

表項目日本語定義文からの集計用データの検索式生成

2M-9

山口 智治 市山 俊治

NEC 関西 C&C 研究所

1 はじめに

自然言語インタフェースの適用例としてデータベース(DB)検索に用いる事例は多い([Waltz78]など)。しかし、最近ではDB検索を単独で行なうよりも他のアプリケーションと連携し、検索結果を活用するという形態が一般的になっている。異種のアプリケーションの操作を高い抽象レベルで記述する試み[難波93]もあるが、アプリケーションの操作を統一的に記述しようとすると、個々のアプリケーションの特徴の活用が困難になる。逆に、アプリケーションの特徴を活かして連携することでDB検索用自然言語インタフェースの利用場面も拡大するが、アプリケーションに依存した構成が必要になり、自然言語インタフェースの汎用性を損ないやすい。

本報告では、スプレッドシート型アプリケーションからDB検索を行なう際に、日本語インタフェースを用いる試みについて述べる。特にスプレッドシートに特徴的な表形式と項目定義を利用する日本語文の入力形態と処理手法について述べる。

2 スプレッドシートからの日本語DB検索

2.1 日本語検索文によるDB検索結果の利用

データベースに蓄積されたデータを利用するために表に取り込む機能は多くのスプレッドシート型アプリケーションが備えている。しかし、実際にデータを取り込むためには複雑なメニュー選択を繰り返したり、DB検索言語を記述したりする必要がある。

日本語DB検索システムへの入出力をスプレッドシートに反映できるならば、このような負担は軽減される。

2.2 項目名からのDB検索

スプレッドシートに特徴的な構造に項目欄がある。項目欄には、表を埋めるデータがどのような属性に対する値であるかを示す情報を記入する。この情報を日本語DB検索に利用すれば、項目欄に項目名を定義するだけで表を埋めるデータをデータベースから検索してることが可能になる。また項目名がデータベースのフィールドやデータと一致しない場合にも、自然言語インタフェースを介することで語彙のゆらぎの吸収が可能である。(図1)

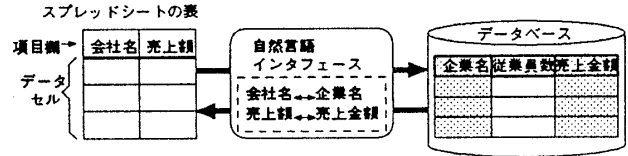


図1: 項目名からのDB検索

2.3 項目の日本語定義文の利用

実際に表を作成する場面では、項目名とデータベースのフィールドやデータとは単純に対応づけられるものばかりではない。データの値によって分類した結果などを表にする場合がある。また、項目名には簡潔な表記や特殊な略語を使用される場合も多い。項目名だけでは日本語DB検索を実行するには情報が不十分である。

このような場合には項目名を説明する定義文を与えることによって日本語DB検索を実行可能にする。例えば、企業情報に関するデータベースから従業員数によって大規模/中規模/小規模に3分類した表を作成しようとするとき、項目名“大規模”に対して「従業員数が5000人以上の会社」というような日本語文による項目の定義を与える。日本語DB検索システムは項目名の代わりにこの定義文を処理して検索を実行する。

3 表項目からの検索式生成

3.1 項目、項目欄の分類

各項目定義文を自然言語インタフェースの言語解析モジュールで解析し、参照するデータベースフィールドとデータを抽出する。フィールドの値がデータで規定されていない場合は、そのフィールドを検索対象とする。データおよびデータで値が規定されているフィールドは検索条件とする。各項目を以下のように分類する。

対象項目: 検索対象のみを含む

条件項目: 検索条件のみを含む(特定のデータに等しいことを条件とする場合は特定条件項目と呼ぶ)

複合項目: 検索対象と検索条件の両方を含む

項目の分類に基づいて項目欄を次のように分類する。

対象項目欄: 対象項目のみを含む

条件項目欄: 条件項目のみを含む(特定条件項目のみの場合は特定条件項目欄)

複合項目欄: 対象項目、条件項目、複合項目が混在

3.2 表のタイプと検索方法

項目欄の分類を用いて次のように表を分類する
タイプ0: カラムとローの一方のみに項目欄を設定

タイプ1: カラム項目欄とロー項目欄の一方が対象項目欄、他方は条件項目欄

タイプ2: 一方が複合項目欄、他方は条件項目欄

タイプ3: 一方または両方が複合項目欄

それぞれ次のような検索方法が可能である。

タイプ0 → 一括検索

タイプ1 → カラム/ロー毎の検索 (ただし条件項目欄が共通なデータベースフィールドを参照する特定条件項目欄の場合に限り、一括検索が可能)

タイプ2 → カラム/ロー毎の検索

タイプ3 → セル毎の検索

検索方法については以下の通り。

一括検索: カラムの項目とローの項目をすべて組み合わせ、ひとつの検索式を生成し、表を埋めるデータすべての獲得を1回の検索で実行する。

カラム/ロー毎の検索: 各カラム (ロー) についてロー (カラム) の項目のみをすべて組合せて検索式を生成し、カラム (ロー) の項目毎に検索を繰り返す。

セル毎の検索: 該当するカラムの項目とローの項目を組み合わせ、検索式を生成し、セル毎に検索を繰り返す。

一般に、一括検索が可能ならばカラム/ロー毎の検索も可能で、カラム/ロー毎の検索が可能な表ではセル毎の検索も可能である。

3.3 中間テーブルを用いる一括検索

タイプ1の表では一括検索可能な場合が限られている。しかし、一時的な中間テーブルをデータベースに作成することで、タイプ1の表の一括検索が可能になる。

中間テーブルには、条件項目を指標とする分類をデータとして参照できるように項目名とキーフィールドとを関連づけて記録する (図2)。

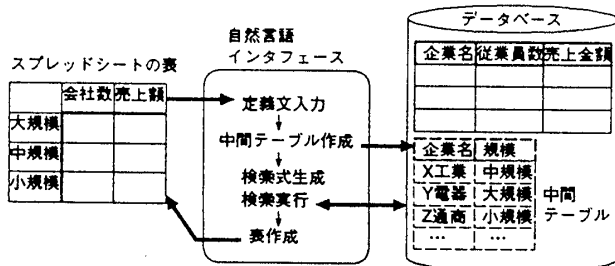


図2: 一括検索のための中間テーブル作成

作成した中間テーブルを既存のテーブルと同様に扱う。分類を表すデータを中間テーブルを用いて参照するようにし、その分類の値でグループ化する検索式を生成する。

4 システムの試作

以上の枠組に基づいて、日本語 DB 検索システムと市販のスプレッドシート型アプリケーション “Wingz” を結合したシステムを試作した。基本的な表の形式であると

思われる、カラムの項目を検索対象としローの項目を検索条件とするタイプ1の表を作成する。

ユーザ入出力部にはスプレッドシートのもつ対話部品を利用した。図3はロー項目定義の様子を示す。

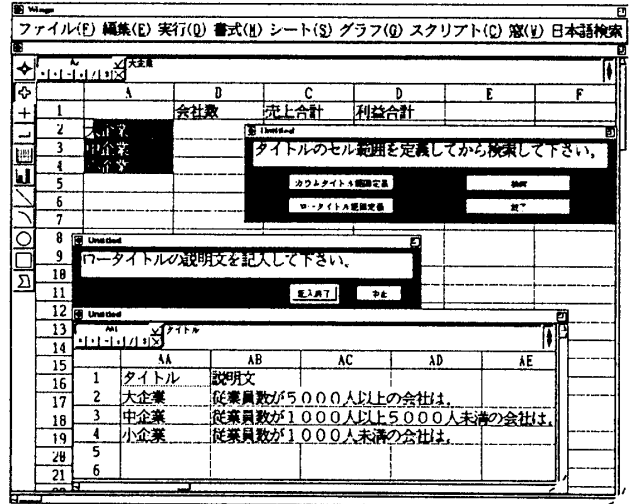


図3: 表作成のための項目定義例

日本語 DB 検索システム部では項目定義文の合成をおこない中間テーブルを参照して一括検索する検索式を生成する。定義文合成部の導入など、日本語 DB 検索システムの処理の枠組を拡張する必要があったが、処理モジュールを部品化した自然言語インタフェース構築キット [市山 91 谷 91] に基づく日本語 DB 検索システムを用いることで処理に柔軟に対処できた。

5 おわりに

個々のアプリケーションの特徴を積極的に活かす形での日本語 DB 検索システムの異種アプリケーションとの結合について述べた。スプレッドシートをアプリケーションの例とすると、項目やその定義文を用いることで日本語 DB 検索システムを介して表を埋めるデータを検索することが可能になる。

スプレッドシートに限らず種々のアプリケーションについても、活用可能な特徴をそれぞれのアプリケーションから見出し、自然言語インタフェースの構成を適応させる手順の定式化が課題である。

参考文献

[Waltz78] D.L.Walts: An English Question Answering System for a Large Relational Database, CACM, Vol.21, No.7, pp.526-539, 1978.
 [難波 93] 難波康晴, 辻洋, 絹川博之: 自然言語インタフェースにおける知識表現の統一とネットワーク探索, 45 回情処全大 2F-8, 1992.
 [市山 91] 市山俊治, 村木一至: 自然言語インタフェースの構築キットの提案, 43 回情処全大 1H-2, 1991.
 [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.