

住所情報入力支援のための自動訂正方式

5V-8

武藤 信夫 川辺 秀樹 岩瀬 成人 福村 好美

NTT 情報通信研究所

1. はじめに

近年のハードウェア技術、データベース技術の進歩により、大規模な顧客情報（名義、住所、職業、電話番号等）を検索するシステムが実現され、電話などを通じて不特定の利用者からの問い合わせをオペレータが受け、目的とする情報を案内するサービスが提供されている。この場合、提示情報誤り、聞き違いあるいはキー入力誤り等により、オペレータの対応の長時間化を招いている。

本報告では、顧客情報検索システム等において目的とする情報を特定するために重要な住所情報を対象に、逐次的な言語解析処理（形態素・構文・意味解析）を適用することにより、「分かち書きを意識しない住所入力」を実現するとともに、含まれる誤りを即時に指摘し、一意に特定できれば正解値に自動訂正する方式を提案する。さらに、本方式を実験システムとして構築しシミュレーションにより自動訂正率と、解析処理速度を評価した結果を報告する。

2. 住所情報に対する入力支援

一般に、データベースに住所情報を格納する場合、格納効率、曖昧性除去のため、JISに準拠した住所コードが使用される。検索時には住所コードの住所階層（都道府県/市区郡町村等）に従ってデータベースに接近するため、住所を入力する場合にはその順序で階層毎の区切りを付加することが必要であった。しかし、広域の住所を対象に電話により問い

合わせを受ける場合、その区切りの挿入位置と階層構造との整合性が不明確である。その地域についての知識がないオペレータにとっては、申し出情報を区切りを意識しないでそのまま入力できることが望ましい。

また、初期の単純な入力ミスが対応の長時間化要因となりうるため、入力した住所情報が誤まっている（解がない）ことを指摘できる場合には、即時にオペレータに割り込んで通知し、その時点で再確認できること、さらに、一意に確定できる場合には、誤りを訂正する等が、住所情報の入力支援に有効である。

3. 自動訂正可能な逐次住所解析方式

前章での住所入力支援機能を実現するため、文字単位横型探索と、係受け関係に住所名の包含関係に対応させた言語解析処理方式を考案した。

図1に、提案の逐次住所解析方式の構成を示す。

Error Correction Method for Assistance of Inputting Address

Nobuo Mutoh, Hideki Kawabe, Shigehito Iwase, and Yoshimi Fukumura

NTT Information and Communication Systems Laboratories

3-9-11 Midoricho, Musashino, Tokyo 180, Japan

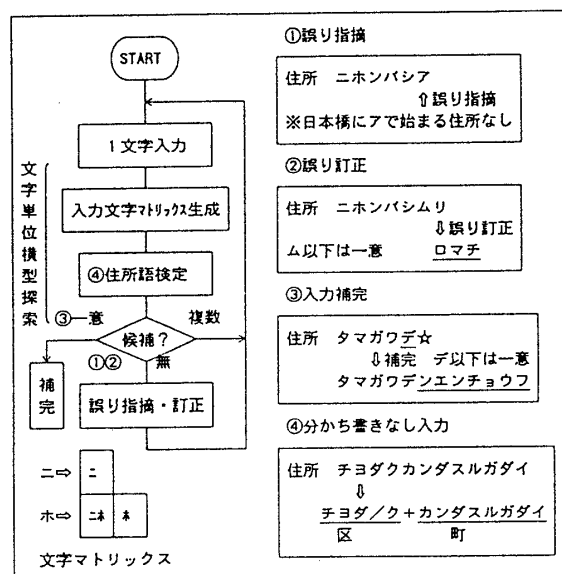


図1. 住所解析方式の構成

本方式は、全ての文字列分割候補の有無を表すマトリックスを作成し、辞書引きの結果を入力マトリックスの各要素に格納する。本マトリックスを逆にたどることにより候補となる住所を捜し出し、同時に、住所の包含関係をチェックすることにより、形態素解析と意味解析を実行する。

本方式では、①自然言語と異なり入力住所は複合名詞である、②入力される住所は予め全て既知である（全国の字名まで含めた住所は約40万件である）、③単語間の係受けは包含関係のみである、④有力な候補を探すのではなく可能性のある候補全てを探す必要がある、等の住所情報の特性に着目し、マトリックスを用いて形態素解析及び意味解析を同時に行うことにより処理の高速化を図っている。

さらに、高速化手法として、存在しない住所情報が入力された時、最後に入力された文字の正誤が確認できれば目的が達成できることから、マトリックスを上位にたどる際、候補が2個検索できた時点でマトリックス検索処理を終了することとした。すなわち、候補が2個以上あれば入力誤りではないが、住所を一意に確定することはできないため、その時点でシステムは次入力を待つこととした。

4. 実験評価

(1) 解析性能の評価

全国の住所（市名+町名）を対象に、次の解析シミュレーションを実施した。

- ①住所階層の順序を逆転した分かち書きなし入力
- ②代表的な誤入力3パターン（文字の倒置／文字の抜け／不要文字の挿入）の誤った町名の入力

①については、100%解析に成功した。

②については、図2が示すように、キーミスの時点で99.5%誤りを指摘でき、キーミスの位置が3文字目であれば約50%、6文字目以降では80%以上誤りを訂正できる結果を得た。

(2) 処理性能の評価

全国／地方（5県）／県の3サイズの住所解析用

単語辞書を用い、問合せの多い「市名+町名」の住所について実測した処理時間を図3に示す。

- ①住所名（市名+町名）の87%が16文字以内であり、地方サイズ以下の辞書の場合キー入力速度を0.4秒とするとキー入力に十分追随可能であると言える。
- ②文字単位横型探索処理はマトリックスによる解析のため、処理時間は入力文字数の2乗に比例すると考えられるが、本方式ではマトリックス検索処理において候補が2個以上あった時点で次入力処理に移行することにより、処理時間をほぼ1乗のオーダーに減少させることができた。

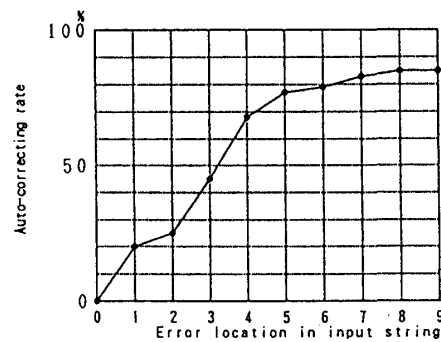


図2. 入力ミスの位置別の自動訂正率

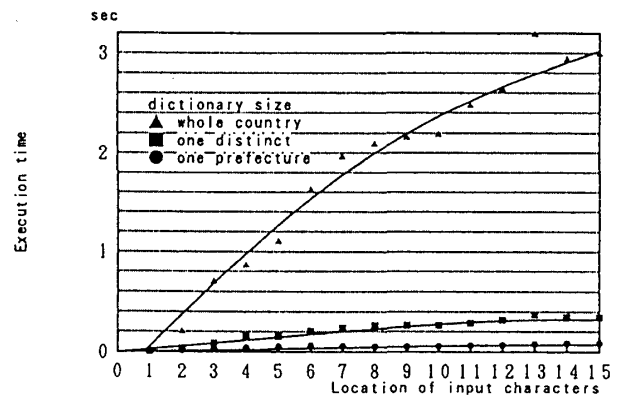


図3. 文字単位横型探索方式の処理速度

5. おわりに

オペレータを対象とした住所入力の支援には、逐次的な言語処理が効果的であり、分かち書きなし住所を1文字入力毎に解析する住所入力方式を考案した。本方式の逐次的解析手法は、一般の係受け文法を適用すれば、自然文の解析にも適用可能である。