

1 L-4 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

関島 章文[†] 清木 康[†] 清水 宏行[‡]

[†]筑波大学、[‡]富士ゼロックス(株)

1 はじめに

データベースの普及に伴い、様々な情報がデータベース化され、オンラインなどで簡単に利用できるようになってきた。これらのデータベースはその種類、情報量ともに年々増加の一途をたどっている。

このようにデータベースの種類が多くなると、一般のユーザにとっては、自分の必要とする情報がどのデータベースに収録されているかを判断することが難しくなってくる。また各データベースの検索方法が統一されていない場合は、それぞれについて検索のための知識を習得しなければならない。

これらを解決するために、一般のユーザの代わりにオンライン情報検索を行ってくれる、データベース・サーチャーという職業がすでに確立されている。しかし今後、あらゆる方面でデータベースの利用頻度が増すにつれて、情報検索をデータベース・サーチャーに頼るだけでは、対処しきれなくなってくる事が予想される。

そこで我々は、多数のデータベースに対して十分な検索知識を有していないユーザでも、効率的な情報検索を可能にする情報検索支援システムの開発を行なっている。

本稿では、この情報検索支援システムの概要と、検索要求に適したデータベース選択の実現方法について述べる。

2 情報検索支援システムの概要

データベース・サーチャーによる検索作業は、利用可能なデータベースの中から検索に必要なものを選択し、実際にデータベース検索を行なって、依頼者の要求を満たす情報を入手するというものである。

これは、以下のような手順に分けることができる。[1]

- 1) 依頼者から検索に必要な情報を獲得
- 2) 検索に適したデータベースの選択
- 3) データベース検索
- 4) 検索結果の評価・加工

我々の情報検索支援システムでは、このうち2)のデータベース選択と、データベース・システムへの問い合わせの発行を、主要な機能として実現する。

検索に適したデータベースの選択を行なうためには、各データベースの収録内容や、内部構造についての情報が必要である。これはすでに、一部のオンライン・データベースの間で、データベース・ディレクトリおよびクロスファイル・インデックスという形で提供されている。データベース・ディレクトリの内容は、各データベースの収録情報の内容や、属性の一覧表といった情報である。またクロスファイル・インデックスの内容は、特定のキーワードを含むレコードの、各データベースにおける収録件数といった情報である。このようなデータベースを選択

する基準となる情報を、以下ではデータベースの「メタ情報」と呼ぶ。

我々は本システムの構成要素として、検索対象となる全てのデータベースのメタ情報を管理する、「メタ・データベース」を設定した。そして、実際に情報検索の支援を行なうモジュールとして、データベース選択部、問い合わせ生成部、問い合わせ発行部、メタ・データベース管理部をそれぞれ設定した。以下に各モジュールの機能を示す。

データベース選択部

検索要求を受けとり、メタ・データベースを用いて、利用可能なデータベース群から、その検索に適したデータベースを選択する。

問い合わせ生成部

検索するデータベースと検索条件を受けとり、そのデータベースの問い合わせ形式に従って、問い合わせ文を生成する。データベースの問い合わせ形式は、メタ・データベースに収録されている。

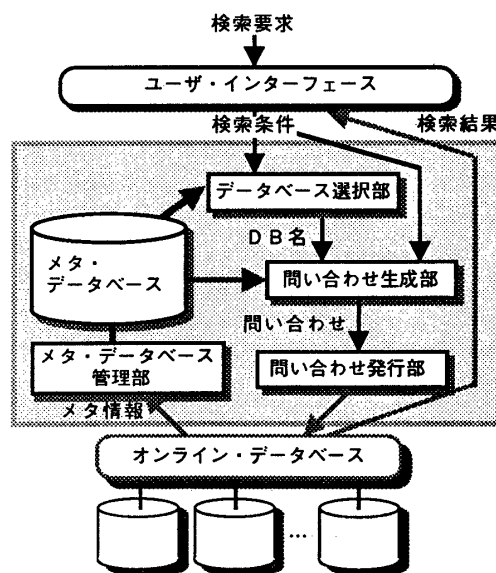
問い合わせ発行部

データベースにアクセスして問い合わせを発行し、得られた検索結果を返す。

メタ・データベース管理部

利用可能なデータベースの変更にしたがって、メタ・データベースの内容を更新する。

また本システムの全体構成を(図1)に示す。



(図1) 情報検索支援システムの全体構成

An Information Retrieval Assistant System for Multiple Databases

Akifumi Sekijima, Yasushi Kiyoki, Hiroyuki Shimizu

University of Tsukuba, Fuji Xerox Co.,Ltd.

実際の情報検索において、本システムによる検索の支援は、以下の順に行なわれる。

- 1) 検索に適したデータベースの選択 (データベース選択部)
- 2) 検索条件からの問い合わせ文の生成 (問い合わせ生成部)
- 3) データベースへの問い合わせ発行 (問い合わせ発行部)

各モジュールへの入力、ユーザ・インターフェースを通して、対話的に与える。また各モジュールの実行結果は、ユーザ・インターフェースに出力される。

3 データベース選択機能の実現

本システムの主要機能のうち、検索要求に適したデータベースを選択する機能を実現した。ここではその実現方法について述べる。

3.1 データベースのメタ情報の設定

データベースの選択において最も重要な要素は、メタ情報の内容である。これは、データベースの収録対象、収録件数、データベース・スキーマなど、様々な種類のものが考えられる。そこで我々は、データベースの構造に着目して、メタ情報を(表1)のように3種類に分類した。

level1のメタ情報 (プロフィール情報)	データベースの収録情報、レコード件数、更新頻度、作成年度、利用料金、作成機関などの情報
level2のメタ情報 (スキーマ情報)	データベースのレコードが持つ属性についての情報
level3のメタ情報 (インデックス情報)	データベースのレコード内の属性値とその件数についての情報

(表1) データベースのメタ情報

ここでlevel2のメタ情報については、前述のデータベース・ディレクトリのように属性の一覧を示すだけでなく、その属性の意味表現をデータとして持つ。これはスキーマが異なるデータベースの間で、属性間の等価性を判断するのに用いる。具体的な内容としては、各データベースのスキーマを意味モデルを用いて抽象化し、その意味モデル表現と各属性の対応を記述したものである。[2]

また、level3のメタ情報は、利用可能なデータベースの数によってはかなり大量のデータになるので、検索条件となる頻度が高い属性のみに限定して用意する。

3.2 メタ・データベースの設定

上記のように分類したメタ情報について、各レベルごとにメタ・データベースを作成する。これは各レベルによって、メタ情報のデータ構造およびデータ量が著しく異なるので、それぞれ別のデータベースとして実現する。例えば、level1、level2のメタ・データベースはオンメモリのデータベース・システム上に作成し、level3のメタ・データベースは分散データベース・システム上に作成する、といった実現方法などが考えられる。

3.3 データベース選択基本機能の設定

次に、各レベルのメタ情報に対応する、データベース選択基本機能を設定する。これは、与えられた検索条件からメタ情報の各レベルに対応するキーワードを受けとり、メタ・データベースを検索することによって、適合するデータベース名を値として返す機能である。これらは論理和・論理積で結合することにより、検索条件の組合せに対してもデータベースの選択を行なうことが可能である。また選択されたデータベースが複数の

場合は、選択に用いたメタ情報の内容により優先順位が決定される。

我々は、各レベルのメタ情報に対して、(表2)のようにデータベース選択基本機能を設定した。

level1のデータベース選択基本機能	
1-1	SelectDB_by_Object (object_name, [Usable_DB_List]) 収録情報による選択
1-2	SelectDB_by_Year (c_op(year), func, [Usable_DB_List]) 作成年による選択
1-3	SelectDB_by_UpdateFreq (c_op(freq), func, [Usable_DB_List]) 更新頻度による選択
1-4	SelectDB_by_RecNum (c_op(number), func, [Usable_DB_List]) 収録レコード数による選択
1-5	SelectDB_by_Charge (c_op(charge), func, [Usable_DB_List]) 検索コスト (接続料金、出力料金) による選択
1-6	SelectDB_by_IncNum (c_op(number), func, [Usable_DB_List]) レコードの年間増加数による選択
1-7	SelectDB_by_Organization (organization, [Usable_DB_List]) 作成機関・作成国による選択
level2のデータベース選択基本機能	
2-1	Select_by_AttName (Att_name, [Usable_DB_List]) 属性名による選択
2-2	SelectDB_by_SDM([SDM-List], [Usable_DB_List]) 属性の意味表現による選択
level3のデータベース選択基本機能	
3-1	SelectDB_by_AttValue (Att_name, value, [Usable_DB_List]) 属性値による選択
3-2	SelectDB_by_AttStat (Att_name,value,c_op(stat),func,[Usable_DB_List]) 属性値の件数による選択

c_op: =,<,>,<=,>=,... func: max,min,average...

(表2) データベース選択基本機能

これらのデータベース選択基本機能は、具体的に以下のように用いられる。

例:

```
SelectDB_by_Object (news, [DB1, DB2, ...]) and
SelectDB_by_Year (>=(1965), [DB1, DB2, ...])
...収録情報が'news'で、1965年以前に作成されたデータベースを選択する。
```

本システムでは、データベース選択部が、与えられた検索条件に対応して、上記のように基本機能を用いることによって、検索に適したデータベースの選択を実現する。

4 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、十分な知識を有していないユーザでも効率的な情報検索が可能な、情報検索支援システムの提案を行なった。また利用可能なデータベース群の中から、与えられた検索条件に適したデータベースの選択を行なう機能を実現した。今後は与えられた検索条件からの問い合わせ文の自動生成および、データベースに対する問い合わせ発行の機能について、それぞれ実現する予定である。

参考文献

- [1] 三輪 眞木子: "サーチャーの時代-高度データベース検索-", 丸善株式会社 (1986)
- [2] R.Hull,R.King: "Semantic Database Modeling: Survey, Applications,and Research Issues",ACM Computing Surveys, Vol.19,No.3 (1986)