

自然言語によるデータベース検索のための対象分野知識入力支援ツール

2F-10

久保 加奈子 市山 俊治

NEC 関西C&C研究所

1 はじめに

データベース検索を行なう自然言語インタフェースにおいて様々なデータベース(以下DB)に対応するためには、それぞれの背後にある対象分野の知識、すなわち

- DBの構造に関する知識
- DB上での表現と自然言語表現との対応に関する知識

を知識ベースとして構築する必要がある。[谷91]

従来、知識ベースの構築では、知識ベースでの知識記述形式や自然言語処理などの専門知識に熟知しているDB管理者が直接その知識記述形式に従って知識を入力していた。しかしこのような方法では、専門知識のないDB管理者には構築が困難であり、多大な労力や時間を構築に費やすことになる。

そこで、これらの専門知識のないDB管理者でも、対象分野知識ベースを容易に構築できる対象分野知識入力支援ツールを試作した。

この対象分野知識入力支援ツールは以下のような特徴をもつ。

- DBMSから自動的にDBスキーマ情報を獲得する。
- 類義語辞書を参照することにより、自然言語表現の網羅性を高める。
- 自然言語インタフェースの解析機能を用いることによって自然言語による知識の定義を受け付ける。
- 自然言語による知識定義より対象分野辞書を自動的に構築する。

2 構成

対象分野知識入力支援ツールは、図1に示すように、DB管理者から知識を獲得する知識ベース作成ツールと、獲得した知識を知識ベースの知識記述形式で表現して編集を支援し、知識ベースに格納する知識ベース編集ツールから成る。

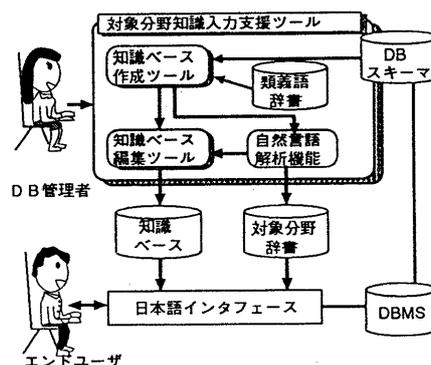


図1 対象分野知識入力支援ツールの構成

3 知識ベース作成ツール

知識ベース作成ツールはDB管理者から知識を獲得するツールであり、図2に示すようないくつかのサブツール群から構成されている。

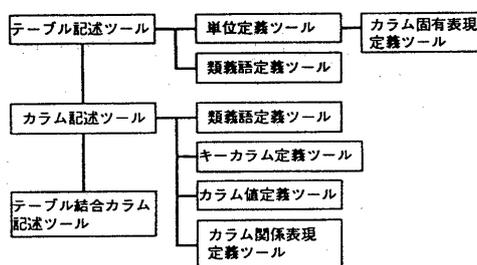


図2 知識ベース作成ツールの構成

1. DBの構成に関する知識

テーブル名、カラム名、カラムのデータ型などのDBスキーマに関しては、DBMSが持つ情報を自動的に獲得する。

他に、テーブル結合時にキーとなるカラムを定義するツールなども用意されている。

2. DB上での表現と自然言語表現との対応に関する知識

- 日本語表記対応付けツール(図3)

テーブル名、カラム名などのDB記述子に対応する日本語表記を定義する。

A Domain Knowledge Input Tool for Database Retrieval in Natural Language  
Kanakano KUBO and Shunji ICHIYAMA  
Kansai C&C Research Lab., NEC Corp.

● 類義語定義ツール (図3)

日本語表記対応付けツールで定義した日本語表記について類義語を定義する。類義語の候補は類義語辞書を検索することによって得られる。

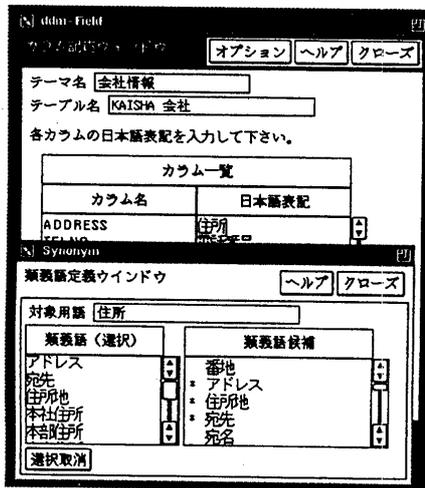


図3 日本語表記対応付けツールと類義語定義ツール

● カラム固有表現定義ツール (図4)

数値カラムにおいて、値が大きい/小さいとき用いるカラムに固有な形容詞表現とデフォルト値を定義する。

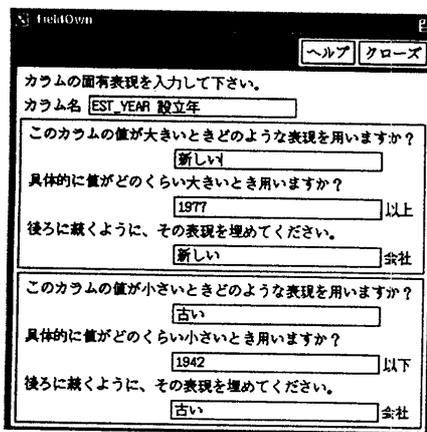


図4 カラム固有表現定義ツール

● カラム関係表現定義ツール (図5)

複数のカラムの間の関係を示す動詞表現を定義する。カラム関係表現は、自然言語インタフェースの解析機能を利用して解析され、対象分野辞書に新たに登録される。

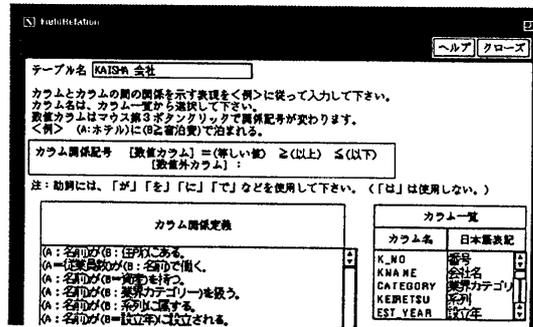


図5 カラム関係表現定義ツール

4 知識ベース編集ツール

知識ベース編集ツールは、知識ベース作成ツールで獲得した知識を知識ベースの知識記述形式であるネットワーク構造として視覚的に表現し、知識の編集を促すツールである。知識ベース作成ツールで定義できない高度な知識はこのツールで入力することができる。

5 おわりに

知識工学、自然言語処理などの専門知識のないDB管理者でも対象分野知識ベースを容易に構築できる対象分野知識ベース構築支援ツールを試作した。

これまで、10テーブル前後からなる2つのDBに対してこのツールを用いることにより、半日から1日で知識ベースを構築することができた。また、知識ベース構築の際、次のような課題が明らかになった。

- 類義語定義において、「自己資本比率ランキング」のような長い複合語の類義語候補が類義語辞書に存在しない。  
→複合語を短く分割したものの類義語候補を表示する。
- カラム固有表現において、「食事ができる」=「設備カラムにカラム値レストランがある」のような数値カラム以外のカラムに関する固有表現が定義できない。また、「規模」=「従業員数 and 売上高 and 資本金」のような表現が、自然言語で入力できない。  
→「○○とは××である(規模とは従業員数と売上高と資本金である)」のような形式の表現にして自然言語解析し、ネットワーク構造に変換する。

今後はこれらの課題を解決するとともに、DB検索以外でのこのツールの応用について検討を進めていく予定である。

[参考文献]

[谷91] 谷幹也, 飯野香, 山口智治, 市山俊治: 自然言語インタフェース構築キット:IF-Kit, 信学技法 NLC91-62, 1991.