

## 知識処理機構を用いた知的文書管理機能

## 6X-4

河合 眞宏、中馬 高彦、櫛木 好明

松下電器産業(株) 情報システム研究所

## 1. まえがき

近年のOA機器の普及はめざましく、その機能も次第に高度化されている。一方、現在のような情報化時代においては、大量の情報の中から、的確な情報を取り出すことがますます重要になってきている。しかしながら、ユーザのこのような要求を十分に満たす高度な検索機能をもったシステムはまだ実用化の域には達していない。

われわれは、ワープロの機能拡張および知的化の観点から、人工知能技術とりわけ知識処理技術および自然言語処理技術を用いた知的文書処理システムの開発を行っている。本稿では、知的文書処理システムにおける文書登録および文書検索機能である知的文書管理機能の文書管理モデル、およびその機能と実現方式について述べる。これは、キーワード付けされた文書の集合として管理されている従来の文書管理システムのひとつの拡張形態として、日本語キーワードシソーラスやキーワード間および文書ファイル間の関連情報などを用いることにより、高度な文書管理機能を実現する。

## 2. 従来の問題点

従来のキーワード検索方式では、

- (1) 文書登録時および文書検索時で時期や作業者の違いにより、検索時に想起するキーワードと登録時に付与されたキーワードが異なることが多い。例えば、キーワードの抽象度の違いや、キーワードの表記の違いなどがそれにあたる。
- (2) 検索したい文書を特定するのに有効な情報(例えば、「ある文書を参照しながら作成した文書」などの関連情報)がわかっている場合でも、それを直接利用できずキーワードそのものがわからなければ検索できない。

などの問題点があるために、検索したい文書を検索することが必ずしもできなかった。

このような問題点を解消するために、本システムは以下の文書管理モデルに基づき、高度な検索機能を提供する。

## 3. 文書管理モデル

図. 1に文書管理モデルの概念図を示す。

分類体系、文書ファイル、および知識情報である日本語シソーラスおよびその探索規則を関連づけている。

分類体系は、共通した性質をもつ文書を複数まとめたもの(分類)の体系であり、通常階層的な構造をもつ。例えば、一般文書全体は、事務文書と技術文書の2つのクラスに細分化され、さらに、事務文書は通達、決裁、手紙などの下位のクラスに細分化される。

文書ファイルは、文書実体データであり、いずれかの文書分類に属する。そのため、文書ファイルは、ある文書分類の性質をもつ一つの実現であると考えられる。このことから、文書分類をクラス、文書ファイルをインスタンスとしたフレーム型知識表現を用いて自然に表現できる。

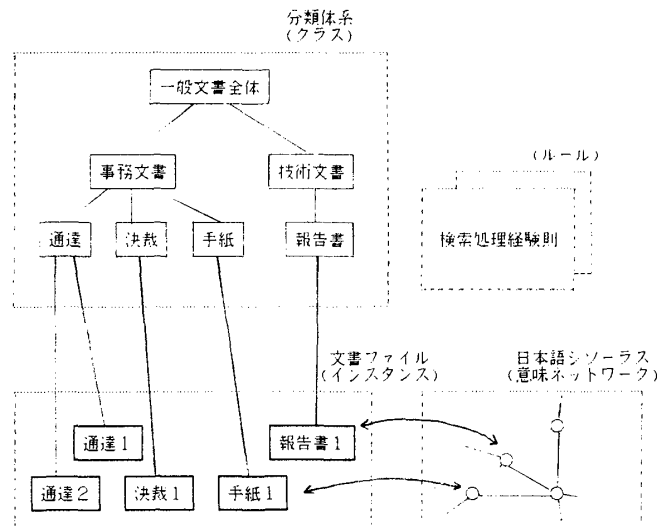


図. 1 文書管理モデル

さらに、日本語シソーラスは、日本語における各概念の階層関係をはじめとする関連情報を保持しており、これはネットワーク構造を表すのに適した意味ネットワーク型知識表現を用いてモデル化する。

また、シソーラスの探索方法などの検索処理の際に用いられる経験則はルール形式を用いてモデル化する。

この文書管理モデルを用いることにより、

- (1) 階層的構造をもつ分類と、その実体である文書ファイルに関連づけて、自然に表現することができる。
- (2) 文書、概念、および検索処理の経験則の追加や修正が容易である。
- (3) 分類-分類間、分類-文書ファイル間の属性値の継承、矛盾性のチェックが可能である。

という特長をもつ。

#### 4. 文書管理機能

本システムでは、以下の文書管理機能を実現する。

##### (1) 文書登録支援機能

###### ●文書付属情報候補提示

キーワードなどの文書付属情報の候補を提示する。

##### (2) 文書検索機能

###### ●拡張検索

設定された検索条件を、日本語シソーラス上での抽象的概念から具体的概念へ拡張して、その拡張結果を用いて文書検索を行う。

###### ●関係検索

日本語シソーラスにおける概念と概念の関係情報（例えば、同義語、類義語、関連語など）、および文書間に存在する関係情報（参照するなど）を利用して文書検索を行う。例えば、「キーワード'調査'と同類語の関係にあるキーワードをもつ文書」を検索したり、「ある文書を参照する文書」を検索する。

###### ●推論検索

日本語シソーラスを、ルール記述に従って幅優先探索の前向き推論により探索し、この推論過程によって生成される概念間関係を用いてキーワードの拡張を行う。また、推論過程の説明機能を用いて、キーワード拡張過程、文書検索理由を説明する。

##### (3) 学習機能

###### ●文書間関係学習

操作者の文書編集動作、履歴、文書内容から文書間の関係情報を自動的に設定する。

###### ●キーワード頻度学習

###### ●文書検索履歴学習

#### 5. システム構成と実現方式

図. 2に、システム構成図を示す。システムは、推論部、知識ベース部、推論制御部、文書管理部、MMI（マンマシンインタフェース）部からなる。知識ベース部には、文書管理モデルを構成するフレーム情報、意味ネットワーク情報、ルール情報を管理する。推論制御部は、推論時の動作環境の設定、推論過程の監視などを行う。推論部は、必要に応じてシソーラスを用いながら検索条件を解釈する。文書管理部は、文書データベースの管理、および、推論部とのインタフェース処理を行う。MMI部は、ユーザからの検索条件入力、検索文書出力、検索過程説明等をマルチウインドウシステムを通して行う。

#### 6. まとめ

知的文書管理機能における文書管理モデル、およびその機能と実現方式について述べた。これは、日本語キーワードシソーラスやキーワード間および文書ファイル間の関連情報などを用いることにより、高度な文書管理機能を実現する。現在、機能評価用プロトタイプシステムを汎用ワークステーション上に試作し、その評価を行っている。今後、高速化、小型化の観点からの改良、ならびに、日本語シソーラスをはじめとする知識データの整備を行う予定である。

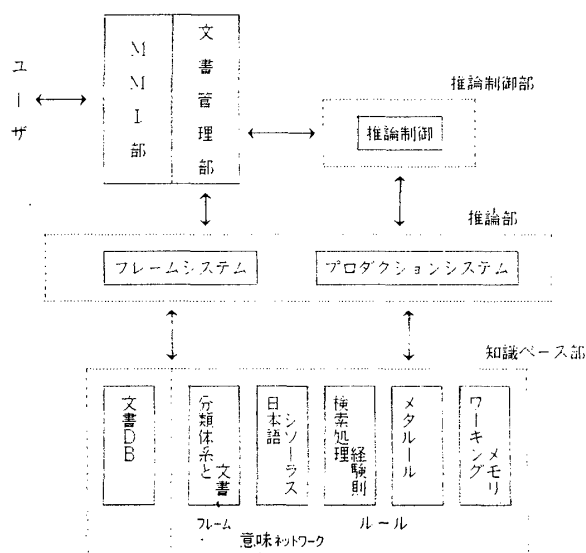


図 2. システム構成図