

データベース検索を行うユーザとコンサルタントの対話解析

2M-7

水谷 研治 平井 誠 松島 宏司
松下電器産業株式会社 中央研究所

1 はじめに

近年のCPUの性能の向上、光磁気ディスクなどの2次記憶媒体の大容量低価格化はデータベース・システムをハード的に大きく成長させた。しかし、ソフト的な側面、すなわち使い勝手においては旧態依然である。多くの商用データベース・システムは複雑なキーワード論理式やメニュー、SQLによる検索方式を採用している。ユーザがシステムに慣れるまでの労力は大きく、システムを使いこなすことができるようになって、自分が意図するデータの集合を検索することは容易ではない。

我々はこのような背景から簡単に効率の良い検索を行える自然言語対話に基づくデータベース検索技術の確立を目指している。本論文ではデータベース検索を行うユーザとコンサルタントの対話の解析結果とコンサルタントのモデル化について述べる。

2 ユーザとコンサルタントの対話

コンサルタントが持つノウハウを実際に検索を行ったコーパスから抽出すべきであるが、技術文献や特許などのデータベース検索を行っているときのコーパスを収集することは困難である。企業の動向などが外部に漏洩する危険性があるので不可能に近い。また、ユーザとコンサルタントの拘束時間、商用データベースのアクセス料金など、現実世界においては常にコストの問題が付随するので、純粋にデータの検索を目的とするコーパスは存在しない。

いくつかのコーパスの観察およびコンサルタントの業務の実態 [1][2] から、コンサルタントとユーザの対話は、制約構成段階、スクリーニング段階、検索結果評価段階の3つの段階から構成されると考えられる。

制約生成段階ではコンサルタントはユーザの要求を分析し、データベースの入力として制約能力の高い制約表現を構成することを目的とする。制約が不十分であると判断される場合はその理由を説明し、ユーザの知識を要求する。ユーザが自分の概念の説明に適切な言葉が見つけれないときはコンサルタントの知識による助言を行う。

制約生成段階で制約が十分であると判断された時点で実際に検索を行い、スクリーニング段階へ移る。この段階ではコンサルタントは検索されたデータの内容に目を通し、ユーザの意図に適合しているか否かを判断する。コンサルタントの知識だけでは判断不可能な場合は、ユー

a dialog analysis of a user and a consultant retrieving information
Kenji MIZUTANI, Makoto HIRAI, Hiroshi MATSUSHIMA
Central Research Laboratory, Matsushita Electric Industrial Co., Ltd

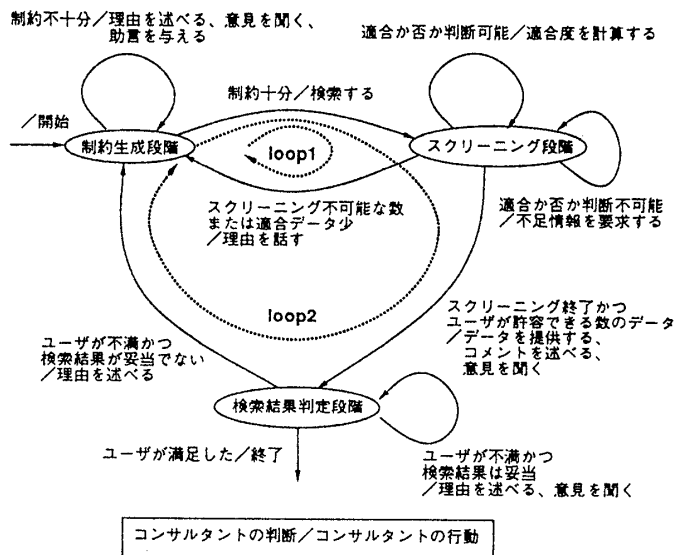


図1: コンサルタントとユーザの対話の遷移

ザに不足情報を要求する。検索件数がスクリーニング不可能な数のとき、あるいは適合すると判断されるデータがユーザが処理できる範囲を越えていたり、少ないときは制約構成段階へ戻る。

スクリーニング段階である程度の数の適合データが得られたときは、検索結果をユーザに提供し、検索結果についてコンサルタントの知識に基づく評価を述べ、ユーザに意見を求める。コンサルタントが検索結果は妥当であると思うがユーザが不満を表明する場合はどちらかが納得するまで話し合う。ユーザが満足すれば検索を終了するが、ユーザが満足せずにコンサルタントも検索結果が妥当でないと判断した場合は制約段階に戻る。

図1に以上の過程を示す。loop1はコンサルタント自身がユーザの意図を満足するデータを獲得していないと判断するために生じる繰り返しである。loop2はコンサルタントがユーザの意図を満足と思っていたデータがユーザを満足させなかったために生じるものである。実際の対話の多くは制約生成段階で終了する傾向にある。制約不十分と判断した場合にはコンサルタント自身で制約の修正を行ってloop1を繰り返すことが多い。ユーザが検索結果に満足しなくても結果を受理することが多く、loop2が繰り返されることは少ない。

3 コンサルタントのモデル化

3.1 コンサルタントのタスク

図1の対話の遷移から、コンサルタントの主な知的作業は、

1. ユーザの要求を適切な制約表現に変換する
2. 得られたデータをスクリーニングする
3. 期待したデータが得られないときに制約表現を修正する

の3つに集約される。この中には2段階の検索が存在する。第1段階は制約生成段階からスクリーニング段階へ移るときの、データベースが入力とする制約表現による表層的な検索であり、第2段階はスクリーニング段階でのコンサルタントによる深層的な検索である。

3.2 自然言語対話によるデータベース検索手法

テキスト・データベースを対象とした検索システムの多くは制約表現としてキーワード論理式を使用し、表層的な検索を行うだけである。スクリーニング段階で排除されるデータが入ること、適合すべきデータが排除されることは避けられない。また、期待したデータが得られなかったときのキーワード論理式の修正はシソーラスを用いた試行錯誤的な方法で行われ、コンサルタントのスクリーニングに頼るところが大きい。このような能力をシステムに持たせるためにはデータのインデックスをキーワードよりも表現能力の高いものにしてシステムに深層的な検索をも行わせる必要がある。

ここでは依存構造を基本的な表現形態としてユーザと対話を進めながらその意図するデータを検索するシステムの枠組を示す。なお、検索対象として特許の抄録データベースを仮定する。

まず、個々のテキスト・データの内容を端的に表す文をあらかじめ解析し、依存構造表現に変換してそのデータのインデックスとする。図2に例を示す[3]。

システムのタスクは以下ようになる。

1. データに対する制約を発話の命題内容の依存構造で表現し、表層的な検索を DAG(Directed Acyclic Graph) 構造のパターン照合で行う。対話によって制約が生成される過程を図3に示す。システムの情報要求の発話 S_1 に対するユーザの情報提示の発話 U_2 からデータに対する制約 C_1 が生成されるが、表層的な検索を行って制約不十分と判断される。システムの知識によりユーザの検索意図を推論して情報確認の発話 S_3 を行い、制約 C_2 を生成する。
2. 深層的な検索はシソーラスとユーザの知識を利用した DAG 間の距離計算に基づくデータの絞り込みとする。
3. 制約を修正するためのユーザに対する質問、助言は複数の DAG 間の構造上の差異から決定する。

[特許抄録の目的項]

単精度の固定小数点の演算回路にプライオリティエンコーダ回路、マルチプレクサ等の回路を付加することにより、効率よく高速にデータ駆動方式パイプライン上で行えるようにする。

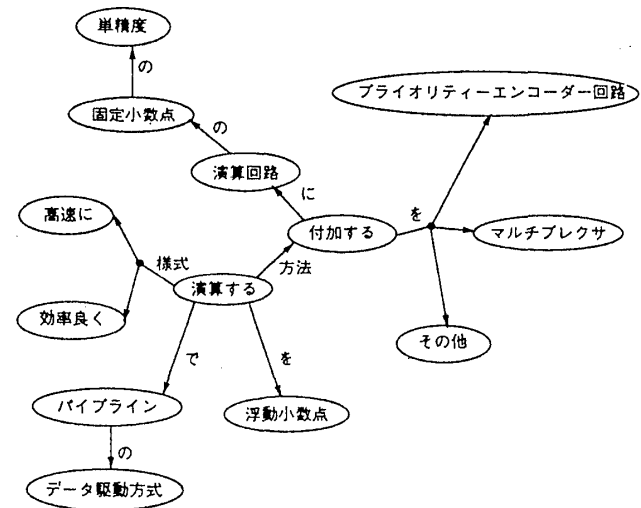


図2: データのインデックス

- S_1 どのような特許をお探しですか。
 U_2 パイプラインの高速化について書かれたものをお願いします。
 S_3 パイプラインで高速に演算するというのでしょうか。
 U_4 そうです。

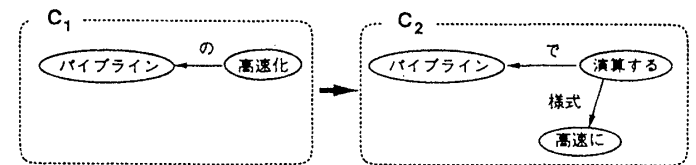


図3: 対話における制約の生成過程

4 おわりに

コンサルタントとユーザの対話から得られたコンサルタントのモデルを示し、自然言語対話によるデータベース検索方法の枠組を示した。今後はテキスト・データベースを対象として手法の具体化を進める予定である。

参考文献

- [1] 情報科学技術協会 編. 新サーチャー入門 (第2版). 日外アソシエーツ.
- [2] J. ヴァーレス 編 池谷 のぞみ 訳. 情報の要求と探索. けい草書房.
- [3] 特許公報 昭 61-204735. データ駆動浮動小数点演算回路. 日本国特許庁.