

大学における生涯 ID のための名寄せ手法

太田 芳博^{†1} 梶田 将司^{†2} 田島 嘉則^{†1}
田島 尚徳^{†1} 平野 靖^{†3}
内藤 久資^{†4} 間瀬 健二^{†2}

本論文では、大学における生涯 ID のための名寄せ手法について述べる。現在、高等教育機関である大学は、これまでの教育研究活動を中核としつつ、入学前や卒業・退職後も含めた関係を構成員との間で保つ必要が生じている。このような生涯化を情報システムにより支援するためには、ユーザ ID の生涯化が必要であるが、その発行・運用のためには、名寄せ問題への対応が必須である。名古屋大学において約 34,500 人を対象に検討を行った結果、名寄せの際、生年月日・氏名（漢字表記）だけでなく、カナ・英字による氏名（別表記）を組み合わせるとともに、外国人を対象とした誤り検出処理を併用することにより、99.3%以上の名寄せ識別率が得られることが分かった。

Name Identification Method for Lifelong ID in Higher Educational Institutions

YOSHIHIRO OHTA,^{†1} SHOJI KAJITA,^{†2}
YOSHINORI TAJIMA,^{†1} HISANORI TAJIMA,^{†1}
YASUSHI HIRANO,^{†3} HISASHI NAITO^{†4} and KENJI MASE^{†2}

This paper describes a name identification method for higher educational institutions. Recently, higher educational institutions have been seeking the way to make relationships among not only constituencies like students, faculty and staff but also alumni and other supporters to realize the concept of 'Cradle-to-endowment' relationship. Although the use of lifelong User ID is required to support such activities using information technologies, name identification problem must be addressed. As the results based on about 34,500 users, it is clarified that the proposed method attains more than 99.3% of name identification rate by using the information about birthday and fullname not only in Kanji but also other expressions like Kana and English, together with an error detection processing for foreign person.

1. はじめに

大学における情報基盤は、ギガビットネットワークによるキャンパスネットワークの高速化や無線 LAN によるユビキタス化をはじめ、学生用教育端末の大規模化・分散化など、そのときどきの最新テクノロジーを組み込みながら発展している。そして、これらを活用したアプリケーションとして、大学・学部・学科・講座の各レベルごとの Web ページや、履修登録・シラバスなどの教務システム、講義などの教育学習活動を支援するためのコース管理システム、出張や物品購入などの決済システムなど、様々な情報サービスが提供されている。しかしながら、これらのシステムは、サービスを提供する学内の組織がそれぞれの方針の下で構築していることが多く、全体として統一感がなく、最悪の場合、それぞれのシステムごとにユーザ ID が異なり、ユーザはシステムごとにユーザ ID を使い分けなければならないという状況さえ生まれ始めている。これは、それぞれの情報システムが、サービス提供者の視点から構築されているためであり、サービス提供者の権益を守りつつ、ユーザの視点に立ったシステムの統合が必要になってきている。

このような背景の下、各大学では、ここ数年、ユーザ ID の統合や、システム間でのシングルサインオンなど、全学的に統一されたユーザ認証基盤の構築がさかんになっている¹⁾。ユーザ ID の発行に際しては、教職員については人事系組織が発行する職員番号を、学生については教務系組織が発行する学生番号を用いて付番することが多く、このため、ユーザ ID の発行・抹消も職員番号や学生番号に連動していることが多い。

しかしながら、学生が、ティーチングアシスタントやリサーチアシスタントとして雇用され職員番号が発行されたり、職員が、専門性を高めるために自学の大学院に入学し学生番号が発行されるなど、1人の構成員が複数の職員番号や学生番号を有している場合も増えてきており、対応するユーザ ID も複数発行されてしまう。一方で、「ゆりかごから寄付まで」²⁾

†1 名古屋大学全学技術センター

Technical Center, Nagoya University

†2 名古屋大学情報連携統括本部情報戦略室

Information and Communications Planning Office, Nagoya University

†3 名古屋大学情報連携基盤センター

Information Technology Center, Nagoya University

†4 名古屋大学多元数理科学研究科

Graduate School of Mathematics, Nagoya University

という言葉に例えられるように、優秀な学生との入学前からの関係構築や、同窓会や職業人大学院への再入学、記念事業や基金構築のための寄付など、入学前や卒業・退職後における関係を保つ方向で各大学の様々な取り組みが始まっている。このため、在籍する学生・教職員だけでなく、入学対象者や卒業生、退職教職員などの非構成員に対するユーザ ID の発行も必要になってきている。このように、生涯にわたって関係が続く多様な構成員に対して、様々なサービスを提供する必要が出てきており、「ユーザ ID の生涯化」が求められている。

そこで、本論文では、名古屋大学でのユーザ ID の生涯化を例に、大学における生涯 ID のための名寄せ手法を提案する³⁾⁻⁵⁾。

本論文の構成は次のとおりである。まず、2 章において、生涯 ID である名古屋大学 ID の導入の背景と付随する名寄せ問題について述べる。そして、3 章・4 章において、名古屋大学での事例をふまえながら名寄せ手法を提案・評価するとともに、5 章において名寄せ処理効率化に向けた考察を行う。最後に、6 章において本論文をまとめる。

2. 名古屋大学 ID の導入と名寄せ問題

2.1 名古屋大学 ID 導入の背景

名古屋大学では、ユーザ認証基盤の統合化の第一歩として、2003 年 2 月に情報連携基盤センター（当時）¹⁾が「全学 ID」の運用を開始した。全学 ID は、情報メディア教育センターが発行していた学生用ユーザ ID をベースとし、新たに教職員を対象に加えることで、名古屋大学の全構成員約 21,000 人に発行された。

しかし、全学 ID の利用が広がるにつれて、多くの問題点が指摘されるようになった⁶⁾⁻⁸⁾：

- (1) 単純に学生番号や職員番号に基づいて全学 ID は付番されていたため、電子メールアドレスや画面表示などで利用された場合、学生番号や職員番号を誰でも容易に類推できてしまう。
- (2) 非常勤雇用や社会人入学などにより、学生番号や職員番号が新たに割り当てられるたびに全学 ID が発行されるため、複数の身分を持つ利用者は複数の全学 ID を使い分ける必要がある。
- (3) 学部を卒業して大学院に進学するなど、身分が変更になる場合は全学 ID も変更になるため、継続的な利用が可能な情報サービスでも、継続利用ができない。
- (4) 学生番号、職員番号を持たない身分の構成員にも学内情報サービスを提供する必要が

ある。

- (5) 構成員の離籍時には全学 ID は無効となるため、卒業生や退職者向けの情報サービスが提供できない。

これらの問題・課題に対応するために、2004 年 12 月から全学情報小委員会の下で抜本的な見直しの検討が行われ、2005 年 9 月には役員会で新しい ID 体系である「名古屋大学 ID」の方針が次のように決定された⁶⁾⁻⁸⁾：

- 名古屋大学に在籍するすべての構成員（正規学生、研究生、聴講生、常勤職員、非常勤職員）、および卒業生・修了生に対して発行する。
- 在籍中に複数の身分を持つ場合でも 1 人 1 ID とし、さらに離籍後も同じ ID が利用できるようにする（「生涯 ID」としての導入）。
- 全学同窓会との協力体制の下、卒業生・修了生に対しても名古屋大学 ID を発行し、卒業生・修了生に対するサービスの向上に対応する。
- 長期（たとえば 100 年間）にわたって利用できるような ID 体系とする。
- 名古屋大学内のすべての情報サービスで用いることができるユーザ認証システムを提供する。
- 学内情報サービス提供者が必要とする、氏名、所属部局など ID に付随するセキュリティ的に問題とならないような基本情報は提供できるようにする。

この方針に基づいて、「名古屋大学 ID」の実装と新しい認証基盤サービスの運用準備が開始された。

2.2 名寄せ問題

名古屋大学 ID には、その設計方針の中で「生涯 ID」という新しい概念が導入された。

しかし、名古屋大学 ID の導入準備が本格的になるにつれて、予想以上に深刻であった問題が「名寄せ」であった。

名寄せとは、同一人物に 1 つだけの生涯 ID を発行するための事前処理である。たとえば、名古屋大学においては「大学院博士課程後期課程に在学中、ティーチングアシスタントやリサーチアシスタントとして雇用された」という場合、全学 ID は合計 3 つ発行されている。このように、同 1 人物に割り当てられた複数 ID を束ねる問題を「タイプ I 名寄せ問題」と呼ぶこととする（図 1 参照）。

一方、名古屋大学では毎年 4 月には約 5,000 人に及び入学者・進学者・新規採用者が発生する。「生涯 1 つの ID」を標榜する名古屋大学 ID の発行過程においては、新しい構成員に対し、名古屋大学 ID をすでに発行しているかどうかを確認する必要がある。たとえば、

*1 2009 年 4 月に情報メディア教育センターと統合され、情報基盤センターに改組された。

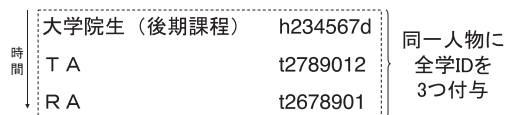


図 1 同一人物に対して複数発行された全学 ID を束ねる問題 (「タイプ I 名寄せ問題」)
Fig.1 Problem of name identification (Type I).

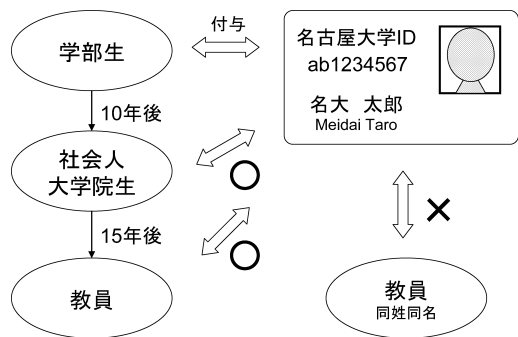


図 2 すでに名古屋大学 ID を発行済みの人物に対して重複して発行しないようにする問題 (「タイプ II 名寄せ問題」)
Fig.2 Problem of name identification (Type II).

名古屋大学を卒業後、民間企業に就職したものの、10 年後に弁護士資格を取得するために法科大学院に入学した場合、仮に本人がすでに発行済みの名古屋大学 ID を忘れていたとしても適切に名寄せを行い、重複発行しないようにする必要がある。このように、すでに名古屋大学 ID を発行済みの人物に対して重複発行しないようにする問題を「タイプ II 名寄せ問題」と呼ぶこととする (図 2 参照)。

2.3 名寄せ問題へのアプローチ

このような名寄せ問題は、企業における顧客情報管理や、地方公共団体における戸籍・住民基本台帳連携、社会保険庁における年金記録管理など、医療・金融・保険業界では古典的な問題である。たとえば、保険契約の顧客情報管理では、漢字氏名と生年月日を利用した漢字名寄せ索引システム⁹⁾が、また、戸籍・住民基本台帳連携では、漢字氏名 (筆頭者および記載されている個人)、生年月日、本籍を利用したマッチング処理¹⁰⁾が提案されているが、大学業界では具体的な研究事例はない。

大学構成員の個人属性情報は、教職員については人事系組織が、学生については教務系組

織が管理している場合が多く、管理されている属性情報項目も統一されていないが、漢字氏名・生年月日はいずれの場合でも管理対象になっているので、大学においても漢字氏名・生年月日は名寄せ処理に利用可能である。しかしながら、これらの属性情報の大学における名寄せ性能は明確ではなく、また、教職員・学生・研究員など多種多様で入れ替わりが激しい多数の構成員が対象となることから、大学に特化したより良い名寄せ手法を開発する必要がある。

そこで、本研究では、タイプ I 名寄せ問題を対象に、大学における名寄せ手法の検討を行うとともに、今後、入学者・進学者・新規採用者に対して継続的に発生するタイプ II 名寄せ問題を対象に評価を行うこととした。

なお、商用の名寄せソフトウェアサービス^{11),12)}を利用することも考えられるが、1) 長期にわたる金銭的なコスト負担、2) 構成員の個人情報を学外へ提供することによる情報漏洩の懸念、から独自対応することとした。これにより、大学に特化した名寄せ手法のあり方を明確にするとともに、他大学との成果共有も行うことができる。

3. 大学における名寄せ手法の開発

ここでは、タイプ I 名寄せ問題を対象に大学における名寄せ手法を検討する。まず、手動により正しい名寄せ結果 (以降、「名寄せグランドトゥルース (Grand Truth)」と呼ぶ) を確定し、そのうえで、改めて名寄せ処理手順を検討する。

3.1 名寄せグランドトゥルースの確定

3.1.1 実験条件

総務部人事労務課が管理する「人事マスタ」および、学務部学務企画課が管理する「学務マスタ」の提供を受けて、氏名、職員番号、学生番号、生年月日などの個人属性情報を抽出した。また、一部の研究生・研究員など、総務部・学務部では在籍状況を把握していない者については、全学 ID の「個別申請マスタ」から同様に必要な個人属性情報を抽出した。ただし、個別申請マスタには生年月日がなかったため、別途、調査を行った。

3.1.2 実験方法

2006 年 11 月の段階で LDAP サーバに登録されていた 44,324 の全学 ID すべてを「有効な全学 ID」と見なし、上記マスタ群から抽出した学生番号・職員番号をもとにデータ結合を行った。そして、結合データに対して担当者 2 人が氏名と生年月日を利用して次のように名寄せを行い、それぞれの結果を比較・修正したうえで最終的な結果を得た：

- 1) 担当者 A が、氏名 (漢字表記) と生年月日を主キーとしてデータ比較を行い、簡易的

表 1 名寄せ結果
Table 1 The results of name identification.

名寄せ状況		ユーザ数	割合
名寄せが不要なユーザ		30,771 人	83.0%
名寄せが必要なユーザ	全学 ID を 2 つ付与	5,496 人	14.8%
	全学 ID を 3 つ付与	745 人	2.0%
	全学 ID を 4 つ付与	75 人	0.2%
	全学 ID を 5 つ付与	4 人	0.01%
	全学 ID を 6 つ付与	1 人	0.003%
小計		6,321 人	17.0%

に名寄せ結果を得た後、生年月日を手がかりに名寄せ誤りを訂正し、名寄せ結果 A を得る。

2) 担当者 B が、氏名(カナ表記)と生年月日を主キーとし、1)と同じ手順により名寄せ結果 B を得る。

3) 名寄せ結果 A・B 間の差分を抽出し、他の担当者 2 人を加えて目視確認^{*1}により修正し、最終的な名寄せ結果を確定する。

なお、各作業の途中で気づいたマスターデータに関する問題点(生年月日の誤りが疑われるもの、全角・半角空白文字の違い、カタカナ・ひらがなの混在の修正、アルファベット表記の全角・半角の違いなど)については適宜、担当者が確認、修正を加えながら名寄せを行った。

3.1.3 実験結果

上記の手順により得た名寄せ結果を表 1 に示す。

表から分かるように、全体の 83.0%のユーザは全学 ID が重複して発行されておらず、名寄せの必要はなかった。一方、全学 ID が複数発行されており、名寄せの必要があるユーザは 17.0%であった。名寄せが必要なユーザにおいては、全学 ID を 2 つ付与したケースが最も多かった。また、5 つ以上付与したケースはきわめて稀であることも分かった。

3.2 名寄せ処理手順の検討

名寄せグランドツールの作成過程を通じて、名寄せを困難にする特徴的な問題点として次の 3 点が分かった：

*1 ここでの目視確認とは、差分を生年月日順にソートし、誤りを発見しやすくしたうえで、名寄せ結果 A・B どちらが正しいかを判定した。

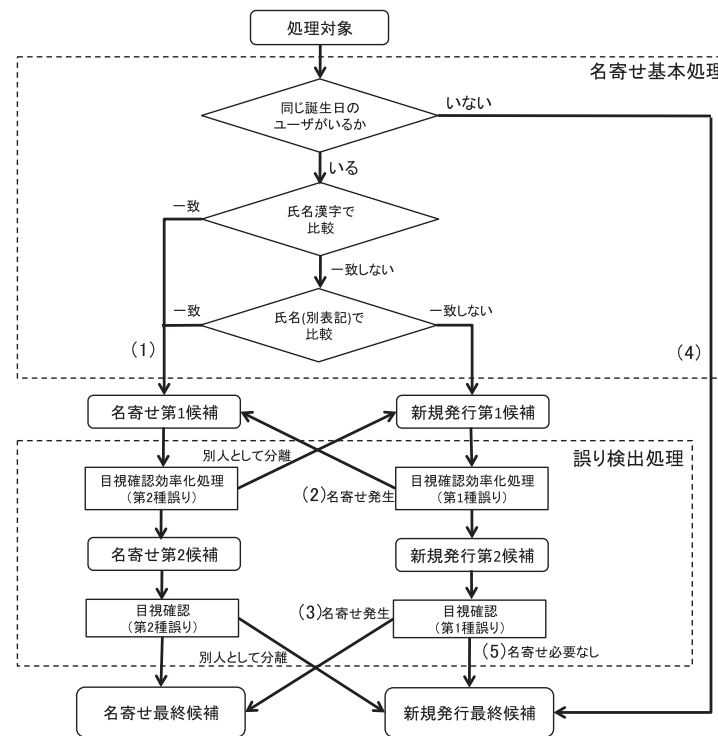


図 3 大学における名寄せ手法
Fig. 3 Name identification method for higher educational institutions.

- (1) 日本人の氏名(漢字表記)の揺れの問題。
たとえば「よみ」が「わたなべ」である場合、「渡辺」「渡邊」「渡邊」「渡部」のように氏名の漢字表記には複数の候補があるため、入力時のカナ漢字変換ミスや漢字表記の簡略化が行われると、氏名の漢字表記を比較するだけでは同一人物とは判定できない。
- (2) 欧米外国人の氏名の揺れの問題。
外国人に関しての氏名登録方法が統一されていないことが主な原因である。具体的には、氏名の漢字表記ができない場合、入力担当者が独自の判断に基づきアルファベットやカタカナで登録していた。さらに、氏名からカタカナの変換方法にも違いが

生じていた。

(3) 改姓・改名・通称使用時の問題。

改姓・改名・通称使用希望の反映により、同一人物であっても氏名が異なる場合があった。

以上のことから、表記に揺らぎの多い氏名（漢字表記）だけを比較するよりも、氏名（カナ表記）や氏名（英字表記）など、氏名に関する利用可能な別表記を併用^{*1}する「名寄せ基本処理」を行うとともに、目視確認対象に含まれる欧米外国人の氏名の揺れや改姓・改名・通称の問題などに対応するための「誤り検出処理」を行う名寄せ手法を提案する（図3参照）。

本手法は、新規発行第1候補・第2候補に分類されたものの中に「誤って他人と判定してしまったもの」が含まれる可能性がある。また、名寄せ第1候補・第2候補には「誤って同一人物と判定してしまったもの」が含まれる可能性がある。これらの誤りをそれぞれ「第1種誤り」「第2種誤り」として以降では扱う。

4. 評価実験

本章では、タイプII名寄せ問題の対象者となる、入学・進学・新規採用時に発生する新規構成員およびすでに卒業・修了・退職した離籍者について、前章で提案した大学における名寄せ手法の評価・検討を行う。

4.1 実験条件

新規構成員としては、2008年4月期に入学・採用された5,507人を、離籍者としては、1989年9月時点で在籍していた教職員などを含む人事マスタ、および1993年4月以降に入学した学生を含む学務マスタから全学ID発行以前の離籍者28,995人を対象とした。また、氏名（別表記）としては、氏名（カナ表記）および氏名（英字表記）（利用可能な場合^{*2}）を用いた。

また、誤り検出処理については、次のように実装した：

目視確認効率化処理（第1種誤り） 外国人の氏名の揺れの問題に対応するため、比較対象ユーザの氏名（漢字表記）にアルファベットが現れた場合は外国人と見なし、そのアルファベット表記の氏名が部分一致する場合は「外国人特例」として名寄せ発生と判定す

*1 たとえば、氏名（漢字表記）に異体字が使用されていても、氏名（カナ表記）が正しく登録されていれば名寄せ検出が可能となる。

*2 2008年1月からICカード化された新学生証・職員証の導入にともない、英字表記が利用できるようになったが、教職員については非常勤職員に対する新職員証導入時期がずれたため、氏名（英字表記）は利用しなかった。

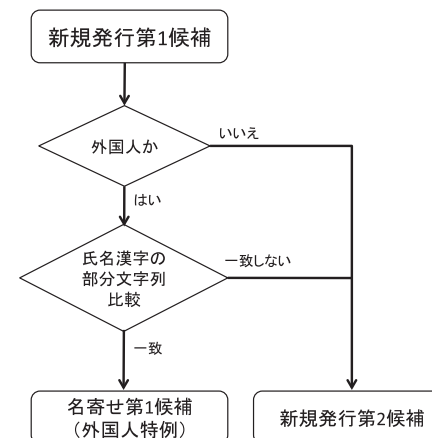


図4 外国人用目視確認効率化処理（第1種誤り）

Fig. 4 Processing for improving visual inspection for foreign person.

ることとした（図4参照）。実験では部分文字列の比較として、予備実験において十分な識別結果が得られることが分かった、先頭のアルファベット5文字を用いた。

目視確認（第1種誤り） 比較対象ユーザと同じ生年月日のユーザの氏名（漢字表記）をすべて抽出し、目視確認を行う。不明な場合は関係部署に調査を依頼した。

目視確認効率化処理（第2種誤り） 目視確認（第2種誤り） 過去数年分の電子化されたマスタデータにおいて同姓同名で同じ生年月日の人はいないことが分かっていたため、特に実装は行わなかった。

4.2 実験結果

4.2.1 新規構成員の場合

名寄せ結果を表2に、第1種誤り数を表3に示す。表2中の出力項目中の番号は図3における番号に対応しており、図3中の「名寄せ最終候補」の総数は表2の(1),(2),(3)の合計であり、「新規発行最終候補」の総数は(4),(5)の合計である。

4.2.1.1 全体性能について

新規発行第2候補の目視確認（第1種誤り）で検出された第1種誤り率は全体の0.7%であり、この目視確認を行わなかったとしても名寄せ処理全体では非常に高い識別率（99.3%）を得ることができた。

表 2 新規構成員に対する名寄せ結果 (単位: 人).

Table 2 The results of name identification (unit: person).

処理内容	出力項目	学部学生		大学院生		教職員		計	
		人	%	人	%	人	%	人	%
基本処理	名寄せ第 1 候補 (1)	7	0.3	1,327	58.1	456	51.3	1,790	32.5
	新規発行第 1 候補	552	23.6	903	39.6	381	42.9	1,836	33.3
	新規発行最終候補 (4)	1,776	76.1	53	2.3	52	5.9	1,881	34.2
目視確認 効率化処理 (第 1 種誤り)	名寄せ第 1 候補 (2)	0	0.0	30	1.3	1	0.1	31	0.6
	新規発行第 2 候補	552	23.6	873	38.2	380	42.7	1,805	32.8
目視確認 (第 1 種誤り)	名寄せ最終候補 (3)	4	0.2	28	1.2	8	0.9	40	0.7
	新規発行最終候補 (5)	548	23.5	845	37.0	372	41.8	1,765	32.1
計	(1)+(2)+(3)+(4)+(5)	2,335	100.0	2,283	100.0	889	100.0	5,507	100.0

表 3 新規構成員に対する名寄せにおける第 1 種誤り数 (単位: 人)

Table 3 The number of type I error for name identification processing (unit: person).

	学部生	大学院生	教職員	計
生年月日および氏名 (漢字表記) を使用 (従来法)	5	94	25	124
生年月日および氏名 (漢字表記・カナ表記) を使用 (提案法)	4	63	9	76
生年月日および氏名 (漢字表記・カナ表記・英字表記) を使用 (提案法)	4	58	-	62

4.2.1.2 個別性能について

学部学生はほとんど名寄せは発生せず (11 人, 処理対象者の 0.47%), 逆に 99.5% が新規発行となることから分かる。これは現役入学者 (生年月日が 1989 年 4 月 2 日以降) は, 過去に在籍したことがないためである。

大学院生については, 名寄せ最終候補が 1,385 人 (処理対象者の 60.1%) と多い。これらは, 学部を卒業して大学院に進学している場合である。教職員については, 名寄せ最終候補が 465 人 (処理対象者の 52.3%) と多いが, 学生のときのようにはっきりとした特徴は読み取れなかった。

表 3 より, 氏名 (漢字表記) のみを用いて名寄せを行う従来法と比べて, 氏名 (カナ表記) を併用することで, 第 1 種誤りを 38.7% 削減 (124 件から 76 件) でき, さらに氏名 (カナ表記), 氏名 (英字表記) を併用することで 50.0% 削減 (124 件から 62 件) することができた。この結果から, 従来方法⁹⁾ と比較して, 生年月日・氏名 (漢字表記)・氏名 (別表記) を併用する名寄せ基本処理は, 有効であることが分かった。

次に, 新規発行第 2 候補に対し目視確認を行い, 第 1 種誤りとされた数とその誤り率を表 4 に示す。全体では, 新規発行第 2 候補の中から, 2.2% の第 1 種誤りが発見されている。

表 4 新規発行第 2 候補の目視確認により発見された第 1 種誤り数 (単位: 人)

Table 4 The results of name identification type I error (unit: person).

	新規発行第 2 候補数	誤り数	誤り率 (%)
学部学生	552	4	0.7
大学院生	873	28	3.2
教職員	380	8	2.1
合計	1,805	40	2.2

4.2.2 離籍者の場合

名寄せ結果を表 5 に, 第 1 種誤り数を表 6 に示す。新規構成員の場合と同じように, 第 1 種誤り率は全体の 0.6% であり, 目視確認を行わなかったとしても名寄せ処理全体では非常に高い識別率 (99.4%) を得ることができた。新規発行第 2 候補からは 0.9% の第 1 種誤り (20,230 人中 174 人) が目視確認で検出されている。

4.3 さらに目視確認効率化に向けた調査

新規構成員の結果において, 目視確認で発見された第 1 種誤りについて日本人と外国人に分類し, その原因を調べた。「日本人」「中国・韓国などの漢字文化圏の外国人」「その他

表 5 名寄せ結果 (単位: 人).

Table 5 The results of name identification (unit: person).

処理内容	出力項目	計	
		人	%
基本処理	名寄せ第 1 候補 (1)	4,645	16.0
	新規発行第 1 候補	20,239	69.8
	新規発行最終候補 (4)	4,111	14.2
目視確認	名寄せ第 1 候補 (2)	9	0.03
効率化処理	新規発行第 2 候補	20,230	69.8
目視確認	名寄せ最終候補 (3)	174	0.6
	新規発行最終候補 (5)	20,056	69.2
計	(1)+(2)+(3)+(4)+(5)	28,995	100.0

表 6 新規発行第 2 候補の目視確認により発見された第 1 種誤り数 (単位: 人)

Table 6 The results of name identification type I error (unit: person).

	新規発行第 2 候補	誤り数	誤り率 (%)
計	20,230	174	0.9

表 7 目視確認で発見された第 1 種誤り数

Table 7 The number of categorized name identification type I error.

	学部学生	大学院生	教職員
	発生数	発生数	発生数
日本人	0	3	2
漢字圏の外国人	4	10	2
その他の外国人	0	15	4
計	4	28	8

の外国人」ごとに分類したものを表 7 に示す。

日本人に関しては、誤りの原因としては以下のようなものであった：

- (1) 改姓, 改名が行われていた。
 - (2) 氏名データの不備 (氏名 (カナ表記), 氏名 (英字表記) の未登録) によるもの。
 - (3) 氏名 (カナ表記) において、促音部が大文字小文字で異なっていたこと (「ヨ」と「ョ」など)。
 - (4) 氏名 (英字表記) のアルファベット表記が異なっていた (「Shou」と「Sho」など)。
- 漢字文化圏の外国人の誤りの原因としては以下のようなものであった：
- (1) 氏名 (漢字表記) に異体字が使用されていた。
 - (2) 氏名 (漢字表記) に英名が使用されていた。

(3) 氏名 (漢字表記) に一部だけカナが使用されていた^{*1}。

漢字文化圏以外の外国人の誤り原因は、以下のようなものであった：

- (1) 通常、日本人では氏名 (漢字表記) が格納されるフィールドに、英名やカタカナで氏名が登録されていた。
- (2) “ ” や “ ” , “ ” など氏名には不適切と思われる文字が入っていた。
- (3) 姓と名が逆順序で登録されていた。
- (4) 姓しか登録されていなかった。
- (5) 名がイニシャルだけしか登録されていなかった。

5. 名寄せ誤りの改善に向けた考察

前章で示したように、我々の提案した名寄せ手法は第 1 種誤り率を全体の 0.6% 以下にすることができた。しかし、誤りをなくすことはできないため、最悪の場合は目視確認を行う必要がある。これを極力回避するためには、次のような対策が考えられる。

第 1 種誤りの改善案：

- 1) 名寄せ基本処理において、氏名によく現れる漢字の異体字変換テーブルを作成しておく、テーブルを参照して漢字氏名の比較処理を行う。
- 2) 改姓・改名についての情報を人事マスタ・学務マスタ側で管理し、目視確認効率化処理において利用する。
- 3) 人事マスタ・学務マスタ側において氏名登録時のルールを策定し、氏名表記方法を統一する。

第 2 種誤りの改善案：

- 1) 在籍情報 (過去に在籍したか、どこに在籍していたかなど) を、目視確認効率化処理において利用する。
- 2) 同姓・同名・通名使用に関する情報を人事マスタ・学務マスタ側で管理し、目視確認効率化処理において利用する。

6. ま と め

本論文では、名古屋大学でのユーザ ID の生涯化を例に、大学における生涯 ID のための名寄せ手法について述べた。その結果、名寄せの際、生年月日・氏名 (漢字表記) だけでな

*1 データ入力時に該当する漢字が見つけれなかったためと思われる。

く、カナ・英字による氏名（別表記）を組み合わせるとともに、外国人を対象とした誤り検出処理を併用することにより、99.3%以上の名寄せ識別率が得られることが分かった。また、さらなる性能向上のための対策についても述べた。

今後の課題としては、さらなる性能向上を旨として、単なるユーザ ID の管理業務にとどまらず、人に関する情報の管理業務を全学的に適切に行うための組織およびルールを整備する必要がある。

謝辞 本研究（の一部）は国立情報学研究所の最先端学術情報基盤（CSI）事業の一環として行われたものである。

参 考 文 献

- 1) 内藤久資, 梶田将司, 小尻智子, 平野 靖, 間瀬健二: 大学における統一認証基盤としての CAS とその拡張, 情報処理学会論文誌, Vol.47, No.4, pp.1127-1135 (2006).
- 2) Katz, R.N. and Associates: *Web Portals & Higher Education*, Jossey-Bass (2002).
- 3) 梶田将司, 太田芳博, 田島嘉則, 田島尚徳, 平野 靖, 内藤久資, 間瀬健二: 生涯利用可能な名古屋大学 ID の導入に伴う名寄せ問題とその解決方法, 情報処理学会研究報告 (分散システム/インターネット運用技術), Vol.2008, No.23, pp.73-78 (2008).
- 4) 太田芳博, 梶田将司, 田島嘉則, 田島尚徳, 平野 靖, 内藤久資, 間瀬健二: 生涯利用可能な名古屋大学 ID の新規発行における名寄せ方法に関する検討, 情報処理学会研究報告 (インターネットと運用技術), Vol.2008, No.37, pp.109-114 (2008).
- 5) 内藤久資: 名古屋大学におけるアイデンティティ管理—生涯 ID としての名古屋大学 ID, 第 23 回インターネット技術第 163 委員会研究会 (オンライン). 入手先 <http://www.itrc.net/report/meet23/data/1p3/naitou.pdf> (参照 2007-06-02)
- 6) 間瀬健二, 平野 靖, 梶田将司: 名古屋大学 ID の導入について—(I) 概要, 名古屋大学情報連携基盤センターニュース, Vol.5, No.4, pp.316-320 (2006).
- 7) 平野 靖, 梶田将司, 間瀬健二: 名古屋大学 ID の導入について—(II) 全学 ID からの移行, 名古屋大学情報連携基盤センターニュース, Vol.6, No.2, pp.140-145 (2007).
- 8) 梶田将司, 平野 靖, 間瀬健二: 名古屋大学 ID の導入について—(III) 将来構想, 名古屋大学情報連携基盤センターニュース, Vol.7, No.1, pp.11-17 (2008).
- 9) 小川孝一郎: 漢字名寄せ索引システム, オペレーションズ・リサーチ: 経営の科学, Vol.23, No.6, pp.366-370 (1978).
- 10) 竹中 稔: 戸籍の電子化にあたっての名前表記問題, 第 23 回インターネット技術第 163 委員会研究会 (オンライン). 入手先 <http://www.itrc.net/report/meet23/data/1p3/takenaka.pdf> (2008.5)
- 11) NTT データ株式会社: 名寄せソリューション OUSIA .<http://bs.nttdata.co.jp/ousia/>
- 12) i's Factory: 名寄せサービス .
<http://www.isfactory.co.jp/solution/identification.html>

(平成 21 年 6 月 15 日受付)

(平成 21 年 12 月 17 日採録)



太田 芳博

昭和 45 年生。平成 4 年福井大学工学部情報工学科卒業。平成 4 年名古屋大学環境医学研究所宇宙医学実験センター技術官。現在、名古屋大学全学技術センター共通基盤技術支援室情報通信技術系技術課技師。



梶田 将司 (正会員)

平成 2 年名古屋大学工学部情報工学科卒業。平成 7 年同大学院工学研究科情報工学専攻博士課程満了, 博士 (工学)。平成 14 年名古屋大学情報連携基盤センター准教授, 平成 21 年同大学情報連携統括本部情報戦略室准教授, 現在に至る。情報技術の戦略的利活用を通じた高等教育機関における教育・研究の質的変革に関する研究・実践に従事。平成 10 年日本音響学会第 15 回粟屋潔学術奨励賞, 平成 13 年電子情報通信学会第 57 回論文賞, 平成 21 年 IBM Shared University Research Award 受賞。電子情報通信学会, 日本音響学会, 日本教育工学会, 教育システム情報学会, IEEE, ACM 各会員。



田島 嘉則

昭和 38 年生。昭和 61 年中部大学工学部電子工学科卒業。平成 2 年学術情報センター事業部システム管理課技術官。平成 7 年名古屋大学医学部附属病院事務官。平成 9 年名古屋大学大型計算機センター技術官。平成 14 年名古屋大学情報連携基盤センター技術官。現在、名古屋大学全学技術センター共通基盤技術支援室情報通信技術系技術課技師。



田島 尚徳

昭和 49 年生。平成 9 年中部大学工学部機械工学科卒業。平成 9 年鳥取大学工学部庶務掛機械実習工場技術官。平成 10 年鳥取大学工学部技術部技術官。平成 14 年名古屋大学情報連携基盤センター技術官。現在、名古屋大学全学技術センター共通基盤技術支援室情報通信技術系技術課副技師。



平野 靖

平成 7 年名古屋大学工学部電子情報工学科卒業。平成 9 年同大学大学院博士課程前期課程（電子情報学専攻）修了。平成 11 年同大学院博士課程後期課程（情報工学専攻）修了。平成 12 年 4 月同大学院工学研究科助手。平成 14 年 4 月同大学情報連携基盤センター助手。平成 16 年 4 月同大学情報連携基盤センター助教授。平成 19 年 4 月同大学情報連携基盤センター准教授。博士（工学）。平成 10 年 4 月より平成 11 年 11 月まで日本学術振興会特別研究員（DC2），平成 11 年 12 月より平成 12 年 3 月まで日本学術振興会特別研究員（PD）。3 次元画像処理とその肺腫瘍の良悪性鑑別への応用に関する研究，および大学内・大学間ユーザ認証基盤に関する研究に従事。電子情報通信学会，日本生体医工学会，日本医用画像工学会，および IEEE 各会員。



内藤 久資

昭和 36 年生。大阪大学理学部数学科卒業。名古屋大学大学院理学研究科前期課程数学専攻修了。理学博士。名古屋大学大学院多元数理科学研究科准教授。主たる研究テーマは微分幾何学と情報ネットワーク。特に，多様体上の変分問題と関連する非線型偏微分方程式，幾何学的オブジェクトの視覚化，幾何学と材料科学との融合，情報ネットワークの認証システム等。日本数学会会員。



間瀬 健二（正会員）

昭和 54 年名古屋大学工学部卒業。昭和 56 年同大学大学院工学系研究科情報工学専攻（前期課程）修了。同年日本電信電話公社（現 NTT）入社。昭和 63 年～平成元年米 MIT メディア研究所客員研究員。平成 7～14 年 ATR 研究室長。平成 14 年より名古屋大学教授。現在，同大学大学院情報科学研究科所属。人工知能学会平成 11 年度論文賞ほか。電子情報通信学会フェロー，IEEE，ACM，電子情報通信学会，人工知能学会ほか各会員。博士（工学）。