

## 簡易住所入力による郵便番号検索に関する一検討

IW-1

加藤厚志 野田智嘉 磯部俊吉  
郵政省 郵政研究所

### 1 はじめに

郵政研究所では、郵便処理システムの情報機械化に関する研究を行っている。その要素技術のひとつに、ビデオコーディングシステムがある。これは、OCRで郵便番号が読み取れなかった郵便物に対して、郵便物に郵便番号情報を付定するシステムで、オペレータが郵便物のあて名イメージ画像を見ながら、住所を打鍵入力し、それにより検索された郵便番号を郵便物に自動付定するものである。

大量の処理件数が見込まれているため、このシステムの実現のためには、効率良く住所を入力する方法を検討する必要がある。その方法として、住所をベタ入力するのは非効率であることから、ある程度省略した住所を入力し、それをもとに郵便番号を検索する方法を考えている。

今回、全国の住所表記をいくつかの階層に分類し、各階層ごとに住所の一部あるいは全部を入力することによる郵便番号の検索効率を、何通りかの入力方法について検討・評価したので報告する。

### 2 入力文字数による住所の確定度

住所の入力方法を検討する準備として、全国約50万件の住所表記を、表1のように7階層に分類し、この各階層について、市郡名は都道府県別、区名・郡内町村名は市郡別と、上位階層の区画別に、漢字を頭から何文字入力すれば住所が一意に確定するかを調査した。

各階層について、その区画別の調査結果を平均したものを見ると、図1、図2に示す。

図1、図2を見ると、上位3階層の住所表記については、比較的少ない文字の入力で住所が一意に確

Finding a ZIP code by Simple Address Inputting  
Atsushi Kato, Chiyoishi Noda, Shunkichi Isobe  
Institute for Posts and Telecommunications  
Policy  
1-6-19 Azabudai, Minato-ku, Tokyo 106, Japan

定し、下位4階層については、多くの文字を入力しなければ住所が一意に確定しない傾向にあることが分かる。特に、第4階層の大字名・市内町名、第6

表1 住所表記の階層分類

階層	住所表記 (カッコ内は表記数)
1	都道府県名 (47) 上位
2	市郡名・都内区名 (1251) ↑
3	区名・郡内町村名 (3372)
4	大字名・市内町名 (105813)
5	通称名 (大字・字間区画) (123285)
6	字名・丁目 (496074) ↓
7	小字名 (497032) 下位

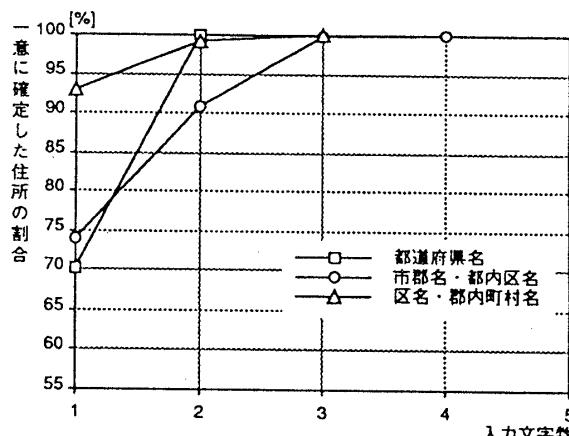


図1 漢字入力数による住所の確定度 1～3階層

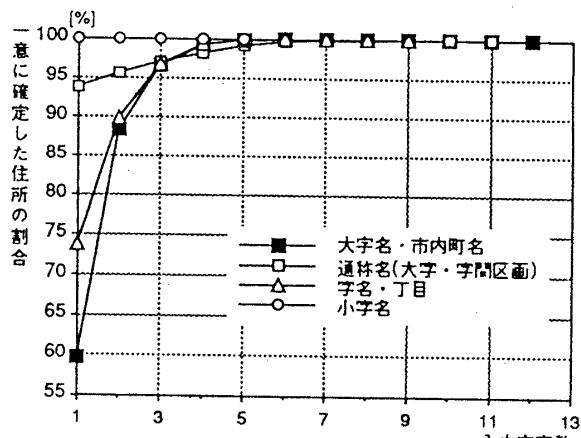


図2 漢字入力数による住所の確定度 4～7階層

階層の字名・丁目については、1～2文字の入力では確定割合が低いものとなっている。

### 3 住所入力方法の検討・評価

郵便番号は、おもに表1中の第4階層の区画ごとに割り振られている。

前節の結果より、この第4階層は、確定度が低いものとなっているため、郵便番号を効率的に検索するためには、この第4階層を確定させやすい入力方法をとる必要がある。

それに効果的であると思われる入力方法として、以下の2つの方法について、全国の住所表記を対象に郵便番号の検索効率を評価した。

#### 方法1：

第1階層（都道府県名）、第2階層（市郡名・都内区名）、第4階層（大字名・市内町名）、第6階層（字名・丁目）の、おのの漢字頭1文字を入力する方法。ただし、表記先頭に「大字」「字」が冠されている場合、それに続く1文字を入力し、また数字で始まる表記（丁目など）については、数字全桁を入力する。

例) 北海道旭川市北門町10丁目 の場合、

「北」「旭」「北」「10」を入力

#### 方法2：

第2階層（市郡名）の漢字頭1文字と、第4階層（大字名・市内町名）を全文字を入力する方法。ただし、表記先頭に「大字」が冠されている場合、それに続く文字を入力する。

例) 福島県伊達郡靈山町大字石田字中田 の場合、「伊」「石田」を入力

データベース検索条件は以下のとおり。

- ①前方一致検索をおこなう。ただし、数字を検索する場合、入力桁数とデータベース内数字表記桁数が等しくなければ不一致と判定する。
- ②階層ごとに検索をおこなう。つまり、方法1で入力される4文字は、それぞれ第1、第2、第4、第6階層の住所表記より検索し、また方法2で入力される文字は、それぞれ第2、第4階層の住所表記よ

り検索する。

- ③データベース内の住所表記先頭に「大字」「字」「小字」が冠されている場合、おののその次の文字から比較をとる。

郵便番号検索評価結果を図3に示す。

図3より、郵便番号が一意に確定する割合は、方法1では全住所表記の約84%であり、方法2では約90%であった。さらに、郵便番号候補数が3つ以下に確定する割合は、方法1では約97%、方法2では約99%であった。方法1では住所入力した各階層ごとに候補が多数発生したため、方法2に比べて郵便番号検索効率が若干低いものとなっている。

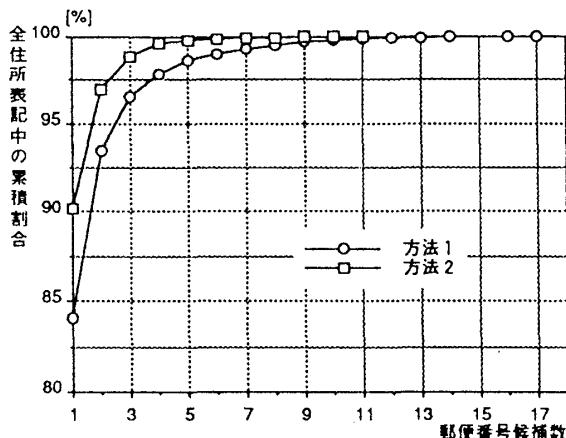


図3 簡易住所入力による郵便番号検索結果

### 4 おわりに

今回の評価した住所の入力方法では、郵便番号検索効率、入力のしやすさなどから、方法2の方が優れていると思われる。この方法では、郵便番号を一意に確定出来る割合は90%であるものの、候補をかなり絞り込むことが出来るため、検索された候補を選択入力することにより、効率良く郵便番号を検索出来るものと思われる。

今後の課題として、今回評価した方法は漢字入力を前提としているため、その入力インターフェースについて、かな漢字変換入力、手書き文字認識入力など入力方法の検討・評価を行う予定である。

### 参考文献

- [1] 丸川、古賀、嶋、藤澤：住所知識を用いた入力文字列の不確実性の解消、平3秋季信学全大、D-51