

文字遷移利用クイック検索システム

2H-3

郡司 圭子 桂 晃洋 葛貫 壮四郎

(株)日立製作所 日立研究所

1. 背景と目的

ペンコンピュータによる伝票処理では、住所や定形句の入力が多い。しかし、住所を全部手書きするのは大変である。メニューで階層の上位から選ぶ場合には、階層の上位が分からないと、下位を選択できない。郵便番号を入力して住所を検索する場合、自分の住所以外の郵便番号は覚えにくいという問題があった。そこで、住所のうちの分かっている一部分（例：「大みか」）を手書き入力するだけで、目的の住所（例：「茨城県 日立市 大みか町」）を簡単に入力できるシステムを開発することにした。

2. 手書き住所入力の課題

住所の一部分を手書き入力するだけで、目的の住所を簡単に入力するためには、次の課題がある。

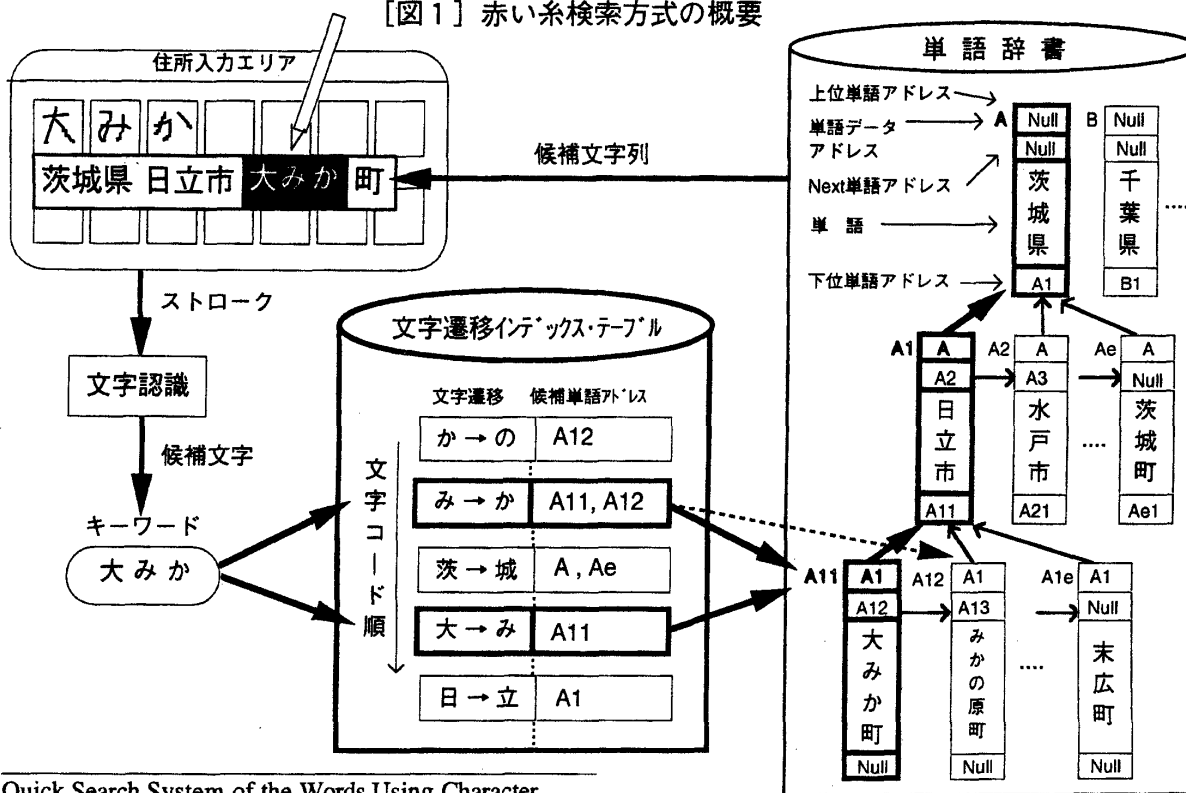
- (a) 曖昧な入力からの高速な検索
- (b) 検索結果を絞る条件の容易な追加
- (c) ユーザの意図を反映した候補表示

3. 文字遷移利用クイック検索システム

曖昧な手書き入力から高速に住所を検索する方式として各種の方式を検討した。曖昧な入力から正しい住所を高速に検索でき、絞り込み条件の追加も簡単なことから、文字遷移インデックステーブルと単語辞書を用い、文字遷移インデックスには文字遷移を含む単語へのポインタを設ける方式（赤い糸検索方式）を採用した。

図1に、赤い糸検索方式の概要を示す。キーとなる文字「大みか」を手書き入力すると、文字認識して文字コードに変換される。キーワード「大みか」は文字遷移「大→み」「み→か」に分解され、これらの文字遷移は、文字遷移インデックステーブルの候補単語アドレスより、単語辞書(A11)の「大みか町」に関係していることが分かる。そこで、アドレス(A11)に直接アクセスし、単語辞書の階層情報を使って「茨城県 日立市 大みか町」を得る。

[図1] 赤い糸検索方式の概要



Quick Search System of the Words Using Character

Transition Information

Keiko Gunji, Kouyou Katsura, Soushirou Kuzunuki

Hitachi, Ltd. Hitachi Research Laboratory

4. 検索結果の優先方法

「大みか町」は、ユニークな地名であった。しかし、例えば、「末広」のつく町名は、全国に150以上もある。このように、複数存在する地名を検索する場合でも、キー文字に市郡・都道府県名などの情報を追加することにより、候補を絞り込むことが可能である。そこで、例えば、「茨城県 日立市 末広町」を検索するために、図2のように、キー文字を「末広日立」と、町名と市名を区切らずに並べて入力できるようにした。町名と市名を区切らなくても、文字遷移に分解すると「末→広」「広→日」「日→立」と、町名や市名の一部になり、これらの文字遷移を含む候補単語に、「末広町」や「日立市」が得られる。単語辞書より、「末広町」は「日立市」の下位の階層につながっているため、「日立市 末広町」と、まとめて考えることができる。

候補単語を表示する際には、ユーザが欲しい情報が上位になるように並べる必要がある。ユーザが入力したキー文字全てを含む結果がもっとも嬉しいが、キー文字全てを含む結果に限定すると、一部誤りがあっても検索不可能になってしまう。そこで、キー文字を多く含む住所ほど優先的に表示することにした。すると、図2の例では、キー文字すべて(4文字)を含む「茨城県 日立市 末広町」が第1候補になり、キー文字2文字を含む地名が第2候補以降になる。(キー文字の数で優先がつけられない場合には、階層順や、単語中のキー文字の位置などに着

目して検索結果をソートする。)

5. 文字遷移利用クイック検索の特長

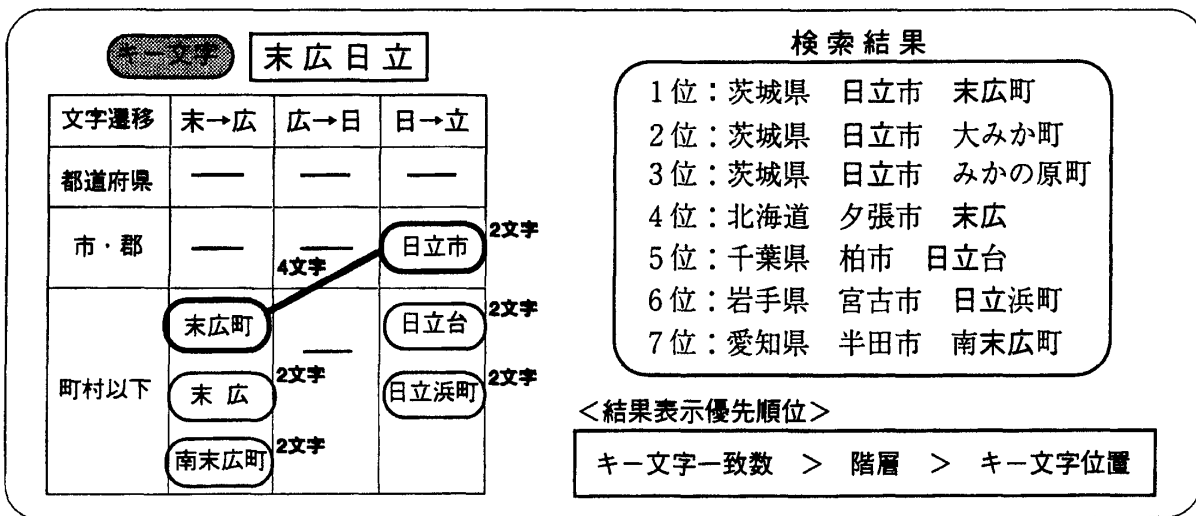
文字遷移利用クイック検索は、赤い糸検索方式を用いるため、次のような特長がある。

- (a) 文字遷移インデックスにより単語辞書に直接アクセスするため、高速な検索が可能
- (b) 文字遷移単位に分解して候補を抽出するため、キー文字に一部誤りがあっても、他の部分で候補を抽出可能
- (c) キー文字を文字遷移に分解し、後から単語辞書を用いてまとめ直すので、キー文字の順番が正確でなくても検索可能(検索絞りキーの容易な追加)
- (d) キー文字を含む数や階層順による候補のソートで検索者の意図する検索候補を優先表示

6. 評価

本方式による住所入力を、従来の住所全文字を入力して認識する方式と比較評価した。評価のために、乱数を用いて無作為に住所のサンプルを10個選び、10人を対象に、従来方式と新方式で、住所を正しく入力できるまでにかかる時間を測定した。その結果、<表1>のとおり、サンプルに用いた住所入力時間を、平均で48.7%に削減することが分かった。(従来方式と新方式の入力時間の母平均が等しいという仮説をたて、有意水準1%で検定したところ、仮説は棄却された。母平均の99%信頼区間が重ならないことから、有意な差があるのは明らか。)

[図2] 検索結果の優先順位決定方法



<表1> 住所入力にかかる時間比

	(新方式) / (従来方式)
10人の平均	48.7%

<参考> サンプル住所入力にかかる時間 (秒)

	(新方式)	(従来方式)
10人の平均	286	586
母平均の99%信頼区間	220~351	489~683