

複数データベースを対象とする検索支援システムの構成

5H-3

大和田 俊和*, 清木 康**

*筑波大学 情報学類、**筑波大学 電子・情報工学系

1 はじめに

本稿では、オンラインで利用可能な多数のデータベースを対象とした、データベース検索支援システムについて述べる。近年、データベースの普及に伴い、データベースの種類およびデータベース数が増加している。そのため、データベース群全体を把握し、検索要求に応じる適切なデータベースを選択することが困難になっている。我々は、データベース群について十分な知識を有さないユーザが、検索要求に適したデータベースを選択するために必要となるメタ情報および基本機能を示した[1][2]。メタ情報は、各データベースの特徴および実際の内容に関する情報であり、メタ情報に対して基本機能を活用することにより、適切なデータベースが選択される。本稿では、メタ情報の分類のための指針を示し、さらに、メタ情報の生成、および、利用方法について述べる。

2 データベース検索支援システム

データベースの普及に伴い、多数のデータベースをより簡単に利用できるようになってきているが、実際には、それらのデータベース群に関して十分な知識の無いユーザがそれらを有効に活用することは困難である。データベース検索支援システムは、メタ情報およびそれらの操作を行う基本機能群により、適切なデータベースの選択を支援するためのシステムである。

3 システムの全体構成

検索支援システムにおいて実現される機能は、次のようにまとめられる。

- 検索要求の獲得
- データベースの選択
- 検索式の作成
- 問い合わせの発行
- 検索結果の出力

データベースの選択機能を実現するためには、検索対象である各データベースの収録内容を判断する基準となる情報が必要である。本システムでは、あらかじめ各データベースから特定の収録内容に関する情報を抽出しておき、それらをメタ情報として利用できる形で保持する。同様に、検索式を作成する機能を実現するために、あらかじめ各データベースの検索方法、索引語を抽出してメタ情報として保持する。また、本システムでは、これらのメタ情報を管理するためのデータベースとして、メタデータベースを設定する。前述のデータベース選択および検索式の作成の各機能は、メタデータベースを用いて、必要なメタ情報を参照することにより実現される。

問い合わせの発行機能は、実際にオンライン・データベース・システムと通信を行うことにより実現する。

本システムにおける検索支援は、次の機能によって実現される。

- データベース選択機能
- 問い合わせ生成機能
- 問い合わせ発行機能
- メタ・データベース管理機能

本システムの全体構成を図1に示す。

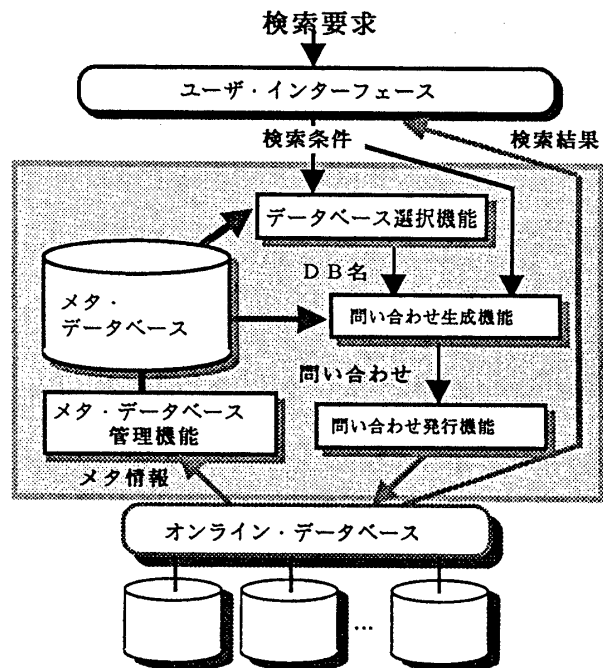


図1 データベース検索支援システムの全体構成

以下では、データベース選択機能に関する部分を述べる。

4 データベースのメタ情報の分類

本システムのデータベース選択機能は、検索対象である各データベースからあらかじめ抽出したメタ情報を用いて実現する。ここでは、そのメタ情報の内容と種類の分類について述べる。

本システムでは、データベース選択のために用いるメタ情報を、データベースの構造との関連に着目し、以下の3種類に分類した。(図2)

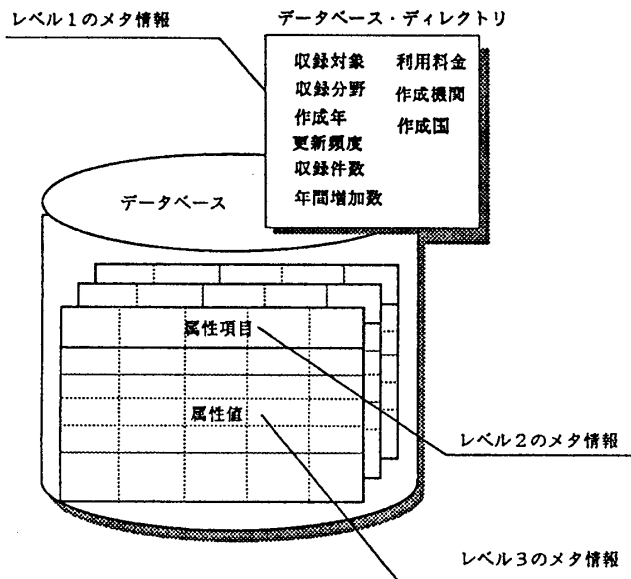


図 2 メタ情報の分類

(1) レベル1のメタ情報

レベル1のメタ情報は、各データベースの外部的な特徴に関する情報より成り立っている。それらのメタ情報は、1データベース内の収録情報に共通した特徴情報である。

(抽出方法) レベル1のメタ情報は全て、データベース・ディレクトリを利用することによって抽出される。

・レベル1のメタ情報

データベース自体が持つ性質に関する情報であり、次のものがある。

- データベースの主題内容
- データベースの収録情報の分野
- データベースの作成・管理に関する情報
- データベースの収録件数に対する統計データ
- データベース利用に関する情報

(2) レベル2のメタ情報

レベル2のメタ情報は、各データベースの属性項目の名前を表す情報である。これは、属性の内容が同じでも、データベースによって異なる属性名を用いている場合がある。したがって、レベル2のメタ情報としては、属性項目の内容に対応した、各データベース間で共通の属性名(グローバルな属性名)を用いるものとする。

(抽出方法) 属性名は、データベース・ディレクトリを利用することによって抽出される。ただし、属性項目の内容に対応して、抽出した属性名をデータベース間で共通の属性名(グローバルな属性名)に変換する必要がある。

・レベル2のメタ情報

データベースの属性に関する情報である。

- データベースの収録情報が持つ属性項目の内容

(3) レベル3のメタ情報

レベル3のメタ情報は、各データベースの収録情報に含まれる値(文字列、数値)を表す情報である。これは、特定のデータがデータベースに収録されているかどうかを判断するための情報である。

(抽出方法) データベースの属性項目のうち、検索時に指定される頻度が高いものは、ほとんどといってよいほど索引化されており、データベース毎に転置ファイルをそのままレベル3のメタ情報として利用できる。また、シソーラスが用意されている場合、それを用いてデータベース間での用語の統一化を行うようにする。

・レベル3のメタ情報

データベースの属性値に関する情報である。

- データベースの収録情報に含まれる属性値の内容
- データベースの収録情報に含まれる属性値の統計データ

5 メタ・データベースの実現

メタ・データベースでは、データベース間で異なって表現されている同義のキーワード(異名同義)群、あるいは、データベース間で同じに表現されているが異なる意味を持つキーワード(同名異義)群について、それらを本来の意味にしたがって、正しく扱わなければならない。

レベル1のメタ情報においては、異名同義および同名異義といった問題は起こりにくい。レベル1のメタ情報は、各データベースの外部的な特徴を表す情報であり、このメタ情報の作成時において、データベース間の表現を統一化することは容易である。

一方、レベル2、3のメタ情報は、すでに作成されたデータベースの内部情報から抽出されるので、データベース間では、異名同義および同名異義の属性名および属性値群が存在する。

そこで、レベル2およびレベル3において、それらの属性名群および属性値群に対応するメタ情報を正しく生成するための方法が必要となる。我々は、情報(キーワード)間の意味の近さを計算によって求め、異名同義、同名異義の関係を動的に解釈する方法の検討を、現在行なっている。

6 おわりに

オンラインで利用可能な多数のデータベースに対し、十分な知識を有さないユーザでも効率的な情報検索が可能で、検索支援システムの提案を行なった。そのなかのデータベース選択を行なう機能を実現し、さらに、データベース選択を行なうために必要となるメタ情報を分類するための指針にもとづき、メタ情報の生成および利用法について示した。今後は、メタ・データベースの実現、および検索支援システムの他の機能の実現を行なっていく予定である。

参考文献

- [1] H.Shimizu, Y.Kiyoki, A.Sekijima, N.Kamibayashi: "A Decision Making Support System for Selecting Appropriate Online Databases", Proc.1st Int.workshop on Interoperability in Multidatabase Systems, pp.322-329, Apr.1991.
- [2] 関島 章文、清水 康、清水 宏行: "複数のデータベースを対象とする情報検索支援システム", 情報処理学会第42回全国大会講演論文集(4), pp.97-98, 1991.