

# 質問の類似性および再利用に基づく データベースにおける対話支援機構

4 G-6

堀川健一 有川正俊 上林彌彦

京都大学工学部 広島市立大学情報科学部 京都大学工学部

## 1 まえがき

近年、ネットワーク上に散在している情報源データベースから、利用者質問に応じて必要な情報を抽出し、合成した情報を提供する環境の実現が重要な課題となっている。利用者による情報の抽出や合成そのものについての研究がなされているが、質問記述の作業を支援する環境についてはあまり研究されていない。

そこで、本稿では、データベースにおける利用者質問の曖昧さ、不十分さ、誤りを補うための対話支援機構の必要性が示される（図1）。本機構では、質問の意味的な類似性を用いて、過去の曖昧さの処理、表示オブジェクトの選択、表示方式などの質問処理の方式を再利用する。

筆者らは、重要な応用例として動的地図合成環境を開発している。本機構は様々な環境に応用できる。

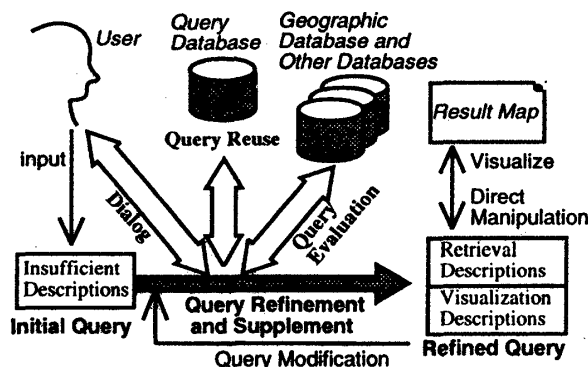


図1: 質問の再利用に基づく対話支援機構の概略図

## 2 地理データベース

地形、道路、橋、建築物、行政界などの地理情報に他の情報を合成して利用者に情報を提供する地理情報システムは、施設管理、都市計画、環境評価、防災、カーナビゲーションなどの分野で実用化されている。

地理データベースは、地理情報システム全般で共有可能な地理情報を一元管理し、地理情報システムに地

理情報を提供するデータサーバである。

地理データベース上の地理データは図形データ・インスタンス名（名前）・クラス名（種別）などを属性として持つ単位で扱われることが望ましい。各属性、特に図形データは表示に依存しない形式であれば、用途に応じて様々な可視化を行なうことができる。このような地理データを地理オブジェクトと呼ぶ。

## 3 利用者質問記述とその不十分さ

質問記述は値と演算子から構成される。値は、主にクラス名（例: 駅, 国道, 大学）とインスタンス名（例: 京都駅, 東名高速, 京都大学）である。他には、属性名（例: 利用者数, 交通量）や文字列や数値が値として使うことができる。演算子としては、「の近くの」、「の北の」、「の値が」、「を満たす」などがある。

例1: (京都大学 の北の 飲食店) の地図

例2: (京都大学本部構内 の北端から北側に 100m の範囲にある 飲食店) の地図

例3: (京都駅 の近くの) (深夜にあいている) 飲食店

$x$  や  $y$  を値や式,  $p$  を演算子とすると, 2項関係質問は  $p(x, y)$  や  $(x)p(y)$  の式で表される。例3は, 例4のように表すことができる。

例4: ( $near$ (“京都駅”, “飲食店”))

$\wedge$  ( $open(10p.m. \sim 4a.m.,$  飲食店 “))

$\wedge, \vee$  はそれぞれ積集合, 和集合を示す。質問内容の処理においては, このように分割して処理されるが, 検索実行時には, 質問最適化の結果によっては2つの検索条件が一括処理される。

### 3.1 質問の曖昧さと質問の明確化

利用者が一般に作成する例1のような質問には, 「京都大学」, 「北側」などの検索処理向きでない曖昧な記述が存在する。演算子に関して例を挙げると, 「の近くの」や「の北の」では, 検索範囲や検索件数の指定が曖昧である。さらに, 「北」には, 緯度的に北, 方角的に北, 北側に接するなどの明確化も必要である。例2は例1の明確化の一例で, データベースでの検索処理を行なうことができる。

Dialogue Supporting Functions of Databases Based on Query Similarity and Reuse

Ken'ichi HORIKAWA†, Masatoshi ARIKAWA‡ and Yahiko KAMBAYASHI†

†Faculty of Engineering, Kyoto University, ‡Faculty of Information Science, Hiroshima City University

### 3.2 質問の不十分さと質問の詳細化

例2のような質問では、可視化の指定や背景の指定などがない。利用者要求に合う地図を効果的に表現するために以下に示す手法の適用が必要である。

**総描** 空間的關係を効果的に表現する手法。その一つの背景とは用途や縮尺に応じて追加表示される地理情報のことで、例えば、例2では京都大学構内、目標となる建物、道路、交差点などである。

**利用者による選択表示** 例えば、飲食店の名前に加え、営業時間や店種を選択表示させることである。

**記号化** 色、線、記号、フォントの指定や凡例の表示。

このように詳細化は背景の追加と可視化の指定に大別される。詳細化された質問は規模が大きく複雑であるために利用者が作成するのは困難である [1]。

## 4 地理データベースにおける対話支援機構

### 4.1 質問改良に基づいた対話支援機構

本稿では、例1のような利用者質問に質問データベース上の過去の質問の処理結果を用いて、明確化、背景の追加、詳細化を行なう手法を述べる。この変換を質問改良と呼ぶ。質問改良では誤りの訂正も行なわれる [1]。本稿では便宜上、主に明確化について説明するが、背景の追加や詳細化についても同様である。

図1に質問改良を含む対話型支援機構を示す。

- i) 利用者が初期質問を作成する。
- ii) 質問改良を実行する。質問改良の候補が複数存在する場合には利用者による選択を要求する。
- iii) 質問改良結果が質問データベースに蓄積される。
- iv) 地図への直接操作や質問文への操作によって、質問の変更を受け付け、ii) の質問改良に移る。

### 4.2 表示限界の考慮

利用者の選択を要求する場合は、地図の表示を並行に行ない、視覚的支援を行なう [2]。可視化されるデータは、ディスプレイの表示限界による量的な制約を受ける。質問から導出される重要度に基づいて重要度の高いものから強調を伴う可視化を行ない、重要度の低いものは、表示が簡略もしくは省略される [3]。表示限界を超える場合には利用者に質問修正を要求する。

## 5 再利用と類似性

### 5.1 再利用による質問改良

質問データベースでは、質問毎に質問、その部分式、値、演算子の各レベルでの  $x$  から  $y$  への改良を  $x \rightarrow y$  の形式で登録し、部分的な再利用に用いる。

質問改良は、利用者質問を質問データベース上の類似性の高い部分質問の集合に分解し、それぞれの改良結果、つまり、 $x \rightarrow y$  の右辺の合成により行なわれる。

### 5.2 質問の類似性

類似性の計算のために、地理オブジェクト  $x$  や  $y$  に対して、意味的類似性、物理的広さレベル、役割的広さレベルを定義する [1]。

**意味的類似性**  $res(x, y)$   $x$  と  $y$  との意味的な隔たり。

分類をグラフ構造化したシソーラス上で定義され、一致を0として非負数値で示すことにする。

**物理的広さレベル**  $lp(x)$   $x$  の物理的な広さのレベル。広さのレベルとは、広さを計算上 (1) 番地, (2) 丁, (3) 町, (4) 区, (5) 市, (6) 県, (7) 地方, (8) 国の8レベルで扱うことにしたもの。

**役割的広さレベル**  $lr(x)$   $x$  の影響圏の広さのレベル

2つの地理オブジェクト  $x, y$  の近さの評価は、以下のように重み  $w_1, w_2, w_3$  により表される。

$$sim(x, y) =$$

$$w_1 res(x, y) + w_2 |lp(x) - lp(y)| + w_3 |lr(x) - lr(y)|$$

演算子  $x, y$  についても、意味的類似性が定義され、 $sim(x, y)$  が定義される。

これらにより2つの式の類似性の定義ができる。

$$sim(p_1(x_1, y_1), p_2(x_2, y_2)) =$$

$$sim(p_1, p_2) + sim(x_1, x_2) + sim(y_1, y_2)$$

以上は、2項関係質問の近さについて述べたが3項以上の多項演算質問も同様に定義される。また、さらに処理方式の再利用も考えられ、処理方式の立場から別の類似性も定義できる。

## 6 まとめ

本稿では、データベースにおいて、質問の再利用により利用者質問を質問結果データを合成する質問に変換する対話機構を提案した。

### 参考文献

- [1] 堀川健一, 上林彌彦, 有川正俊: “地理データベースにおける質問再利用の基本的な機構”, 電子情報通信学会技術研究報告, Vol.94, No.224, pp.47-54, 1994.
- [2] 有川正俊, 甲斐宏, 上林彌彦: “情報表現能力と応答の実時間性を考慮した動的な文字配置の実験および評価”, 情報処理学会第48回全国大会, No.4, pp.179-180, 1994.
- [3] 有川正俊, 河北秀世, 甲斐宏, 上林彌彦: “データベースビューに基づく動的対話地図オブジェクト”, 電子情報通信学会第5回機能図形情報システムシンポジウム, pp.107-112, 1994.