

印刷文書認識システム AutoReco/2

2L-1

—システムの構成と概要—

加藤美治[†]、山下晶夫[‡]、平山唯樹[‡]日本アイ・ビー・エム株式会社[†]大和研究所/[‡]東京基礎研究所

1 はじめに

電子ファイル・システムによる保管と検索、さらにはCD-ROM等の新しいメディアの登場に伴い、オフィス環境は急速に変化している。しかしながら、オフィス情報の大部分は依然として『紙』のままであり、情報の分類、保管、検索、加工、および再利用を困難にしている。すなわち、情報を蓄積して活用するためには既存の情報の再入力が必要であり、そのデータ再入力コストが電子ファイル・システム等を構築する際に大きな障害となっている。

既存の印刷文書の再入力に関しては、OCRによる文字認識技術に大きな期待が寄せられているが、テキスト領域やレイアウトの構成要素をいちいち指定するのはユーザーにとって手間のかかる作業である。さらに、将来的には文書をイメージ・データとしてではなくマルチメディア・データとして管理するデータベースが普及すると予想される。このマルチメディア文書データベースでは、テキスト情報、イメージ情報、図情報、表情情報等を蓄積し、文書のレイアウト/論理構造も合わせて記録しておくことが必要とされる。また、認識結果の修正に要する時間の短縮は、認識処理のスループットを向上させるためには重要である。

筆者らは文書のレイアウト解析、認識結果の文脈後処理、およびグラフィックス表示によるユーザー・インターフェースを備えた日本語印刷文字認識システムの開発を通して上述の問題点の解決を図ったのでその概要を報告する。

2 システム構成

図1にシステム構成を示す。文字認識以外の全ての処理はパーソナル・コンピュータ PS/55 上で行われており、OS/2の下でマルチスレッド・アプリケーションとして実現されている。

多様な要求に対応するために、スキャナーからの読み

PS/55

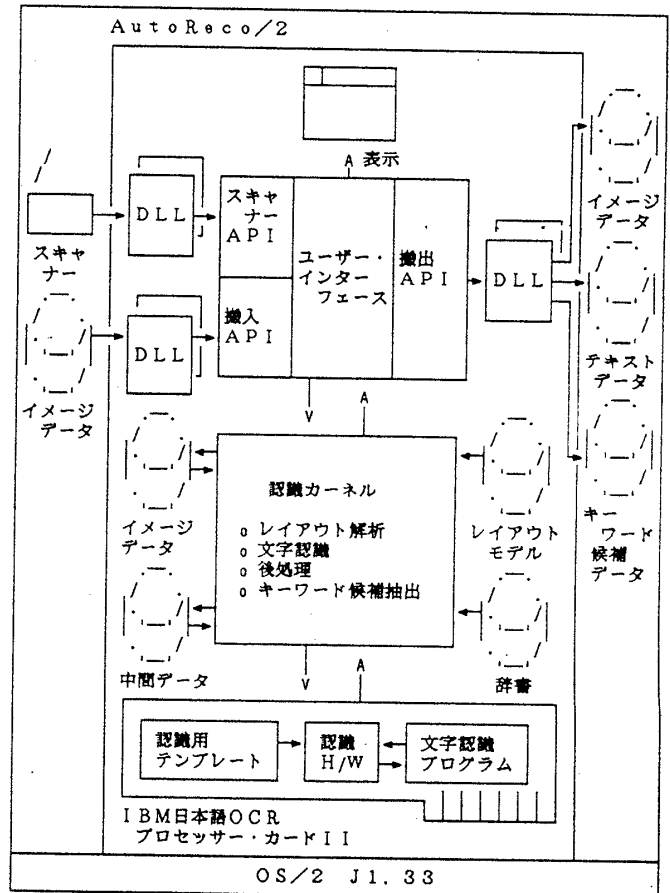


図 1. システム構成

込み、ファイルからの搬入、そして処理結果の搬出については汎用的なインターフェースを設定している。

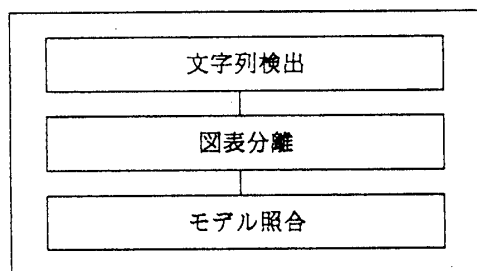
文字認識は漢字 OCR カード上で行われ、認識速度を向上させるだけでなく、メイン・プロセッサが認識後処理およびグラフィックスを使用したユーザー・インターフェースの実現のみに専念できるように、機能分散を行っている。漢字 OCR カードは32-ビットのマイクロプロセッサ、イメージ・データと認識用のテンプレートを格納するためのメモリー、および特徴抽出と識別を行うための専用ハードウェアで構成されている。

Document Recognition System: AutoReco/2 - System Configuration and Overview -

Y.Kato[†], A.Yamashita[‡], Y.Hirayama[‡]

[†]Yamato Laboratory, IBM Japan, Ltd./ [‡]IBM Research, Tokyo Research Laboratory

イメージ・プロセス



テキスト・プロセス V

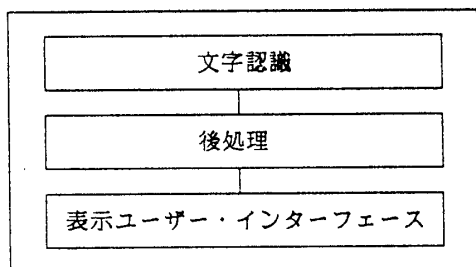


図 2. プロセスの概要

3 プロセスの概要

図2にプロセスの概要を示す。プロセスは大別するとイメージ・プロセスとテキスト・プロセスから成る。

イメージ・プロセスにより、スキャナーあるいはイメージ・ファイルから取り込まれたイメージ・データに対してレイアウト解析が行われる。最初に文字列が検出された後、テキスト部分とそれ以外の図表部分とに分離され、さらにレイアウト・モデルとのマッチングを行い、文字列矩形に対してラベル付けを行う。

テキスト・プロセスは、レイアウト解析によってテキスト領域と判断された部分に対して文字認識を行う。認識処理は漢字 OCR カード上で行われ、カードから得られた認識結果に対して認識後処理を行っている。後処理結果は、認識結果の検証及び修正が容易に行えるようにグラフィックス表示される。

認識されたテキスト・データはイメージ・データおよび日本語後処理の過程で抽出されたキーワード候補データと共に搬出インターフェースを通して搬出される。搬出の際に、レイアウト解析中に文字列矩形に対してラベル付けが行なわれた場合には、電子ファイル・システム等への自動索引付けのために、特定の索引項目に指定されたラベルを持つ矩形内の文字列を割り当てることができる。

4 データ・モデル

処理の多くは並行動作する処理モジュールとして実現されているので協調動作が必要になっている。各モジュールはいずれも一つのオブジェクトとして実現されており、メッセージによって起動やパラメータ設定が行われる⁽¹⁾。認識カーネルはこれらのパーツと呼ばれる処理モジュールと処理結果を格納しているデータ・オブジェクトおよびマップと呼ばれるオブジェクトから構成されている。

マップはデータ・オブジェクトの集合を管理しており、データ・オブジェクトの登録の機能や各種の問い合わせに対して該当するデータ・オブジェクトの集合を返答する機能を持っている。

認識処理中に生成されるデータ・オブジェクトとしては、ページ・イメージ、文字列矩形(行あるいは行の集合でブロックと呼ばれる)、文字ごとの認識結果、および後処理によって抽出されるキーワード候補データ等がある。

これらのデータ・オブジェクトの管理のために生成されるマップとしては、ブロックとブロックの関係を表す木構造、ブロックとそのブロック内に含まれる文字の関係、キーワード候補の管理などに用いられるものがある。

インプリメンテーションはC言語によって行い、メッセージは構造体に貼り付けられた関数の呼び出しとして実現され、並行動作はOS/2のマルチ・スレッド機能を利用している。

5 まとめ

文字認識においては認識率のみが議論の対象になる傾向があるが、文字認識処理を効率的に行うためには、文書の構造解析、認識結果の検証および修正に要する時間の短縮、そして他のアプリケーション・プログラムへの接続も重要である。本システムはこれらの観点を考慮に入れ、生産性の高いシステムを実現することができた。

参考文献

- [1] 天野ほか: マルチメディア文書入力のための文書画像システム: DRS, 情報処理学会マルチメディア通信と分散処理研究会, 48-6, pp.41-48 (1991).