

4Y-8

知的ファイリングモデルシステムの開発(その1)
 - システムの設計思想と実現方法 -

藤縄雅章 藤澤浩道 中野康明 畠山 敦 羽田野英一
 (株)日立製作所 中央研究所

1. はじめに

情報の価値と得られる情報量が増大しており、その情報を効率よく分類・蓄積し、欲しい情報を簡単な対話で、即座に取り出す高度な文書ファイリングシステムが求められている。このため文書ファイリングの本質的な意味を考察して、ファイリングシステムの理念とその技術課題を明らかにした[1, 2]。この理念を基本としファイリングの理想を追究する、実際に試用できる知的ファイリングモデルシステムを構築した。この追求には奥深い人間的課題が多いため、試用を通じて容易に改善して行けるシステム構成方法を開発した。

2. 文書ファイリングの課題

ユーザ側から見た文書ファイリングの課題は、ファイリングする文書画像の品質向上、文書登録の効率向上および文書検索の容易化、高機能化である。

(1) 画像の高品質化：写真などの中間調画像の品質向上および文書のカラー化[3]と共に、長い時間凝視しても疲れにくい紙に近い特性を有する表示方式が求められている。

(2) 登録の効率向上：文書画像から検索に必要な表題や著者名を含む書誌的事項や抄録などの本文内容に関する記述を、人手に頼らず自動的に抽出する技術が求められている。また、これまでに創られた書誌的

事項のデータベースである2次情報と、原文書である1次情報を自動的に対応付けする技術が必要である。

更に、本文を認識することによりフルテキスト・サーチが可能となり、将来的には対象世界の専門知識を自動的に獲得する技術が必要と考える

(3) 検索の容易化、高機能化：従来、紙の世界で行っていたページをぺらぺらめくるなどの探索的な検索機能を、ファイリングシステムの上で実現するマンマシン性の向上、キーワードなど限定された用語だけでなく、日常使用する言葉による自由語検索が必要である。

更に、曖昧な知識からの検索を実現にする意味的内容による検索が必要と考える。

これらの課題を表1に纏める。

3. モデルシステムの設計思想

3. 1 知的ファイリングシステム

文書の登録および検索に対する真のニーズに応えるため、文書理解技術による文書の自動ファイリングと概念ネットワークを用いた知的ファイリングを提案した[1]。図1にそのシステムの概念図を示す。

3. 2 モデルシステムの基本方針

ファイリングの理想イメージを実現するために、本モデルシステムは以下の機能を有する。

表1 文書ファイリングの課題

項目	具体的な内容
画像	●画像の高品質化 中間調画像の品質向上 入力画像のカラー化
登録	●登録作業の効率向上 書誌事項と原文書の対応付け ページ属性(表紙, 図面)の自動識別
検索	●検索作業の容易化, 高機能化 マンマシン性の向上 日常語による検索 意味的内容による検索

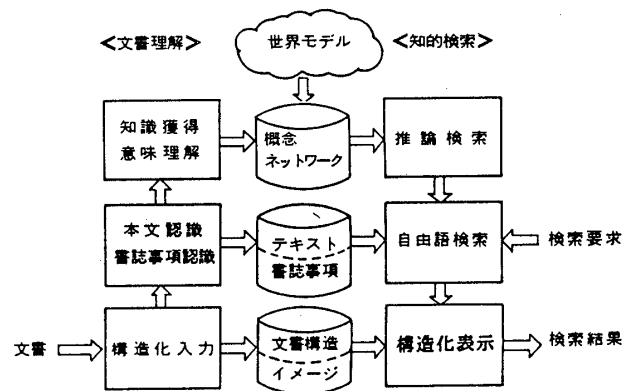


図1 知的ファイリングシステムの概念図

(1) 登録のための文書理解 [5]、概念ネットワークを用いた推論検索 [1]、日常的な言葉による自由語検索 [4] などの高度な知識処理を可能とする。

(2) 検索された書誌的事項および原文書の提示をマンマシン性良く行なう [6] ために、高速なファイル・アクセス機能を持ち、マルチ文書表示や多重な文書表示などの機能的な画面表示ができる。

(3) 実際の業務に試用できる量の原文書イメージ (数万ページ) が蓄積できる。また、試用者の登録・検索に対する要求に柔軟かつ迅速に対応できる。

4. モデルシステムの構成

本システムは文書の登録・検索を実行する大形計算機、文書画像の蓄積・表示を行なう画像ファイルおよびこれ等を連結するためのパソコンから成る (図 2)。

ユーザは大形計算機の端末として機能するパソコンから文書の登録や検索を指示し、パソコンは検索結果である文書番号を受け取って、光ディスクを用いた画像ファイルから原文書を読み出しディスプレイに表示させる。

(1) 大形計算機 (M-680H) : LISP プログラム支援環境の上に構築された知識ベースを作成・編集する概念ネットワーク・エディタ、書誌的事項を自由な言葉で検索する自由語検索プログラム、および登録作業の効率向上を図る文書識別プログラムを核とする。試用によるシステムの改善を容易にすることも考慮して、一部の画像処理を除いて LISP 言語を使用している。

(2) パソコン (B-16/MX) : 原文書の入出力・蓄積・表示などの画像管理を行なう共に、大形計算機による画面の直接操作を可能にするファイル制御機能および TSS 機能を有する。

検索結果の機能的な画像表示、印刷、蓄積文書の読出しなどを行なうため、パソコン上に画像ファイル装

置を制御する仮想的な中央制御装置 (CPU) を設け、画ファイル装置の入出力デバイスを直接制御する仮想計算機制御方式を開発した。

本仮想計算機はインタプリタ形式の制御用言語 (VCL) のもとで稼動する。VCL の命令群はメモリ間の転送命令、演算命令、プログラム制御命令などの基本命令と文書表示、印刷、蓄積文書の読み出しなどの基本動作命令から成る。新しいファイリング機能はこれ等の命令をプログラミングすることにより実現できる。また、命令の拡張も容易である。

(3) 画像ファイル装置 : 画像を入出力制御機能だけを有する。

5. まとめ

新聞記事、英文雑誌記事、特許情報などの文書ファイリングを行ない概念ネットワークを用いた推論検索、自由語検索および高度マンマシン・インターフェース技術の有効性を確認した。また、柔軟なプログラム構造を採用することにより新しいアイデアの確認を迅速に行なうことができた。

今後、新しいコンセプトの試用による検討を進め、使い易いファイリングシステムの実現を目指したい。

6. 謝辞

東京大学和田研究室で開発した Utilisp を利用したことを記し感謝の意を表わします。

- [1] 藤沢他: 情処研報、日本語文書処理 7-4 (1986.7)
- [2] H. Fujisawa, et al., Proceedings of First Int. Conf. Expert Database Systems, 1986, pp31-41
- [3] 藤縄他: 信学全大 1253, 1986
- [4] 島山他: 第 33 回情処全大 4Y-9, 1986
- [5] 中野他: 第 33 回情処全大 4Y-10, 1986
- [6] 藤沢他: 第 33 回情処全大 4Y-11, 1986

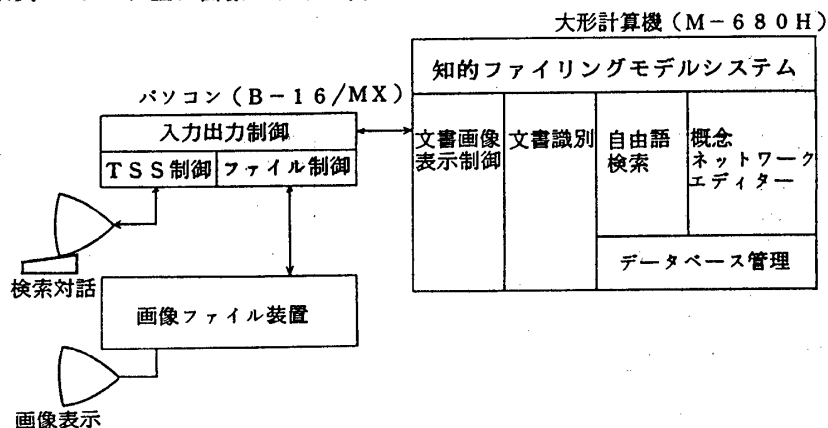


図 2 モデルシステムの構成