

PB入力型電話番号検索実験システム

デモ1

佐藤 亨 林 智定 永井良史 藤岡健吾 野田良輔

NTT 情報通信研究所

1. はじめに

プッシュボタン（PB）信号を送出可能な電話機を入力端末とした、電話番号検索方式の研究を進めている。^{[1][2]}

本方式は、以下の要素技術から成る。

- ① PB電話機の12個の限られたキーを用いて、情報を入力するための日本語入力技術
- ② 情報検索に必要な情報を聞出す対話誘導技術
- ③ 入力情報から膨大な電話番号DBを検索する検索技術

本稿では、これら要素技術を概説し、これら技術を統合したPB電話機による電話番号検索システムを紹介する。

2. システム構成

PB入力型電話番号検索システムは、C/S形態のセンタと電話回線を介して接続されたPB電話機からなる。（図1）

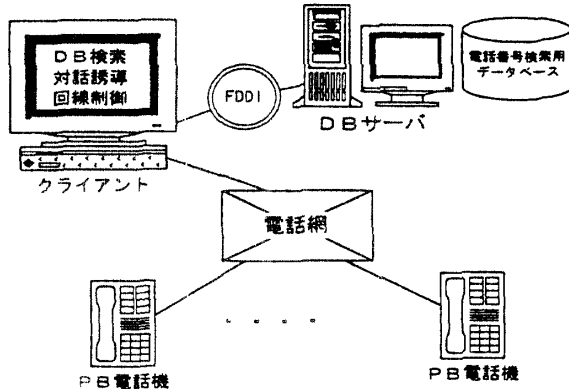


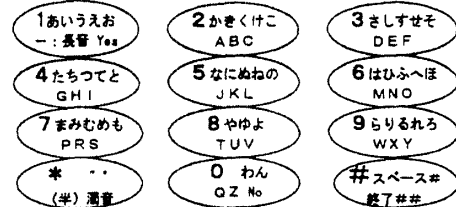
図1 システムの構成

3. PB電話機による日本語入力技術^[3]

キー数が限定される電話機からの入力方法としては、各情報にコードを付与し、コード番号を直接入力する方法や、情報を仮名読みし、その仮名読みをキーの組合せで入力する方法が一般的に採用されている。

このため、コード表の照会が必要であったり、かな入力に複雑なキー操作を必要とし、利用者に負担とストレスを強いてきた。

そこで当システムでは、図2に示すように1つのキーに複数のかな文字を割り振り、1回のキー押下でかな1文字を入力させる、文字情報縮退法^{*1}を考案した。この方法によれば、「よこはまし」なら、「よ」が「8」、「こ」が「2」、「は」が「6」、「ま」が「7」、「し」が「3」であるから「82673」として入力することができる。



投入例：よこはまし - 82673##
やまだらう - 874**491##

図2 PBボタン・文字割り振り

一方、この文字情報縮退法を用いた場合のPBキーによる表現能力と日本語の46音に濁音・半濁音を加えたかな71文字の表現能力を図3に記す。

さらに現実世界で使用されている表現において、PBキーによる表現能力とかな文字の表現能力の差を同じく図3に色付けして示す。この図から12種類のPBキーによって、かなりの表現が可能であることが判る。

文字情報縮退法は、キー押下数も少ないことから誤入力の発生頻度も低く、軽快に入力することが可能であり、会話用の入力手段として優れている。しかし、例えば“2”の押下で“か”～“こ”の、どの1文字を入力されたのか特定できないために、曖昧性を伴う。

そこで、検索条件となる入力情報の持つ付随情報^{*2}を利用した実在検定を実施し、目的外の候補を除外したり、後述の対話誘導により利用者から自然な対話の中で追加情報を入手し、減少した情報を補完している。

A Fully Automated Directory Assistance System using a Telephone Keypad
Toru SATO, Tomosada HAYASHI, Ryoji NAGAI, Kengo FUJIOKA, Ryousuke NODA
NTT Information and Communication Systems Laboratories

1: 仮名以外に、濁音や半濁音用に「」、入力の区切り／入力の終了用に「#」／「##」を割り振っている

*2: 横須賀市は神奈川県にあるといった相関関係等

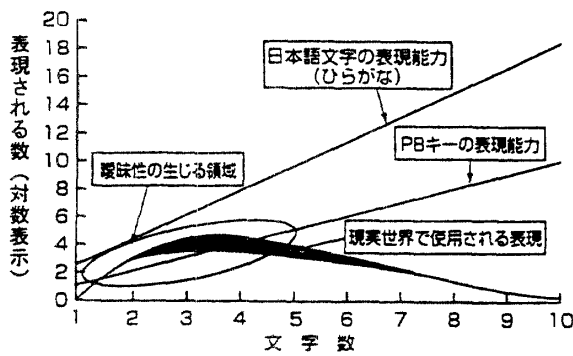


図3 日本語の表現能力の冗長性^[6]

4. 対話誘導^[4]

電話番号の検索に必要な情報の入手に加え、文字情報縮退法による曖昧性、場合によっては利用者の記憶違い／曖昧さを解消する手段が必要である。

そこで、電話番号検索システムでは、利用者の求める解に効率的に到達できることを目的としたガイドランスを用いており、必要な情報を聞き出す知的作業である対話誘導を、以下の四つの戦略を立てて実現している。

- ① 利用者の入力が最小となるようにする。
- ② 利用者には、文字情報縮退法に起因した曖昧さがあることを意識させない応答をする。
- ③ 利用者には要求されても当然と思われる質問をする。
- ④ 一度要求した情報、既知の情報から推定できる情報について質問しない。

5. 検索法

5.1 形態素解析を利用した検索高速化^[5]

入力された断片的な情報が、検索対象に含まれるか検索する場合、検索対象データの先頭から1文字ずつらしながら一致確認する方法が広く利用されている。この方法では、検索対象データが大規模になると、検索効率が著しく低下するという問題があった。

正式名称	関	東	不	燃	建	築	工	房
生成した インデックス	関	東	不	燃	建	築	工	房
	不	燃	建	築	工	房		
	建	築	工	房				
検索キー	建	築	工	房				

図4 階段状インデックスの構成

そこで、検索対象データを形態素解析し、意味のある単語単位に分解し、この単語単位で始まる階段状のインデックスを生成させ、インデックスの前方から一致を試みる検索法を導入している。

例えば、図4に示すように、検索対象の正式名称に対し、単語単位に始まる4つのインデックスを生成し、入力された検索キーでこのインデックスを検索することで検索の高速化を図っている。

5.2 単語区切りを用いた絞り込み^[6]

入力された断片的な読み仮名情報を元に、検索する場合、入力された読み仮名を含む全てを検索候補とすると、目的以外の候補が多数挙がってしまう。そこで候補となったものから不要な候補を効率的に排除する必要がある。

そこで「単語」に対応した読み仮名に区切り記号を付けてDBに登録し、利用者の入力情報がDBに登録された区切り記号の位置と一致しない候補を排除し、候補を絞り込む手法を導入している。

5.3 漢字情報を用いた絞り込み^[7]

日本人の人名には、同音異字（ヒロシ=弘、浩、宏、博、寛、洋等）が多数存在する。このため、電話番号の名義を確定しようとする、この漢字情報を有効に使う必要があるが、音声を前提とした従来型の電話機では、視覚を前提とした漢字の識別は容易ではなかった。

そこで、漢字の構成を簡易な説明文（例：ヒロシという字は、太平洋の洋ですか）で利用者に音声で説明し、選択させる処理を導入している。

6. おわりに

本稿では、PB電話機を入力端末とした電話番号検索システムの要素技術を概説した。

今後は、さらなるシステムの高度化を目指し、住所の代わりに著名な建物等の目標物名を入力させる方法、無駄な対話を少なくする手法、等について検討する予定である。

参考文献

- [1] M.Higashida, Fully Automated Directory Assistance Service that Accommodates Degenerated Keyword Input Via Telephones, PTC, (1997).
- [2] 林他, PB電話機を利用した電話番号案内技術, 97信学総大, 1997.
- [3] 佐藤他, PB電話機を利用した日本語入力技術, 97信学総大, 1997.
- [4] 奥他, PB電話機を利用した電話番号案内に適した対話誘導戦略, 97信学総大, 1997.
- [5] 野田他, 形態素解析を利用したデータベース検索高速化方法, 97信学総大, 1997.
- [6] 永井他, 文字区切り単語区切りを用いた検索解の絞り込み効果, 97信学総大, 1997.
- [7] 藤岡他, PB電話機を利用したデータベース検索における漢字情報の利用方法, 97信学総大, 1997.
- [8] NTT・情報通信研究所・東田正信の個人情報