認証作業の軽減も実現できない．情報システムの管理者側にも問題がある．一人の利用者の身元情報を情報システム毎に個別管理していることになり，身元情報の管理コストも増大してしまう．識別子の配布作業にもコストがかかる．更に情報管理の責任者である大学本部から見ても，複雑化によるコスト増，利便性低下による非効率化，各システムのセキュリティレベルの差異による情報漏洩の危険性といった問題がある．

3. 識別子の決定方法で考察するべき用件

識別子の決定や利用に関しては３つの立場が考えられる．情報システムや身元情報データベースを管理運営し提供する提供者，これを利用する利用者，そして第三者である．特に，第三者として情報システムへの攻撃を行う悪用者への対策を考慮する必要があり，そのため悪用者からの攻撃に対する耐性について考察す

る．以後それぞれの側面を情報システムからみた「管理性」，利用者からみた「利便性」，悪用者からの攻撃に対する「耐性」と呼ぶ事とする．

3.1. 管理性

管理側の立場では，情報システムや身元情報データベースのシステム構築，情報管理，運用，保守のコストが主に問題となる．情報システム側で利便性を維持する為の管理コストも問題となる．例えばクレジットカード番号の様に桁数が多く憶えにくい識別子は，利便性のために識別子を記載したカードの配布が必要になり，そのためのコストが必要になる．

どのような文字列を識別子として決定するかにより，初期導入・管理運営・保守のコストは異なってくる．例えば，4.1節で述べる「通し番号」を識別子に用いる場合，システム構築の難易度が低いため，初期導入コストが削減される．

3.2. 利便性

識別子の役割として，第一に利用者の身元情報をデータベースから参照するためのポインタの役割がある．データベース参照に用いるポインタであれば，一意な文字列であれば良い．しかし，利用者が識別子を用いる場面における利便性を考えると，記憶の容易さ，入力の容易さ，判読のしやすさ，誤指定のしにくさ，といった側面も考慮しなければならない．

まず，利用者認証を行う情報システムでは，利用開始時に識別子（およびパスワード）の入力を要求するものが多い．そのため，容易に記憶および入力できる識別子が望ましいことになる．

他にも，利用者へのメッセージ送信の宛先としての利用も想定する必要がある．電子メールアドレスの@マークより前の文字列に利用者識別子を使う場合がしばしば存在する．他にも，チャットやTV会議の宛先として用いる場合も多い．識別子をメッセージの宛先として用いる場合は，判読のしやすさや，誤指定のしにくさ等の面を考慮すべきである．

記憶の容易さ，入力の容易さ，判読のしやすさ，誤指定のしにくさ，といった問題を回避するためには，利用者による直接入力を行わせないようにする必要がある．そのためには，ICカード等の記憶媒体に識別子を格納し，その媒体を介して情報システムを利用するようにする必要がある．

3.3. 耐性

管理者と利用者の他に，悪用者から攻撃に対する耐性を考慮する必要がある．識別子の決定の際に耐性の面から考慮すべきことは，利用者識別子の推測が容易かどうか，という点だけである．耐性としては，認証時に用いる秘密情報（パスワードなど）の脆弱性などがあるが，これは本論文の対象外である．

識別子を機械的に推測して攻撃を行う方法としては，総当たり攻撃（ブルート・フォース攻撃）と辞書攻撃が代表的である[2]．他にも，別の個人情報（氏名など）からの推測や，物理的に利用者の背後から識別子を盗み見るといった攻撃もありえる．

ある識別子の決定方法を選択すると，識別子として使用可能な全文字列の集合である識別子空間が決まる．また，その空間内の，どの部分の文字列を使うかの偏りも決まる．識別子空間の大きさと，実際に利用される識別子の存在率や偏りにより，総当り攻撃や辞書攻撃への耐性が決まる．例えば，通し番号を識別子に用いると，初期値から値を増加させていくだけで全利用者の識別子を推測できる．そのため，総当たり攻撃に対する耐性は低い．

4. 識別子の作成について

識別子を作成するに当たり以下に示す6つの方法について検討した．

(1) 通し番号（シリアルナンバー）
(2) ランダムな文字列
(3) 利用者が希望する任意の文字列
(4) 行政など別の枠組で設定された識別子の流用
(5) 氏名からの自動生成
(6) 部分毎に意味を持たせた文字列

以下にそれぞれの方法について管理性，利便性，耐性の面から考察する．

4.1. 通し番号

通し番号とは初めから終わりまで一続きの番号をいう．例えば，0001の次は0002，といったものになる．

（１） 管理性

通し番号を識別子に利用する場合，身元情報データベースに利用者が追加されるたびに次の番号を割当てる事になる．そのため衝突回避・解決のためのコストはかからない．データベースシステムは，エントリ毎に自動で通し番号を割当てる機能が搭載されている．そのため構築コストも低く抑えることができる．

実際には通し番号だけで識別子を作ることは少ない．4.6節で述べる部分文字列の一部に通し番号を用いる方法が多い．

（２） 利便性

通し番号では，1つ違いの番号も有効な識別子である．識別子を認証に用いる場合，誤った識別子を入力すると，他人の身元情報を照合する事になる．また，識別子を宛先として利用する場合，気づかずに異なる相手に通信してしまう事になる．また，長い通し番号は記憶しづらい．

（３） 耐性

総当り攻撃が容易であるため，耐性は低い．

（４） 実例

オンラインショップの予約番号などで使われる.

4.2. ランダムな文字列

無作為に作成された文字列のことを言う．通常，意味のある文字列になる可能性は低い.

（１） 管理性

ランダムな文字列を識別子にする場合，システム側で識別子の衝突回避をする必要性がある.利用者が識別子を忘れても,空間を広く取っておけば識別子の再割当てが容易にできる.その為,管理コストを抑えることができる.

（２） 利便性

ランダムな文字列は記憶しにくいため，入力の際の利便性は低い．しかし，誤入力による他者の識別子入力の危険性は，識別子空間が広くなるため低くなる．一方，ランダムな文字列は言語的な意味がないため，入力誤りを利用者が気づかないという問題もある.

（３） 耐性

ランダムな文字列では識別子空間を，使用する文字種数の文字列長による冪乗にできる．そのため，実利用する識別子空間を最大限にできる．また空間内の分布も均等にできる．そのため，総当りの試行回数に対する存在識別子の発見率が低くなるため,耐性は高い.

（４） 実例

クレジットカード会社提供の web 情報サービス用における識別子として利用されている.

4.3. 利用者が希望する任意の文字列

利用者が任意に希望する文字列を識別子として利用する.

（１） 管理性

利用者が希望する任意の文字列を識別子とする場合，他の決定法よりも複雑な衝突回避の仕組が必要となる．衝突した際には，利用者が指定した任意の文字列に似ていて衝突しない識別子の候補を挙げるなどの回避策が必要である．また，使用してはいけない文字列などの処理が必要となってくるため多少の処理コストがかかる.

（２） 利便性

利用者自身が選択した文字列なので憶えるのが容易であり，利便性が高い．但し，希望する文字列が既に利用されていて取得できなかった場合，その人にとっては利便性が低下する事になる.

（３） 耐性

総当り攻撃には強い．しかし，一般的な単語を利用しがちである為，辞書攻撃には弱い.

（４） 実例

Hotmail，Yahoo メール，mixi など，登録型の情報システムで利用されことが多い.

4.4. 行政など別の枠組で設定された識別子の流用

行政などで利用されている識別子を，そのまま内部の情報システムでの識別子として用いる．例えば，情報システムの利用者識別子に，パスポート番号を用いるような方法である.

（１） 管理性

行政など別の枠組で設定された識別子を流用する場合，既に識別子が決まっている為，識別子を割り当てる手間を省くことができる．しかし，情報システム側で自由に識別子の変更ができない為，管理はしづらい．また，流用元で識別子の変更がなされるとシステム側でも識別子の変更を余儀なくされる.

更に，外部の枠部で決められた識別子を持たない人が存在する場合に問題となる.図４の(A)の場合のように，ローカルな情報システムの利用者 L が，行政など別の枠組で設定された識別子の空間 G に収まる場合は良い．しかし，図４の(B)のように，収まらない人が居る場合は何らかの対策が必要となる．例えば，パスポート番号を流用する場合，パスポート未取得者や外国人の扱いが問題になる.

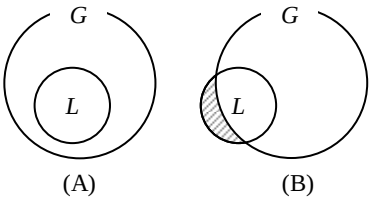

図 4　流用の際の問題点

また、流用の可否も問題になる．住民票コードは日本人全員が取得可能であるため,図４の(B)の問題は発生しづらく，流用に適している．しかし，住民基本台帳法によると住民票コードは「何人も，自己と同一の世帯に属する者以外の者（以下この条において「第三者」という.）に対し，当該第三者又は当該第三者以外の者に係る住民票に記載された住民票コードを告知することを求めてはならない」と記載されている[3]．そのため，他のシステムでは流用できない.

（２） 利便性

行政など別の枠組で設定された識別子を，普段からよく使っていれば利便性が高くなる．しかし，普遍的かつ流用しやすい識別子はあまり存在しない.

（３） 耐性

総当たり攻撃や辞書攻撃への耐性は，流用元の識別子に依存する.

（４） 実例

パスポート番号や免許証番号の流用が想定される.

4.5. 氏名からの自動生成

自分の氏名をそのまま識別子とする.

（1） 管理性

氏名から識別子を自動生成すると，ある程度の規模の組織の場合，同姓同名問題が発生する．その場合，衝突回避の為のルールとシステム作りのコストが発生する．単純に名前の後に数字を付与する方法が一般的だが，同姓同名に順序が発生し，それが問題になる場合もある．また，結婚などによる氏名の変更があった場合,旧姓をそのまま利用することもあるが,場合によっては識別子の変更が発生するという問題もある．

（2） 利便性

自分自身の氏名を識別子とする為，忘れる事がほとんど無く，利便性は非常に高い．識別子を宛先として利用する場合，同姓同名の他者への通信が発生する可能性がある．

（3） 耐性

総当り攻撃は困難である．しかし，名前がわかってしまうと識別子の推測も容易である．また，名前辞書による辞書攻撃への耐性は低い．

（4） 実例

企業の出退勤システムでの利用例がある．

4.6. 部分毎に意味を持たせた文字列

桁数ごとに何らかの意味を持たせた文字列を意味する．例えば先頭 1 桁に部局を意味する文字列を使用するなどした文字列である．

（1） 管理性

利用者が介入して識別子を作成すると衝突回避が必要となる．部局コードの決定などの意味を持たせる為の管理コストが発生する．部分文字列に利用者の属性を示す部分を採用する場合，属性の変化（人事異動等）に対応する為の管理コストが大きい．これについては5.3節で述べる．

（2） 利便性

1 桁目が部局を表す，といったルールを使用することで利用者の属性を判別することが可能になる.また，記憶も比較的容易である．

但し，ルールを知ることで識別子と利用者との対応が容易にわかる為，利用者のプライバシー保護が困難になる．

4.1節，4.2節，4.3節で示した方法と組み合わせて用いることが多い．

（3） 耐性

識別子のルールを知っている場合，総当り攻撃が可能な場合がある．

（4） 実例

九州大学の学籍番号で利用されている．

表 1 各方法の比較

	管理性	利便性	耐性
シリアル番号	◎	×	×
ランダムな文字列	○	×	○
希望文字列	×	○	△
行政など別の枠組の流用	△	△	△
氏名からの自動生成	×	○	△
意味付けされた部分文字列の集合	△	○	△

表 1 に，6 つの方法を管理，利用，耐性からみた評価を記載する．どの方法も一長一短がある．サービスの特徴によってどのように識別子を付けるかを決めるべきであろう．

5. その他識別子の考察

5.1. 文字種

4章において識別子は文字列であると述べた．実際にはどのような文字種で構成するかという事も議論の対象となる．例えば，4.1節において通し番号を利用した場合，これは数字のみで構成された識別子となる．

使用する文字種によって，同じ識別子長でも空間の広さが異なる．その為，利用者にとっての利便性（判読のしやすさ，覚えやすさ，入力のしやすさ）や，総当たり攻撃への耐性などに違いが出る．

通常よく使われる文字種としては，以下に示す a～e の 5 つがある．後の物ほど空間が広くなり，また言語的な意味のある文字列を作る事が容易になる．その一方で，形状の酷似した文字の取り違えなどの問題も発生しやすくなる．

a.数字，b.英字，c.英数字，
d.英数字＋記号文字*1，e.多言語文字
（*1: ASCII コードで表現可能な物）

情報システムが身元情報も保持するような独立したサービスの場合，識別子はその情報システムで扱える文字種を採用すればよい．しかし，統合認証システムなど，組織全体で身元情報を一括管理し，各情報システムに対し認証・認可処理を提供するシステムを用意する場合には，身元情報を保持する統合認証システム側と，それを利用者認証のために利用する情報システム側で，扱える識別子の文字種に齟齬が発生する可能性がある．この場合，情報システム側のソフトウェアの修正や，識別子の一意性を保持したまま別の文字種集合に変換するラッパーの作成が必要となる．

英字については，大文字と小文字を区別するシステムと区別しないシステムが存在する．それらのシステムの認証を統合する場合には衝突解決のために何らかの措置が必要となる可能性がある．

多言語文字では，文字種を ASCII コードから拡張し，言語固有の文字種を利用可能とする．例えば日本であれば日本語文字（仮名・漢字）を識別子として使用できる事になる．Windows XP 等は識別子が Unicode で多言語化されているため[4]，日本語名を利用している利用者は多い．可読性は上がるが，同時に誤記の可能性も高まる．多言語文字には，同じ字形を計算機で表現するための文字コードが多くの言語で複数あるという問題がある．例えば日本語文字については代表的な文字コードだけでも 4 種類あり（JIS・シフト JIS・EUC・Unicode），互換性がない[5]．このためコード変換が必要となりシステムを複雑化させる事がある．特に認証システムと情報システムが独立の場合は，齟齬を拡大させ対応コストを増大させる事になる．

5.2. 固定長と可変長

文字種と並んで考慮するべき事項として，識別子の文字列長がある．文字列長には，固定長か可変長かという要素と，文字数を最小長・最大長の要素がある．

固定長の場合，識別子はシステムで指定された文字列長より長くなる事も短くなる事も許されない．このため，文字種と文字列長により空間の大きさが決定される．クレジットカード番号は固定長の識別子と言える．固定長の識別子はシステム構築時に実装が容易になる利点がある．しかし，想定していた範囲の識別子を消費した場合（名前空間内におさまらなくなった場合），大規模な改修が必要となる．過去に予想外の利用者急増で携帯電話の番号の桁数が足らなくなり，いっせいに桁数を増やすという事が発生した[6]．システム改修に加えて利用者全員への周知が必要となり，非常に大きなコストがかかる．

可変長の場合は，通常最小長と最大長がシステムで決定されている．歴史的な理由などにより，識別子の最大長が 8 文字など短いシステムが現存していることがある．このようなシステムの識別子を統合する場合に，最大長より長い識別子を通知なく短縮される事による衝突の危険性が発生する．

5.3. 所属が変わる場合

多くの場合，利用者の属性（所属部局等）が異動等により変化した時には，身元情報に含まれる属性情報が変更される事になる．識別子が身元情報の内容と無関係に決定されていれば，異動者も情報システムの利用が継続できる．

しかし，4.6節で述べたように，利用者属性の一部を織り込んだ識別子を採用している場合，利用者の属性変更に伴い識別子も変更されてしまう．識別子に属性を織り込む事には名前空間の分割管理といった利点もあるが，異動などによる属性の変更が多い組織では管理コストの増大を招くため注意する必要がある．

筆者らが所属する九州大学では学生の転学部により学籍番号の変更がある．また，複数部局に所属する教員もいる．その為，識別子には柔軟性が必要になる．

6. まとめ

本論文ではさまざまな識別子について考察した．方法として，(1)通し番号（シリアルナンバー），(2)ランダムな文字列，(3)利用者が希望する任意の文字列，(4)行政など別の枠組で設定された識別子の流用，(5)氏名からの自動生成，(6)部分毎に意味を持たせた文字列，の 6 つの方法を考察した．また，考察の要件として，管理性，利便性，耐性を挙げ，(1)～(6)の各方法を評価した．(1)～(6)の方法はそれぞれ一長一短があり，どれが最善であると決めることはできない．また，識別子を作成する場合に文字種と文字列長（固定長または可変長）についても考察した．

著者らが所属する九州大学では現在学内統一の識別子を作成する方向で検討している．今回の考察を基に九州大学に所属する学生，教職員，外部の研究員等にそれぞれ一意な識別子を作成していくことを予定している．例えば，初期案では，学生の場合は学籍番号，正規教職員の場合は共済番号などが候補に挙げられているが，この論文で挙げた基準を考慮して検討中である．非常勤教職員，留学生などの識別子の作成方法についても，現在検討中である．

文　献

[1] UPKI イニシアティブホームページ（https://upki-portal.nii.ac.jp/）

[2] Simson Garfinkel, Gene Spafford, 山口英（監訳），谷口功（訳），“UNIX&インターネットセキュリティ第 2 版”，オライリージャパン，東京，p. 245, 1999.

[3] 住民基本台帳法　第四章の二，本人確認情報の処理及び利用等　第四節本人確認の保護　第三十条の四十三．

[4] Microsoft TechNet，“Windows XP Professional の多言語機能”．（http://www.microsoft.com/japan/technet/prodtechnol/winxppro/plan/multilingual.mspx）

[5] 芝野耕司著，“漢字・日本語処理技術の発展：漢字コードの標準化”，情報処理，43 巻 12 号，pp. 1362-1367，2002.（http://www.ipsj.or.jp/katsudou/museum/paper/magazine/IPSJ-MGN431217.pdf）

[6] 総務省東北総合通信局，“大阪・兵庫０６地域の市内局番４桁化及び携帯・自動車電話，ＰＨＳの番号１１桁化”，1998.（http://www.soumu.go.jp/joho_tsusin/pressrelease/japanese/denki/981201j601.html）