

電子化辞書管理のための自然言語インタフェースシステム

9B-7

— システム概要とその構成 —

落合 尚良[†] 森 義和[†] 奥井 伸司^{††}
[†](株)日本電子化辞書研究所 ^{††}NEC技術情報システム開発

1. はじめに

大規模なデータベースやテキストベースでは、これらのデータの維持管理を行なうためのシステムは必要不可欠である。これらのシステムには、逐次検索をはじめ体系的なチェック、一括した更新や、バックアップなどの機能を要求される。また、これらの機能を扱いやすくするため日常使用している自然言語をインタフェースとして用いることは有望な手段であると考えられる。

本稿では、現在の自然言語システムの能力を踏まえ、自然言語をインタフェースとして用いた、管理作業を指示できるシステムを試作した。その概要と構成を述べる。本システムは(株)日本電子化辞書研究所(以下EDRと略す)で開発している大規模電子化辞書を対象としている。

2. 自然言語インタフェース実現の課題

データベースの管理者にとって、データの内容を詳細にチェック、更新するといった管理要求に応えるシステムは当然必要であるが、管理者以外の利用者也、日々の作業としてデータベースの内容を複雑な条件によって、情報を得る作業を行なっており、その作業は管理者と何ら変わらない。従って、管理者以外にも管理に必要な機能を利用することがあり、それらの操作は、利用者にとっても理解しやすいことが重要である。

インタフェースとしての自然言語は、その表現の自由度が高い点や習得の必要がないなどの利点も多い。しかし、問題もいくつかある。そのため、管理に必要な機能を自然言語を用いた場合どのような問題があるかを検討する必要がある。このための方法として、実際に操作を行なっている利用者に対し、管理業務を自然言語で記述した例文を集めて分析し、自然言語を用いた管理システムで要求される機能と、実現可能な方策をまとめた[松山93]。

- ・大規模電子化辞書を管理していくには、逐次的な検索、削除などの機能と、一括した内容の検証と追加など、多くの条件を記述したプログラムを解釈し実行する機能
- ・ファイルの書き出しや、データのバックアップなど管理する上で必要な機能
- ・定型作業は、名前をつけることで自然言語の入力における手間を軽減できる機構
- ・入力文形式に簡単な規約を付けることで、解析の曖昧性を排除でき、むだな問い合わせを軽減できる。
- ・長い条件文は利用者自身でも内容の検証が難しい。また、このような文はいくつかの文に分割し、適切な指示代名詞によって表現することが可能である。

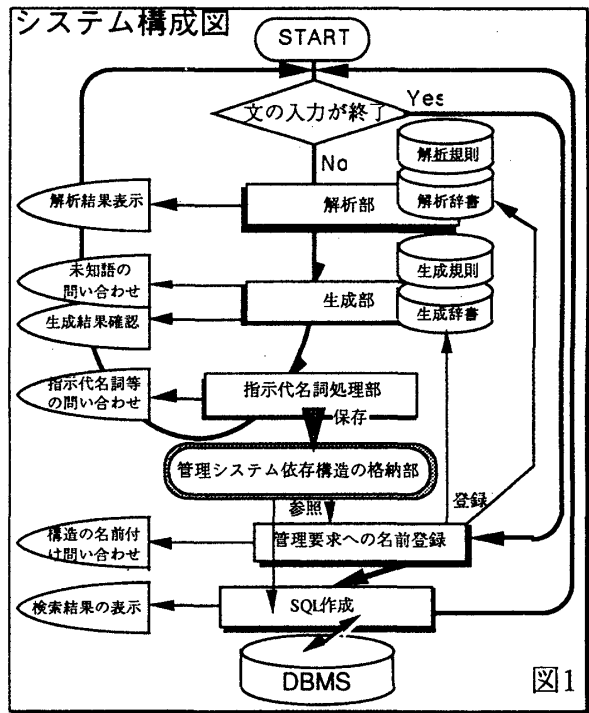
3. 実現への方策

実現すべきシステムでは、ユーザとの使い勝手を考慮し、現在構築している自然言語モデルの能力とマシンの能力を検討し以下のような機能を持つシステムを設計した。

- 1) 入力文の処理機能
処理を自然言語で一文ずつ入力し、解析後、DBMSの処理言語に生成する機能。
- 2) 繰り返し指示や条件を入力できる機能
複雑な条件を入力するために、複数の文に分割して入力する事を可能にする機能。
- 3) 指示代名詞の処理機能
複数の文に分割された条件は、それぞれ指示代名詞を用いて表現されている。この関係を決定する機能。曖昧な場合、ユーザに問い合わせる機能がある。
- 4) 未知語等の語彙獲得機能
未知語に対し、その掛かり受け関係などから未知語の推定を行ない、ユーザからのインタラクションを減らす機能。

4. システム構成

試作システムの構成図を図1に示す。



一文ずつ入力された文は、解析部では解析辞書と解析規則により中間構造を生成する。中間構造をユーザに示し、曖昧な中間構造を生成した場合にシステムが選んだ構造以外を指示することも出来る。

次に、生成部では、生成辞書と生成規則により管理システムに依存した構造に変換する。この際、曖昧なもの未知語については、必要最小限の問い合わせにより、未知語の語彙獲得を行なう。作成した管理システム依存の構造は中間構造と同様にユーザに内

Natural Language Interface for Electronic Dictionary System Manager
 - An overview of system -

Takayoshi OCHIAI, Yoshikazu MORI, Shinji OKUI
 Japan Electronic Dictionary Research Institute, Ltd
 NEC Scientific Information System Development, Ltd
 *本研究はEDR在任中に行なったものである。

容を表示しユーザからの指示によって、別の解を求めることが出来る。

さらに、指示代名詞処理部では管理システム依存の構造に指示代名詞がある場合、指示先の決定を行なう。これらの処理後、管理システム依存の構造を保存し、次の入力文を処理する。

管理要求の入力が終わったら、保存されたすべての管理システム依存の構造を取り出し、DBMSにアクセスし指示を実行するためのコマンドを生成する。さらに、DBMS等に接続しコマンドの実行を行なう。また、実行する前に、入力された管理要求に対し、ユーザは名前を登録することが出来る。以後この処理を示すマクロとして扱うことが出来る。

実行終了後、新たな管理要求の入力を行なうことが出来る。

5. 動作例

本システムで動作する例を以下に示す。ここでは、ユーザからの入力や指示を「入力：」と示し、システムが表示する内容を四角で囲って示している。

入力：動詞と形容動詞を教えてください

(1) 教えて(形容動詞² and 動詞³)
 (2) 教えて(形容動詞² or 動詞³)
 > 解析結果は(1)です。いいですか?

yesまたは、指定の文を入力する。
 表示された構造が正しければ、そのまま実行し
 正しくない場合正しい候補を選ぶ

入力：2

単語¹を表示
 品詞が(形容動詞² or 動詞³)
 > 生成結果です。いいですか?

入力：yes

一文の入力が終了し、指示文の入力待ちに戻る。

入力：それは5つ以上の概念を持つ単語です。

(1) だ(単語² 持つ³ 概念⁷(5つ以上⁵)), それ³)
 > 解析結果です。いいですか?

入力：yes

それ(単語²(5つ以上(文字数(概念⁷))))
 > 生成結果です。いいですか?

入力：7

上記の生成結果は間違っており、その間違いは概念⁷にあることを示している。なお、入力した番号はシステムが単語に付けた番号である。

1) 概念コードの文字数
 2) 概念の数(単語に対する)
 3) 概念の値(数値化した)
 4) 解析構造の間違い
 > 現在は1)です。いいですか?

概念に対するシステムの予測を表示します。

入力：2

それ(単語²(5⁶>=(数(概念⁷))))
 生成結果です。いいですか?

入力：yes

1) 単語¹を表示
 品詞が(形容動詞² or 動詞³)
 > 指示代名詞は「単語」を指示していますが、いいですか?

入力：yes

一文の入力が終了し、指示文の入力待ちに戻る。

入力：その頻度は1以上です。

(1) だ(頻度¹²(1以上¹¹)), それ¹⁰)
 > 解析結果です。いいですか?

入力：yes

それ¹⁰(1¹¹<=(頻度¹²))
 生成結果です。いいですか?

入力：yes

1) 単語¹を表示
 品詞が(形容動詞² or 動詞³)
 2) それ¹は
 5⁶>= 数(概念⁷)
 > 指示代名詞は「概念」を指示しています。いいですか?

入力：1

指示代名詞の候補を示し、間違っている場合指し示す単語を選択する。

入力文が終了したので、実行に移る。

> この一連の作業に名前を付けますか?

入力：「対象活用語」

「対象活用語」として名前を付けました。

1) 単語¹を表示
 品詞が(形容動詞² or 動詞³)
 2) それ¹は
 5⁶>= 数(概念⁷)
 ・ それ¹は
 1¹¹<= 頻度¹²
 > 実行の内容です。いいですか?

入力：yes

見出し 読み 品詞……
 ……………
 実行結果の表示です。次に入力しますか?

実行し、結果を表示する。また、スタックした文構造を破棄します。

その後、指示文入力に戻ります。

入力：対象活用語を検索します。

単語を表示
 「対象活用語」¹
 > 生成結果です。
 > 「対象活用語」は登録された処理です。表示しますか?

入力：no

解析結果表示を省いて生成結果を表示します。マクロ登録された単語を含むので、内容表示を行なうことが出来る。

6. おわりに

自然言語をインタフェース用いた辞書管理システムの構成と、機能について述べた。試作したシステムを作業者に利用してもらうことで、での操作では、今後は、本システムの処理速度や、使用メモリの削減などを中心に利用しやすいシステムとして改良を行なう必要がある。また、現在の解析辞書などはEDR辞書から抽出したカスタム辞書であるが、これをEDR辞書を利用した解析システムに変更し、本システムの動作によるEDR電子化辞書の評価を行なう予定である。

謝辞

本研究の機会を与えて下さった横井所長に深く感謝いたします。

参考文献

[松山93]「辞書管理における自然言語インタフェース -コーパスの分類-」46回情処全大
 EDR電子化辞書：EDR,TR-016 1989