

## オンライン手書き文字認識 (特徴点逐次対応法) における辞書DB編集ツール

6E-1

植田 郁子      大橋 勝之      石垣 一司

(株) 富士通研究所

### 1. はじめに

オンライン手書き文字認識方式「特徴点逐次対応法」  
[1] で使用する認識辞書は、様々な属性を付加した特徴点 (端点、屈曲点) の列である。先に、辞書を DAG (Directed Acyclic Graph) 構造で表現し、認識処理を高速化する方式 [2] [4] を示した。この認識辞書は、特徴点を DAG のノードに対応させ、複数の筆順からなる辞書パターンを1つの DAG で表現したものである。DAG のノードには、特徴点の座標値と属性値が記憶されている。  
認識に使用する辞書は、収集パターンより自動合成するのではなく、マニュアルで辞書DBを作成し、このDBから合成する。このような方式を採用したのは、以下の理由による。

- ・追加・削除・編集が容易に行える。
- ・認識アルゴリズムの変更に対応できる。
- ・認識辞書が DAG 構造であるため、収集データからの自動合成が困難である。

先に辞書の基となる情報を格納する辞書DBを作成・編集するツールを作成した。[3] 本報告では、作成の手順を中心に述べる。

### 2. 辞書DBとは

辞書DBは、部分パターンを単位に階層構造で情報を管理している。部分パターンは、さらに分割可能なMOL型と、それ以上分割できないATOM型に分類され、それぞれ次のような情報で構成される。

#### (1) MOL型

- ・部分パターンコード
- ・部分パターン間の筆順 (最大4個)
- ・大きさ、位置情報 (最大14個)

(1つ上位のパターンを±50で正規化した際に、各部分パターンがどこにどれくらいの大きさで位置するかを、そのパターンを構成している部分パターンの外接枠で表現する。)

#### (2) ATOM型

- ・特徴点列の座標点(x,y)情報
- ・ストロークを単位とした筆順 (最大8個)
- ・筆順毎の各特徴点列の属性

特徴点の属性には、次の5つがある。

<u>省略不能点</u>	書き方によらず安定に抽出される点
<u>方向不定点</u>	次の特徴点への方向が不安定な点
<u>長さ不定点</u>	次の特徴点への距離が不安定な点
<u>例外点</u>	優先順位を最下位にして対応づける点
<u>しきい値</u>	例外的に標準値以外の対応づけしきい値 (許容する対応づけ距離 [1] の最大値) を格納する

- 部分パターンを単位に管理すると、部分パターンを複数の漢字で共有することができるため、次のような利点がある。
- ・部分パターンの情報に対する追加/削除/変更は、その部分パターンを含む全ての漢字に直ちに反映される。
  - ・各部分パターンに高々数個の筆順を登録するだけで膨大な筆順を表現できる。
  - ・辞書DBのサイズを削減できる。

### 3. システム構成とMMI

- ・本システム構成を図1に示す。本システムは、UNIXマシン(SUN3)とパソコン(FMR70)をイーサネットに接続した構成で、パソコンには座標入力用にマウスとタブレットを接続している。データの作成や編集の図形処理と主制御はパソコン側が行い、UNIX側ではパソコン側の要求によりload/storeなどのDB管理を行う。

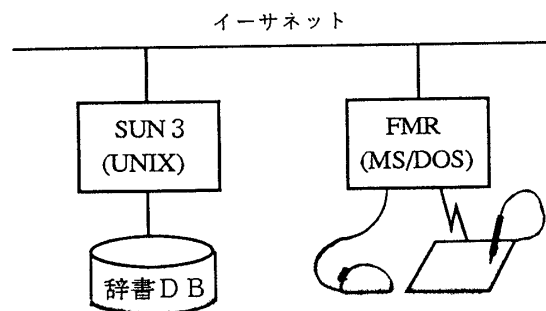


図1 システム構成

A Dictionary Data Base Management System  
for An Online Hand Written Character Recognition

Ikuko UEDA, Katsuyuki OHASHI, Kazushi ISHIGAKI  
FUJITSU Laboratories Ltd.

また、全ての操作は視覚的・直感的に行えるものでなければならない、という方針に従い、日本語メニュー選択を基本として、マウスとタブレットのみですべての操作が行えるようにした。これは、オペレータは辞書DBを作成期間システムの前にかなりの時間拘束されるため、疲労などによる誤操作を最小限にする、という配慮によるものである。

#### 4. 辞書DB作成手順

本ツールによる辞書DB作成手順を図2に示す。

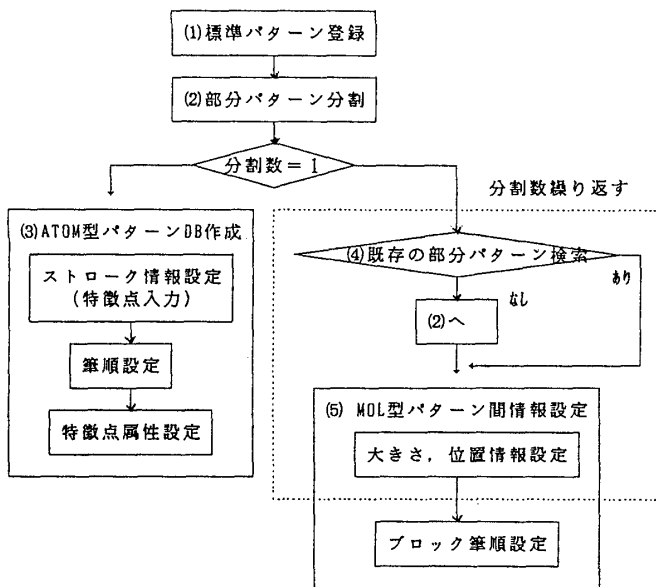


図2 辞書DB作成手順

(1)標準パターン（お手本パターン）登録は、タブレットより入力して登録する。

(2)部分パターン分割は、(1)で登録した標準パターンのストロークをマウスクリックで選択して、連続して選択されたストローク集合で部分パターンを構成する。これを繰り返し過不足なく部分パターンに分割する。

(3)ATOM型パターンDB作成では、マウスで特徴点を入力して、その特徴点の集合でストローク情報を作成する。次に登録すべき筆順を手書きで入力してストロークに順番をつけて筆順を登録し、筆順毎に特徴点に必要な属性を設定する。ATOM型DB編集画面を図3に示す。

(4)複数の部分パターンに分割した場合には、分割した各部分パターンに対して既存のDBの中から検索する。DBの中から検索されなかった場合には、部分パターン分割を再帰的に呼び出しDBを新規に作成する。同じも

のがあった場合それを共有する。

(5)部分パターン間の情報の設定は、部分パターンの外接枠の左上と右下の座標をマウス入力して大きさ・位置を設定した後に、並べて表示してある部分パターンをマウスクリックした順で部分パターン間の筆順を設定する。

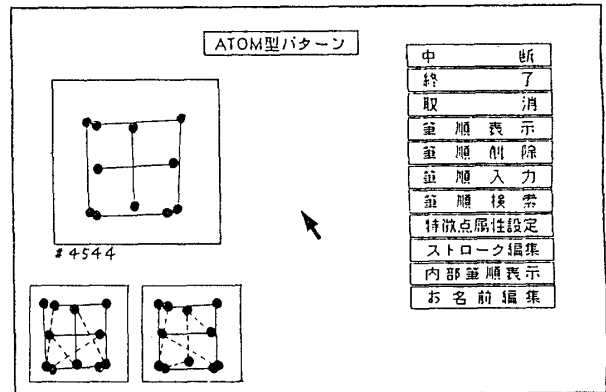


図3 ATOM型DB編集画面

#### 5. まとめ

本ツールを使用して、JIS第一水準および第二水準漢字3000字以上と非漢字の辞書DBを短期間に作成することができた。

現時点での辞書DBの諸量を表1に示す。

登録カテゴリ数	3495 (漢字 3317, 非漢字 178)	
部分パターン数	ATOM型 810	MOL型 3641
平均筆順数	2.28	1.01

表1 辞書DB

性能を向上させるため、現在以下のような改良を行っている。

- ・ATOM型パターンに対する最大筆順数の制限を緩める。(8から16に)
- ・部分パターン間での例外情報設定を可能にする機能を追加する。

#### 【参考文献】

- [1] 石垣他: 昭和61年後期、情処全大、1633
- [2] 石垣他: 昭和62年前期、情処全大、1843
- [3] 石垣他: 昭和62年後期、情処全大、1953
- [4] 大橋他: 昭和63年後期、情処全大、1640