

D-4-10 メタオブジェクトを用いた柔軟な検索エンジン

井上, 良太郎 / 滝沢, 晃崇 / 三浦, 孝夫

(出版者 / Publisher)

一般社団法人電子情報通信学会

(雑誌名 / Journal or Publication Title)

電子情報通信学会総合大会講演論文集

(号 / Number)

1

(開始ページ / Start Page)

71

(終了ページ / End Page)

71

(発行年 / Year)

1999-03-08

D-4-10 メタオブジェクトを用いた柔軟な検索エンジン

Flexible SearchEngine Using MetaObject- An Experimental HOME Application

井上良太郎 滝沢晃崇 三浦孝夫
法政大学工学部電気電子工学科

1 前書き

Web上の情報ベース(インターネット情報源)の保持するスキーマ表現や情報の形式に対して柔軟に対応できる検索エンジンの実現と画像や音声等の非テキスト情報を検索に用いたユーザインタフェースの実現について述べる。

本検索エンジンでは、各情報ベースのスキーマ表現や保持する情報の形式を、必要に応じて取得し評価するため、Web上の情報ベースのようにスキーマ表現や保持する情報が不確定な環境に対しても十分に対応する。これに対し、従来の検索エンジンは、あらかじめスキーマ表現や情報の形式などが細部まで明確に決められているため有効に各情報ベースを利用する事ができない。また、従来の検索方法の様に、画像データ等の非テキスト情報に二次的なテキスト情報(内容を間接的に表すような表現)を付加して検索に利用する場合、付加された二次的なテキスト情報を正確に検索条件として与えなければならない。これに対し本検索エンジンでは、画像などの非テキスト情報を不完全な質問、例えば「画像」「音声」といった大まかな質問でも検索する事が可能である。また、非テキスト情報を直接検索に用いるため、利用者ごとにばらつきが生じにくく、検索条件として与えられた非テキスト情報と情報ベース中の非テキスト情報との一致を的確に検査できる。

2 検索エンジンの概念

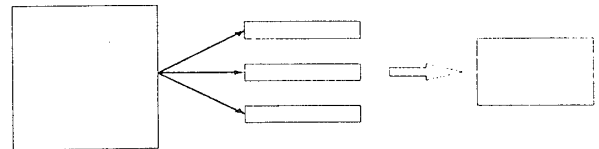
本検索エンジンの実現のため、幾つかの考慮すべき問題がある。まず、画像情報の表現、例えば同一人物の正面顔と横顔の区別は、画像の内容判断に関わるためここでは扱わない事とする。また、情報の柔軟な表現のために共通(核)情報を最小限にすべきである。これについては、スキーマの保持方法、名称(NAME、名前、氏名)、データの表現形式(数字、文字、画像)、データ処理機能等に柔軟性を与える。本論文では、HOMEを用いて検索に重点を置き本検索エンジンについて解説する。

このような検索エンジンを実現するために、各情報ベースの保持するリレーション等のレイアウト情報を、事前に細部まで認識する事無く、いかにして取得し検索に用いるのが問題となる。また、どのようにして非テキスト情報の詳細なデータ部分の解析をする事なく各情報ベースが保持する非テキスト情報の形式を検索条件ごとに認識するのが問題となる。

本検索エンジンの持つ機能の特徴としては、まず、検索対象としているデータの型を認識し、そのデータの型を持ちうる可能性のある情報ベースとリレーションの特定をメタオブジェクトを用いた検索により可能としている点にある。また、検索に用いられた質問が不完全なものであったり、検索条件が完全に情報ベース中のある特定のオブジェクトと一致するような条件でなくとも、メタオブジェクト部分での一致を検査する事により、より柔軟な範囲の結果を得る。即ち、不完全な質問を繰り返しながら的確な質問を確定させる。更に、本検索エンジンでは、非テキスト情報そのものを直接検索対象としたり検索条件に用いるために、情報ベース中の非テキスト情報と検索条件として与えられた非テキスト情報との一致を検査する機能とデータベース中に格納されている非テキスト情報をWWWに表示する機能を持つ。

本検索エンジンの処理の流れの具体例としては、WWWブラウザから、利用者が非テキスト情報を検索したい場合、{画像、音声、...}の中から選択し明示的に示す必要がある。例えば「画像」が選ばれたとする。更に「住所=東京 年齢=22」といったテキスト形式の入力により検索を開始。次に入力情報から検索対象が画像データである事を認識する。また、画像データを保持する可能性のある情報ベースとその情報ベース中のリレーションを特定する。それから、検索条件を満たす画像データを検索する。検索結果として得られた画像データ、例えば属性「趣味」の個々の値、読

書、スキー、テニス、温泉...等をそれぞれ表現したアイコンを、WWWブラウザに画像形式で表示する。更にそれらのアイコンをもとに検索を行い、その画像に付随する情報を得る事ができる。この時、テキスト入力の検索条件も付随する事ができる。この検索結果もWWWブラウザに示される。



メタオブジェクト ⇒ リレーション ⇒ オブジェクト

3 システムの実現方法

HOME環境下の情報ベースは、保持するリレーション、属性、属性の領域等のメタオブジェクトの情報を記述したカタログ:RelationCatalog,AttributeCatalog,DomainCatalogを持つ。これらのカタログを利用して検索を行う。検索対象としているデータの型と情報ベース中の対応する領域との対応は、対象となるリレーションの特定を行う時に用いる。例えば、画像データである属性に用いられている領域名との対応表(仮にImageRel)を定義。新たに画像データに対応する領域名が追加され得た場合以下の表に追加すれば良い。

image	photo	picture	...
-------	-------	---------	-----

各情報ベースのAttributecatalogにおいて、属性の領域名を検索対象としているデータに対応する領域名と一致する行を選択し、リレーション名を記述した列を射影する。この結果、予め決められたリレーションのみを要するのではなく、関連情報をその都度自動的に入力することになり、システムの柔軟性を得る。以下にこのデータ操作をデータベース中に保存し、これをQ1とおく:

$$Q1 = \Pi_{RelName} \sigma_{DomName = \dots} (AttributeCatalog)$$

具体化機能と検索条件を満たすデータを得るためのデータ操作をQ1に行い、検索結果を得る。

$$:\{ \Pi_{Attribute Name} (\$Q1) \} \bowtie \sigma_{Attrn = value n \dots \sigma_{Attrl = value l} (\$Q1)$$

情報ベース中の非テキスト情報を比較したり表示するために、データ種別情報とデータ部分の組であるタグ付きデータを取り出す。逆に、非テキスト情報をタグ付きデータとして情報ベースに格納する機能も有する。取り出したタグ付きデータの情報記述部分を解析し、その記述に基づきファイルを生成しブラウザに表示する。次に検索条件として与えられた非テキスト情報と比較するためには、データ部分の比較を行う。ここでフォーマット形式が異なる場合、動的に形式を変換し一致を検査する。

4 むすび

Web上の情報ベース(インターネット情報源)の保持するスキーマ表現や情報の形式に対して柔軟に対応できる検索エンジンと画像や音声等の非テキスト情報を検索に用いたユーザインタフェースを実現した。

5 参考文献

- [1] 北川 博之著「データベースシステム」、昭晃堂、(1996)
- [2] Miura, T., Matsumoto, W. et al.: Harmonizing Objects and MetaObjects for Data Warehousing, to appear