

漢字姓名による検索

田 中 康 仁 (日本ユニバック㈱)

1. はじめに

漢字情報処理システムのサブシステムとして漢字ディスプレイのハードウェアとこれを動かすベシック・ソフトウェアの開発が各社で行なわれている。しかし、この漢字ディスプレイをいかに使うか、漢字ディスプレイを使用することによりどのようなことが解決し、どのような問題点を研究し解決しなければならぬか明確でない。

ここでは、どのような問題点があるか述べ、その解決の一端を示したい。その前提としては銀行、証券、生保、損保業界で考えられている、また一部試行されている漢字の名寄せシステムを例として考える。データ量は数百万件で、人手による検索は不可能な量と仮定する。

2. 問題はどこで発生するか

漢字によるマスター作成は終わり、姓と名による名寄せ作業は終了しているとする。しかも銀行の支店または生命保険の窓口には漢字ディスプレイが設置されていると仮定する。そこへ1人の客が現われ、銀行の窓口で或る事務処理を行なおうとしたとする。ところが、客は顧客名のメモを無くしてしまっており、漢字での索引を行なわなければならなくなった。窓口係が漢字ディスプレイで顧客の姓、名を入れたが、そのような顧客との取引はないと返事が返ってきた。客はしだいに銀行員の態度にいらだたしくなり、銀行に文句を言い、二度とこの銀行とは取引するものかと心に言いきかせて出て行ってしまった。銀行は1人の顧客を失ってしまったのである。

これは表面に出た場合であるが、クレジット・カードや割賦販売のブラック・リストの検索では犯罪を防止できず、見逃すことになり問題である。

実際の検索には、姓名と住所コード、姓名と生年月日などの別の附属情報とともに検索を行なうが、ここでは附属情報についてはあまり重点をおかず、漢字の姓、名について問題を分析する。

いままでの漢字で管理されている顧客ファイルは、厚紙でできたカードとか台帳とかに記入されており、これを人の目に頼って、または簡単な機械によって、検索している。それゆえ少しぐらいの相違でも別の手段で本人が確認できればその顧客名を見つけだせた。また一つの検索で失敗すれば長年の勤で数枚のカードを探し出し、その中からその顧客を確認し顧客名を見つけていたわけである。しかし機械には自分で考えつく能力がないため検索を不可能にしてしまう。さらに困ったことは機械化とかシステム化ということは、今まで行なっていた手作業により管理されていたビジブル・ファイルから解放されることに意味があるわけである。それゆえ漢字ディスプレイ以外に頼るものがなくなってしまう。このため漢字ディスプレイを使っていかに検索するかは重要な問題である。

3. 誤りの発生する事例

誤りの発生する原因を考えながら解決方法を調べてみる。これらの原因としては次の7個のものがある。

- | | |
|----------------------|--------------|
| ① 同一文字の取り扱い | ⑤ 小文字と大文字の表現 |
| ② 書き誤りやすい文字，見誤りやすい文字 | ⑥ 特殊な文字の使い方 |
| ③ くり返し記号 | ⑦ 変体仮名について |
| ④ かな使い上の問題 | |

これらについて個々に述べる。

3.1 同一文字の取り扱い

漢字の中には同一文字でありながら新字体，旧字体の文字がある。たとえば人名に使われる頻度の高い文字で「沢」という文字と「澤」，「辺」という文字と「邊」等がある。このような文字は一方的に新字体に統一しようとしても顧客の反対が強く，統一は不可能である。このような同一文字で字体の異なるものについては，それらがどのような対で発生するか調べ，プログラムで同一文字として取り扱うという処理をほどこせば問題がない。

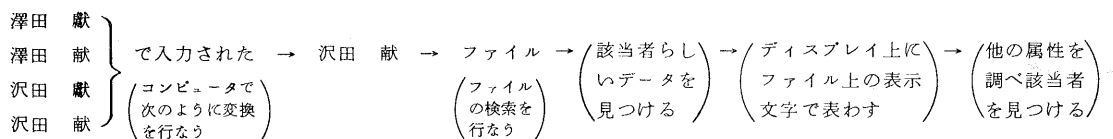
具体的に例をあげて考えてみる。たとえば「澤田 献」という人がいたと仮定し，銀行のファイルに登録されている文字は上記文字で行なわれているとする。しかしこの代理人が窓口に残われて伝票に「沢田 献」と書き，書類を作成したと仮定する。窓口係は「沢田 献」という人は見つからないため，このような人で当銀行に口座を持っている該当者はいませんと言ったとする。これでは何のための名寄せシステムか，また顧客サービスかわからなくなる。そのためにはファイルのアクセス回数が多くなっても，また似た該当者が現われる確率が多くなっても，次の4種類の索引で該当者の検索が行なわれるようにすべきである。

- | | |
|-----------|-----------|
| (1) 澤 田 献 | (3) 沢 田 献 |
| (2) 澤 田 献 | (4) 沢 田 献 |

別の解決方法としては表示する文字（領収書や案内に記入されたり，郵便物の宛名等の文字）とファイルに一定順序でデータを貯えておくための文字（検索に使用する際にコンピュータで表わす内部文字）と区別すべきである。

上記4件のケースはすべて「沢田 献」の領域にファイルする。このようにすれば同一文字による検索ミスは減らすことができる。

コンピュータで次のように検索



3.3 くり返し符号

名前の中にはくり返し符号がしばしば使われている。たとえば「佐々木」とか「すゞこ」等がある。くり返し符号は前の文字がくり返される際に同一文字を書くことを省略する意味で使われる。このくり返し符号には次のようなものがある。

- | | | | | | | | |
|---|---|---|----|---|---|---|---|
| ① | 々 | ③ | ゞ | ⑤ | く | ⑦ | 〃 |
| ② | ゝ | ④ | く' | ⑥ | ゝ | | |

このうち人名に用いられるのは ① ② ③ である。

くり返し符号が常に使われていれば処理しやすいが、使用したり使用しなかったり漢字を書いたりひらかなを書いたりする場合がある。このためファイルを正しく検索することはできなくなる。

くり返し符号の使い方についての基準として「文部省刊行物表記の基準」をあげる。

くり返し符号の使い方（昭和 25 年 9 月「文部省刊行物表記の基準」による）

現代口語文においては、くり返し符号（反復記号）は用いないのが一般的であるが、文芸作品その他では現在でも用いられているのでここに掲げた。なお、「々」が当用漢字音訓表にないため、「村むら、雄おしい」等の表記が行なわれているが、「々」は記号なのでして避ける必要はない。

くり返し符号は、「々」以外は、できるだけ使わないようにするのが望ましい。

「々」同の字点は、漢字一字のくり返しの場合に用いる。

（例）人々 国々 年々 日々

ただし次のような場合には「々」を用いない。

（例）民主主義 大学学術局 当用漢字字体表

「ゝ」一つ点（一の字点）は、かながきの一語の中で、同音をくり返すとき。

（例）あゝ たゝみ とゝのえる じゝ

ただし、次のような場合には「ゝ」を用いない。

（例）バナナ ココア かわいい くまざさ 手がかり そのち いままで
あわてて てりはいうものの …のことと …とともに

「ゞ」は、かな書きの一語の中でくり返された下の音が濁るとき。

（例）たゞし かゞみ すゞり さゞ波

ただし、次のような場合には「ゞ」を用いない。

（例）読んだだけ すべてです。

「く」の字点は、二字のかなをくり返すとき。

（例）くくく ぐぐぐ ぐぐぐ

ただし、三字以上にわたる場合、および二字以上の漢語や、横書きの場合などには用いない。

「ゝ」（二の字点）は、用いないのを原則とする。

「〃」ノの点は、表や簿記などには用いる。

このくり返し符号は戸籍上では使用は許されているか調べてみると、法務省の見解は次のようになっている。

同音の繰り返しに用いる「ㄨ」又は「ㄨ」及び同字の繰り返しに用いる「々」の符号は子の名に用いて差しつかえない（昭和26年3月30日民事甲第674号回答）

同音の繰り返しに用いる「ㄨ」ならびに「ㄨ」及び同字の繰り返しに用いる「々」の文字を子の名に用いた出生の届出は、これを受理してよろしいか。前二者は50音図の範囲内には含まれていないが「現代かなづかい」において使用を認められていることにより、後者は当用漢字中にはないが日常慣用されていることにより、いずれも可否両説に岐れ、戸籍事務取扱上疑義を生ずるに至りましたから、至急何分の御回示に接したく稟請いたします。

おって受理できないとすれば、すでに戸籍の記載を了したこれらのものについてはいかに措置するか、併せて御指示願いたく申し添えます。

回 答

出生届は、受理して差しつかえない。

このくり返し符号についても何か対策が必要である。

変体仮名について調べてみると，日本加除出版の「人名用当用漢字と命名関係の先例」には次のように述べている。

“変体がな”の解説と対照一覧表

“戸籍法”の第50条を受けた“戸籍法施行規則”の第60条には、「常用平易な文字の範囲」として、第3号に「片かな平がな」と規定し、カッコ書きして“変体がなを除く。”とある。したがって、これが施行された昭和23年1月1日以降は、子の名に“変体がな”は認められていないが、それ以前の命名には“変体がな”がなくはない。とりわけ、古い以前の出生の女子名には、戸籍の上でも“変体がな”が散見される。

“変体がな”は、また“草がな”ともいわれ、現今通常に用いられている“ひらがな”とは趣のちがったところの、一種のひらがな様の字体のものである。たとえば、「あ」、「お」、「ま」、「ぢ」……といった字体のものである。一般世上で最も多く見かける変体がなは、ソバ屋の看板の「出汁」^{しゆぢ} という — これも変体がなである。

漢字は、真名^{まな}ともいわれ、俗に本字^{ほんじ}とも称された。種々の漢字の楷書体がしだいに行書体へ草書体へとくずれてゆき、続け字がさらに度を増して各種のひらがな様の字体のものを多数に生じた。漢字の真名・本字^{ほんじ}に対して仮名^{かみな}といわれるゆえんである。明治33年(1900年)8月に文部省が省令として小学校令施行規則を公布し、そうした各種の字体のものの中から50音に対して1字あてを統一規定して、翌年の明治34年から小学校教育に実施したのが現在の「ひらがな」である。

このとき、世上には通用していながら、その「ひらがな」として公に採用されなかったひらがな様の字体のものが“変体がな”だと解することができる。だから、変体がなには、こんにちのひらがなの一步前の漢字に近い字形のものもあれば、全然別の漢字の形の変化したのものもある。変体がなを見て、その字源の漢字がすぐ想像できるものもあれば、たゞすくは字源を推測しにくいものもある。

古来、漢字は男子のもの^{おとこ}とされ、漢字の楷書体・行書体を男手^{おとこ}、草書体を「女手」と呼称したが、各種の変体がなは女手として和歌や女性の手紙等の手書きなどに用いられ、女性の名をしるす文字としても用いられてきた。昭和34年以後、義務教育からは変体がなは姿を消したが、女子の名に変体がなを用いることは大いに減少の傾向はたどったものの、全くあとを断ったというわけではなかった。明治34年前および後のそれらのことがこんにちに尾を引いて、現今の戸籍のうえで変体がなを散見するゆえんであるが、変体がなを普通のひらがなと対比させ、かつ、それぞれの字源の漢字とを対照して知っておくことは、変体がなを用いた名を読解したり、転記したりする際、有効だと思われる。

見かける度合いが多いと想像される変体がなについて、一覧表を作り、「変体がな・ひらがな対照表 — それぞれの字源つき」として別表にかかげておく。

3.4 かな使い上の問題

女性の名前の中にはかな文字だけの場合もある。この場合によく注意しなければならないことがいくつかある。

- ・新旧かな使い
- ・同音表記の誤り
- ・古いかな，カタカナ

旧かな使いが名前の中には残っている「かほり」「かほる」などは「かおり」「かおる」などと書かれる場合もある。また旧かな使いはあまり使わないため代理人が窓口で伝票を書いた場合などにも検索ミスが起きる。

同音の表記の誤りとしてはジ↔ヂ，ズ↔ヅ等がある。本人による誤りは少ないが，代理人（家族など）などがついうっかり誤って書く，オペレータが誤って入力をする場合がある。

古いかな，カタカナの書き方としては，

い↔ゐ え↔ゑ オ↔ヲ イ↔ヰ エ↔ヱ

がある。この古い字を使った名前があるが，日常生活で古い字を使わないため新しい字を使っている場合もあり，誤りの原因となっている。

3.5 小文字と大文字の表現

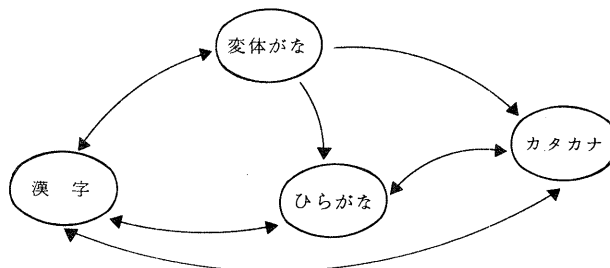
かな文字の名前の中に「や，ゆ，よ，つ」などのように拗音，促音などに小文字を使う場合がある。これを小文字として入力処理している場合がある。このような時大文字での検索を行ってもファイルに該当することができない。これは小文字の使用制限をするとか，小文字は検索用文字としては大文字と同じあつかいにするとかの処置が必要である。

3.6 特殊な文字の使い方

姓の中に3文字で表現するものがいくつかある。この中に中間の文字が添字的な役割の文字を使う場合がある。たとえば、「井之上」とか「井ノ上」「井の上」等の例がある。「之，ノ，の」はしばしば同一として我々の意識の中にある。これは注意しなければならないものである。これと同じような文字としては「ツ，ッ，つ，っ，津」「ケ，ヶ」がある。

3.7 変体仮名について

変体仮名といっても知らない人が多勢いるのではないかと思われる。しかしながら，変体仮名を使っている名前が多量のデータの中には数多くみつかるところ。これは法律で使用を禁止する以前に使用されていたのであろう。変体仮名は使われないので，カタカナ表記したり，ひらがな表記したり，漢字表記等ですましてしまうこともある。変体仮名，ひらがな，カタカナ，漢字は次のように変化する。



変体仮名はほとんど使われることはないが，市区町村の戸籍業務，除籍簿の検索には必要である。株式業務の中にも変体仮名が幾つか使われている。変体仮名は漢字表記で代行されたり，ひらがな，カタカナ表記で代行されるので注意する必要がある。変体仮名を使用しない場合でも女性の名前はひらがな表記をしたりカタカナ表記をするので，漢字ディスプレイの検索ではこれらについても対策をたてなければならない。

表2. 変体がな・ひらがな対照表 — それぞれの字源づき

へ	ふ	ひ	は	の	ね	ち	た	そ	せ	す	し	お	え	う	い	あ	が
部	不	比	波	乃	祢	知	太	曾	世	寸	之	於	衣	宇	以	安	な
部	不	比	波	乃	祢	知	太	曾	世	寸	之	於	衣	宇	以	安	な
魚	不	比	波	乃	祢	知	太	曾	世	寸	之	於	衣	宇	以	安	な
辺	婦	飛	八	北	能	知	島	不	世	老	之	於	衣	宇	以	安	な
邊	婦	飛	八	北	能	知	島	不	世	老	之	於	衣	宇	以	安	な
婁	婦	飛	八	北	能	知	島	不	世	老	之	於	衣	宇	以	安	な
婁	婦	飛	八	北	能	知	島	不	世	老	之	於	衣	宇	以	安	な
																	変体仮名とその字源
も	め	む	み	ま	ほ	ぬ	に	な	と	て	つ	さ	こ	け	く	き	か
毛	女	武	美	末	保	奴	仁	奈	止	天	州	左	己	計	久	幾	加
毛	女	武	美	末	保	奴	仁	奈	止	天	州	左	己	計	久	幾	加
免	馬	牟	見	未	保	奴	爾	奈	止	天	州	左	己	計	久	幾	加
免	馬	牟	見	未	保	奴	爾	奈	止	天	州	左	己	計	久	幾	加
茂	多	牟	三	満	本	奴	爾	奈	止	天	州	左	己	計	久	幾	加
茂	多	牟	三	満	本	奴	爾	奈	止	天	州	左	己	計	久	幾	加
業	而	舞	美	万	豊	奴	爾	奈	止	天	州	左	己	計	久	幾	加

			ん	を	魚		ら	よ		ゆ		や
			无	遠	惠		良	与		由		也
			元	遠	惠		良	与		由		也
			免	遠	惠		羅	与		遊		夜
				乎			頭	代				耶
							を	わ	乃	れ	る	り
							為	和	呂	礼	留	利
							井	和	呂	連	留	利
							為	王	路	礼	流	里
							遣	王	樓	麗	類	理

(日本加除出版「人名用当用漢字と命名関係の先例」より)

4. 検索システムを考えるにあたって

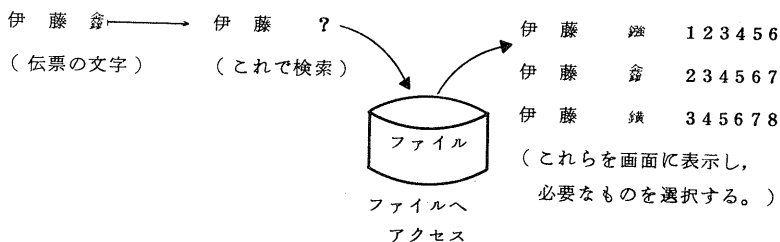
検索ミスを減少させるために原因をあげその改良方法について述べてきたが、それらの部分的改善以外に、運用とか教育とかで誤りの発生を防止することができる。ここではそれらについて4つの対策を考える。

1. 文字種の制限
2. ガイダンス機能の強化
3. ハンドブックの配布と教育
4. 漢字COM, ハード・コピーによるバックアップ

4.1 文字種の制限

銀行の窓口業務を考えると、ある一定時間内に何件かの照会応答を行なわなければならない。このような場合、見なれない文字の名前があると文字表から探さなければならない。たとえば、**鍛**とか**鑫**、**鑽**などの文字が現われると、この文字を見つけるだけで時間がかかる。これらあまり使用しない文字が現われた時は、対応する文字に?を表示し検索を行なう。このようにして、ある一定文字以外の例外的文字を使用した名前を検索し表示する。可能性のあるいくつかの名前の中より本人を確認し顧客名を見つける。

次に例をあげて説明する。



検索に使用できる文字は高々4,000文字ぐらいに制限すべきであろう。あまり多くの文字を許すことはあいまいさを多くしかえって検索に時間がかかる。またオペレータの教育訓練にも期間がかかる。

この文字種を制限するということは、検索用の文字を制限するのであって、宛名に使用しているような表記文字を制限するという意味ではない。

4.2 ガイダンス機能の強化

漢字姓名を使用して検索を行なって該当するものを見つけ出すことができない場合、オペレータがあるファンクション・キーを押せば、入力した漢字を分析し考えられる場合を想定した条件を漢字ディスプレイ上に表示し、オペレータに次に行なうべき方法を示すべきである。

このようにしてオペレータの質に関係なく、ある一定限度までは検索が成功するようにすべきであろう。

4.3 ハンドブックの配布と教育

漢字姓名による検索では、カナ文字を使用していた場合と異なった問題点が発生する。このため起こりうる問題点を列挙しハンドブックにまとめ、オペレータの教育資料とすとか、問題が発生した時解決方法を探しだす手引書とすると、たいへん便利である。

これら手引書は新しい事例が発生する都度内容を追加しファイルできるようなものであれば、オペレータの個々の経験が全員の経験となり、ミスの防止にもつながるであろう。

4.4 漢字 COM, ハードコピーによるバックアップ

いろいろな対策を考えても、やはり人間の目による検索に頼らなければならない場合がある。このために、漢字 COM による検索または漢字プリンタで印刷された索引簿を準備し、人間の目による検索も可能にすべきである。この索引簿も 1 種類用意するのではなく、顧客順、口座順、カナ索引順、漢字順等のものも用意すれば安全である。

このようなバックアップは各支店別に持つか、各支店はその支店だけのもので他店の顧客内容はセンターで管理し、検索の行えなかったものについての問い合わせを処理する等という方法が考えられる。

ハードコピーによるバックアップは、たいへん良い方法であるが、この内容の更新をどの程度のサイクルで行なうかということも考えておかなければならない。

[おわりに]

漢字姓名による顧客の索引システムは同姓同名が発生しにくくたいへん便利なものであるが、システム・デザインに十分検討を加えておかなければ失敗をまねくであろう。今後はこのような問題点を定性的に分析するばかりでなく、定量的に分析し解決方法を進めてみたい。

《参 考 文 献》

- 1) 校正ハンドブック 戸 台 俊 一 著 ダヴィッド社
- 2) 人名用当用漢字と命名関係の先例 日本加除出版
- 3) 昭和 50 年度国立国語研究所研究発表会要旨 国立国語研究所
(用語用字調査と機械処理)