

簡易住所入力による住所確定の検討

3Q-8 尾坂 忠史[†] 中村 嘉明[†] 上原 仁[†] 田中 照隆[†] 加藤 厚志[‡][†]郵政省 郵政研究所 [‡](株)日立製作所 応用システム本部

1. はじめに

現在、郵便物はOCR（光学式文字読取装置）によりあて名を読み取り区分しているが、郵便番号及びあて名を読み取れなかった郵便物の処理は、手作業に頼っている。そこで、OCRで読み取れなかった郵便物に対してオペレータがあて名情報を補完するビデオコーディングシステム（VCS）について検討している[1]。VCSは、記載された郵便番号がOCRで読み取れなかったため、オペレータが郵便番号を打鍵入力するものと、郵便番号が未記載であるためあて名情報を打鍵入力し、郵便番号に変換して付定するものの2種類あるが本研究では後者を対象とした[2]。

VCSは、大量の郵便物を処理するため、効率良く住所を入力する方法について検討する必要がある。

郵便番号未記載郵便物は、VCS処理においてオペレータが住所を入力しなければならない。しかし、あて名住所を全て入力する場合、入力時間がかかるため、簡易的に住所の一部を入力（簡易住所入力）することにした。

本報告では、この簡易住所入力を、効率的に郵便番号を確定できる入力方法について検討した結果を報告する。

2. 全国住所表記の検討

住所の簡易入力方法を検討するに当たり、全国住所表記の統計処理を試みた。まず、全国の住所表記を表1のように7階層に分類した。住所表記は、1階層から7階層に向かって樹木状に分岐している。

表1 住所表記分類

階層	住所表記分類
1	都道府県
2	市郡、都内の区
3	市内の区、郡内の町村
4	大字、市内の町等
5	町内の大字、字の付かないもの
6	字
7	小字

住所表記には東京都府中市、広島県府中市のように市郡名等が重複したものがある。よって、住所を検索するためには、複数の階層の文字を入力することが必要である。

1階層から3階層までの表記については、階層区切りを判別可能であるが、4階層から7階層の表記については、階層区切りを判別するのは困難である。（例：熊本県阿蘇郡蘇陽町大字塩出迫下塩出）このため、4階層以下の階層区切りをオペレータが判別しなくてもよい入力方法が望ましい。

また、4階層の使用文字数は住所表記全体の97%が4文字以下であるため、表記5文字目については、それが5階層以下である可能性が高い。

3. 住所入力評価結果

上記検討結果に従い、入力方法を選定し結果の良好だった5つの入力方法について評価した

Study on the Determination of a Address by Simple Address Inputing

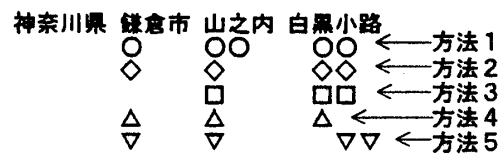
T.Osaka[†], Y.Nakamura[†], H.Uehara[†], N.Tanaka[†], A.Kato[‡]

[†]Ministry of Posts and Telecommunications

(1-6-19 Azabudai, Minato-ku, Tokyo, 106, Japan)

[‡]Hitachi, Ltd (1 Ikegami, Haruoka-cho, Owariasahi-shi, Aichi-ken, 488, Japan)

結果を述べる。(入力方法は、図1に示す。)



- 方法1：2階層の1文字目、4階層以下
1、2、5、6文字目を入力
- 方法2：2階層の1文字目、4階層以下
1、5、6文字目を入力
- 方法3：4階層以下1、5、6文字目を入力
- 方法4：2階層の1文字目、4階層以下
1、5文字目を入力
- 方法5：2階層の1文字目、4階層以下
1文字目、最後の2文字を入力

図1 入力方法

なお、4階層以下に「大字」「字」「小字」が冠されている場合は、これを除いて数えた文字を入力することとし、住所表記が各入力方法規定の文字数に満たない場合、繰り上げた位置の文字を入力することとした。

以上の入力方法を用い、全国約44万件的住所表記全てに対して、同じキーワードを持つ住所表記に分類した。そして、それぞれのキーワードに対応する住所表記数を計上する統計処理を行った。住所の検索効率を評価した結果を図2に示す。

方法5は比較的検索効率が良いが、あて名の記載傾向として、字、小字以下の表記は省略される場合があるため、データベース上の住所表記の最後の2文字を必ず入力することが不可能となる。実際の郵便物を考慮すると、必ずしも実態を表現しているとは言い切れない。方法2と方法3を比較すると、2階層の入力の有無で検索効率に大きな開きが出ている。このため、2階層は入力する必要がある。4階層以下の入力文字数が3、4文字である方法1、方法2では1件に確定できた住所が70%台であるが、2文字である方法4では、1件に確定できた住所は40%台まで落ち込んでいる。これにより、4階層以下は最低3文字入力しなければ住所候補

が絞り込めないことが分かる。

方法2により入力すれば、約44万件的住所表記に対し約94%の住所は住所候補数を5件以内に絞り込める。

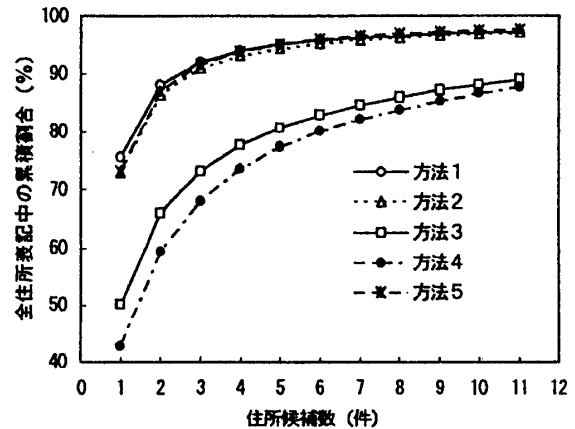


図2 住所候補数からみた入力方法の評価結果

4. おわりに

郵便番号未記載郵便物に対するVCSの住所入力方法について検討し、以下の結論を得た。住所表記の4文字を入力することにより、約44万件的住所表記に対し約94%の住所は住所候補数を5件以内に絞り込めることが分かった。

今後の課題として、以下を挙げる。

- ・入力実験による処理速度、効率の測定。
- ・実際の郵便物量と組み合わせた住所検索。
- ・あて名住所の誤字脱字、旧書体、旧住所表記についても柔軟に検索できる検索アルゴリズム、住所データベース構造。
- ・OCRの認識結果に今回の入力方法を組み合わせる方法。

5. 参考文献

- [1] 加藤、野田、磯部：簡易住所入力による郵便番号検索に関する一検討、情報処理学会第47回全国大会、1W-1、1993
- [2] 田中他：簡易住所入力によるビデオコーディングシステムについて、第7回郵政研究所研究発表会論文、1995-02