

仮説概念を用いた広域情報検索

3P-5

平田好祐
千葉工業大学宮崎収兄
情報工学科

1はじめに

インターネットが個人で利用できるようになった現在、世界各国への情報検索の需要は急速に高まりつつある。現在はWWWなどによる方法が主流であるが、複数のデータベースに対する統一的な情報検索が可能になればネットワーク利用の可能性は飛躍的に高まると考えられる。このような環境では問い合わせたい内容によって対象データベースが変化するため、あらかじめスキーマの統合を行い登録することは困難である。我々は散在するデータベースの全体を仮想的な1つのデータベースとし、各データベースはその一部のデータを格納した不完全データベースと考えることにより広域データベース検索を実現する方法を提案した[宮崎96]。本稿では問い合わせにQBEを用いることによりユーザインターフェイスを改良する方法を提案する。

2 不完全データベースを用いた広域情報検索

インターネットのように不特定多数のデータベースが存在する環境で、サイトの自律性を確保し広域情報検索を実現するためにはマルチデータベースなどで研究されてきたような技術が適用できる。

例えば、構文の違い、名前の違いなどはある程度ディクショナリ、シソーラスなどで吸収可能である。しかし、属性の有無などスキーマ間に変換しきれない不一致がある場合もある。

このため我々は不完全データベース概念によりデータベースを拡張することにより、差異が存在するまで広域検索を行う方法を提案した。この方式では問い合わせはそのまま対象に送られデータベースは答えられるものだけ答える方法をとる。この方法では問い合わせ中の属性がデータベースに存在しない場合、部分解や条件付解（仮説付解）が得られる。複数のサイトから得られた部分解や条件付解を自然結合を拡張した方法によってマージし完全解を得ることも可能である。

3 QBEによる広域情報検索

[宮崎96]で提案した方法ではユーザは仮想的なスキーマを想定し、それに基づいて問い合わせを記述する。本稿ではスキーマ情報をQBE上で表示することにより、より使い易くする方法を検討する。

Global Information Retrieval using Concept of Hypothesis

Yoshihiro Hirata, Nobuyoshi Miyazaki
Department of Computer Science, Chiba Institute of Technology

2-17-1 Tudanuma Narashino Chiba 275 Japan

3.1 QBE形式の長所

- 表の骨格が表示されるため、ユーザは関係リストやスキーマの属性名などを記憶せずにすむ。（広域情報検索では特に意識せずに検索出来ることが重要である）
- 2次元のインターフェイスを使用するので、コマンドなどの入力順序は無関係になり、ユーザは質問の全体像を把握することが容易である。
- ユーザは直感的に検索可能である。

3.2 QBEによる広域情報検索の方法

仮想的スキーマによって問い合わせを入力する方法では曖昧性のため、記述や処理が困難な場合もある。このため、ユーザーが問い合わせを記述し易くするために以下のように処理を行う。

ユーザの知りたい情報（属性名まで）を入力



情報のありそうなサイトに問い合わせ、関連するスキーマ情報を得る。



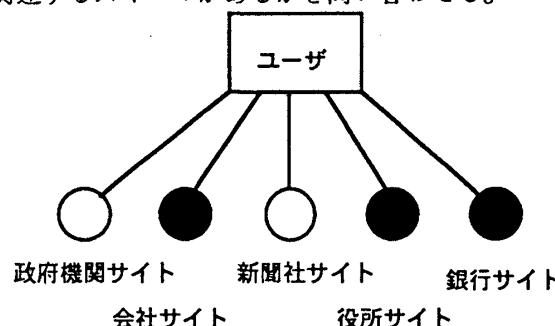
スキーマ情報を統合しQBE上で表示する。
(より良い属性名がわかれればそれを表示する。
属性名が曖昧なため処理が困難な場合はその旨を表示し再入力する。)



ユーザが問い合わせを行う

例 ユーザは 名前=千葉太郎という人物の住所
役職、銀行の預金を知りたいとする。

問い合わせを行う前に情報のありそうなサイトに、
関連するスキーマがあるかを問い合わせる。



以下のようなスキーマを持つサイトから情報を得る
ことができたとする。

銀行サイトのスキーマ：
預金者 {名前、口座、預金}

役所サイトのスキーマ：
住人 {姓名、住所、電話}

会社サイトのスキーマ：
社員 {氏名、役職、年齢}

- ・システムはスキーマ情報を統合しQBE上で表示する。
- ・ユーザがQBEにより問い合わせを入力する。
- ・3つのサイトへ問い合わせを送り、各サイトから送られた解を表示する。

人物	氏名	役職	住所	預金
	千葉太郎			



人物	氏名	役職	住所	預金
	千葉太郎	P.X	P.Y	P.Z



会社のサイト 市役所のサイト 銀行のサイト



人物	氏名	役職	住所	預金
	千葉太郎	課長		
	千葉太郎		千葉県	
	千葉太郎			300,000

各サイトからの情報は部分解なので得られた情報だけが表示されている。

問題点

- ・1つの対象に関係する情報が別々に表示され統一性がない。

3・3 複数解の統一

1つの対象に対する情報が複数のサイトから得られた時は拡張自然結合によってマージすることがで

きる。

例 マージ処理による解の統一
上記の例から

人物	氏名	役職	住所	預金
	千葉太郎	課長		
	千葉太郎		千葉県	
	千葉太郎			300,000

マージ



人物	名前	役職	住所	預金
	千葉太郎	課長	千葉県	300,000

解をマージすることでこのような完全解を得ることができる。

問題点

- ・同一対象のものでない情報がマージされることがある。このため、解の統一や部分解、複数解表示はユーザの指定によって変更可能とする。
- ・この問題の本質的解法には広域IDの検討が必要。

6 まとめ

インターネットのような広域情報検索では、異なるデータモデル同士の問い合わせが発生し、スキーマなどの問題によって実現が困難である。また、検索を行えたとしても多量の解の生成による対処の問題がある。ここでは関係データベースのインターフェイスであるQBEによる広域情報検索を示した。

技術的な部分以外にも、インターネットには広域情報検索を実現させるための問題以外にも誤った情報や倫理上問題のある情報も多数存在するなどの問題がある。実現方法と共にインターネット上の法整備なども合わせて検討していく必要もあるだろう。

参考文献

[宮崎96]宮崎、不完全データベースと広域データベース検索 情報処理学会 DBS研究会、1996年1月