

日本語対話システム「Anyone」 自然言語によるエンドユーザコンピューティング

板橋美子 大塚仁司 野村千佳子 高梨郁子 辻秀一 樋口雅宏
三菱電機株式会社

自然言語解析技術を用いたエンドユーザ向けビジネス系データベース検索システム「Anyone」を開発した。「Anyone」の開発にあたっては、応答速度とユーザからみた入力文の組み立てやすさを重視して、名詞句中心の単純な文型に入力を制限するという方針をとった。

本稿では、販売管理分野での情報系処理の「Anyone」上での構築における、エンドユーザの検索要求を実現するためにとった方策と、「Anyone」の概要を報告する。

"Anyone"-An Interactive System In Japanese End User Computing Featuring The Natural Language Analysis

Yoshiko Itabashi Hitoshi Ohgashi Chikako Nomura
Ikuko Takanashi Hidekazu Tsuji Masahiro Higuchi

Mitsubishi Electric Corporation

We have developed a business database retrieval system named "Anyone", that features the natural language analysis technology to respond to end users. In developing "Anyone", we adopted a method that limits input-statements to a certain style features some noun phrase, so that users could get more rapid responses and feel easier to input a statement.

We report the overview of "Anyone", as well as the methodology that we take on this system, "Anyone", to establish an efficient data processing for sales management and handle the requests from end users effectively.

1. はじめに

現在、コンピュータの利用は、単なる業務の合理化にとどまらず、自社の競争優位性を確保するためのSIS（戦略情報システム）、「情報」資源の有効活用のためのシステムなど様々な側面で進んでいる。他方、パソコンの普及に代表されるダウンサイジングの進展は、情報システムを情報システム部門からエンドユーザの手に取り戻そうとする動きとも考えることができ、コンピュータの利用技術は、分野や利用方法、利用人口の数と質の拡大に直面している。（参考文献1）

エンドユーザコンピューティングはこのような背景から重要度を増している。いわゆる基幹業務は、データの保全や整合性の保持の必要性が高いものが多いため、従来の情報システム部門主導のシステム構築が多いが、蓄積されたデータを情報として活用する部分は、情報を必要とする人達、すなわちエンドユーザ主導の形態が適している。そのような情報システムを構築するには、次のような方法が考えられる。

- (1) 基幹業務のデータのうち必要な項目／レコードをあらかじめ選択して、部門サーバに定期的にあるいは随時転送し各個人のWSからOAソフトなどを使用して参照／加工する。
- (2) 各個人のWSから基幹データを直接検索し、必要ならばOAソフトなどで加工する。

(1)の方法では、選択したデータが適切か、追加や変更が必要な時の対処等が問題となる。(2)の方法では、エンドユーザが基幹データの構造やそれを操作するためのプログラム言語の知識を持っていないという点が問題となる。

私達は、「ワープロは打てるがOAソフトや4GLを使いこなせない」ユーザを想定して、(2)の基幹データを直接検索する方法をとって、基幹データの構造やプログラム言語の知識を辞書として備えた日本語対話システム「Anyone」を開発した。

日本語文を入力手段として用いることにより、自然言語の次のような特長をいかしたエンドユーザコンピューティングを実現できる。

- (1) 共通性…操作する人あるいは対象システムの変化に対する許容度が高い。
- (2) 非手続き性…検索したい項目や条件を日本語で入力するだけでシステムが自律的に動作する。
- (3) 抽象性…コンピュータ内の正確な表現でなくても、日本語による抽象的な表現でよい。
- (4) 曖昧性…ユーザの要求が明確でないときも、入力を受け付けられる。

反面、現在の自然言語解析技術の限界から、処理機能や性能が制限される（参考文献2）。また、システムの機能の限界が、エンドユーザからみてわかりづらいという欠点がある。

私達は、このような問題点を解決するために、検索要求と表示／解析パターンの指示を表現する単純な文型に入力を制限するという方策をとった。また、日本語対話システムを構築するには、基幹データの構造やプログラム言語の知識を辞書として登録することが必要である。私達は、最初の試みとして、販売管理の分野で、実際に販売されて実稼働している販売管理のパッケージの情報系処理を対象とし、日本語対話システムを構築した。その過程で、情報系システムの検索ニーズをユーザが制限された文型で要求できるように、様々な機能を開発した。

以下、2章で、販売管理パッケージの情報系システムの日本語対話システム化の報告をおこなう。

3章では、日本語対話システムの構成と対話処理の動作、検索要求を表現する文型を説明する。

2. 販売管理パッケージの情報系処理の日本語対話システム化

2.1 非定型処理のニーズ調査

Anyoneは4GLなどと違って、適用分野、適用システム別に辞書を構築する必要がある。この辞書に記憶された用語や処理パターンが、システムの検索レパートリになる。

そのため、システムを構築する前にニーズを調べ、必要なデータや用語を洗いだし、重要度を評価する必要がある。調査は販売管理パッケージの設計者1名、個別の販売管理システムを担当しているシステムエンジニア2名に、インタビューする形で実施した。

得られた主な意見について以下にあげる。

(1) システムの有用性

ユーザや適用システムは異なっても、知りたいことは言葉にしていれば同じことが多い。そういう意味でこのようなシステムは有用になる可能性が高い。

(2) 非定型検索のニーズ

販売管理の分野でユーザが知りたいことは、第一には「売上金額」「粗利金額」「粗利率」を様々な角度から様々な条件で調べたいということに尽きる。第二の重要度で、「在庫」「売上数量」などのパラメータがある。主な検索パターンについては、パッケージ自身に用意されているので、それからはみ出した条件やキーが現状のパッケージの非定型検索ニーズである。

(3) データの必要性

(2) であげられなかったデータは、現在、非定型検索のニーズを調査してもあがってはこないかも知れないが、データとして存在するものは、原則として検索できるようにして欲しい。条件を指定して商品名を検索したり、商品情報を検索したりは、現在、適当なツールがないので具体的な要求がないが、簡単にできるなら潜在的なニーズはあるのではないか。

(4) データの見方

検索結果は単に値を表示するだけでなく、簡単にグラフ化したり、「目標値」「予想値」「前年の値」などと対比させたりしてみたい。

(5) 操作性

検索には図1に示すような流れがある。まず、全体の傾向を調べ、それから原因となる要因を調べていく。そのための操作の流れをサポートするようなインタフェースにして欲しい。一つの問い合わせを一度ずつプログラムを起動したりプログラミングするのは面倒である。また、経営者、部門管理者などのユーザの立場によっても、操作の流れは異なる。

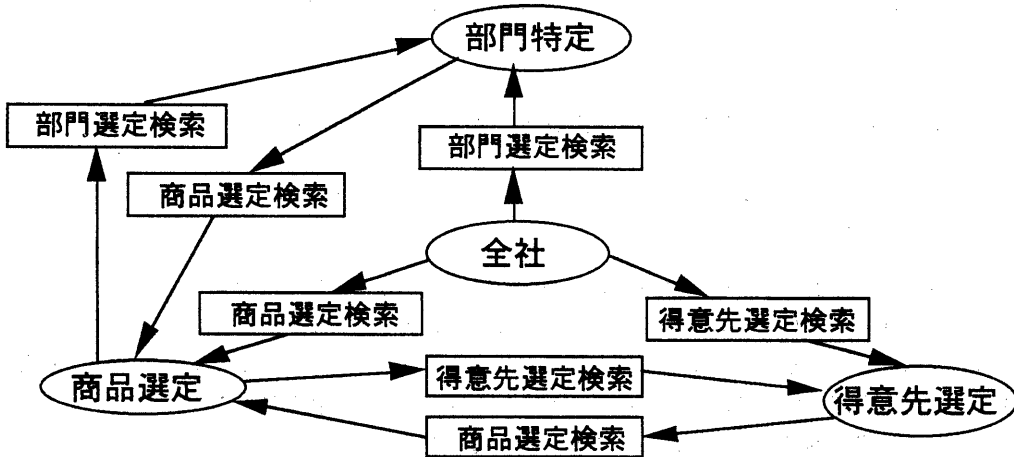


図1 経営者の分析視点

2.2 システム設計の方針

ニーズ調査に基づき、Anyoneは、次のような方針で設計した。

- (1) システムの主題となるような言葉、販売管理の場合は「売上金額」「粗利金額」「粗利率」「在庫」「売上数量」について、全体の傾向と分析の視点別に整理した例文集を作成し、ユーザはそれを参照しながら、視点を順次かえながら、データの検索/分析をできるようにする。
- (2) 例文集の構成は、ユーザ毎に変更可能にする。
- (3) 主題となる言葉に対応するデータについては、データの構造上可能な検索は、すべて可能になるよう辞書を作成する。
- (4) その他のデータは、可能な限り検索できるような辞書を作成する。

2.2 辞書の設計の方針

Anyoneは、基幹データの構造やプログラム言語の知識を辞書（意味辞書、手続辞書）として実装している。辞書の適用システム毎に登録すべき情報の中で、最も重要なものは基幹データの構造情報である。プログラミング言語としては、標準データベース検索言語であるSQLを使用することとし、その知識はシステム内部で記憶している。

Anyoneでは、基本的に各単語間の意味的な関連付けができるかどうかという問題を、各単語を組み合わせた検索が可能かどうかという問題に置き換えて、入力文を解析している。そのため、データの構造情報は、辞書のなかの意味辞書に記憶されている。

データの構造情報を辞書に記述するときは、図2のように各データを次のような分類でクラス分けして再構成した概念スキーマの形式で記憶している。

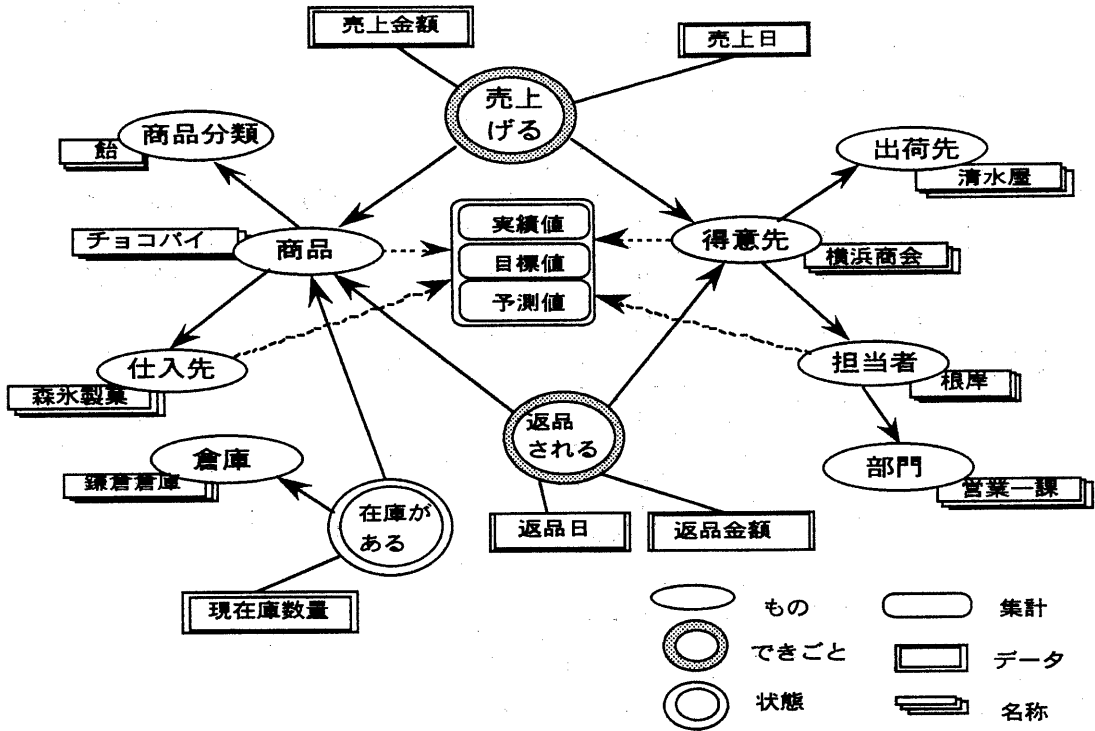


図2 概念スキーマの例

- (1) マスタ (もの) …ものとその性質を記憶
- (2) できごと…トランザクション情報
- (3) 状態…ものどうしの関連した状態を記憶
- (4) 集計…集計済みの値とそれと対比される「目標値」「予想値」など

実際のデータの構造（実スキーマ）が、この概念スキーマと同じなら、辞書の作成は難しい作業ではないが、実際にはかなりくいちがう点があるので、SQLにおけるViewの生成の方法を援用して、再構成しなおしている。

3. 日本語対話システム「Anyone」の概要

3.1 システム構成

Anyoneは、次のような部分から構成される。

- (1) 対話プログラム
- (2) 辞書
- (3) 辞書構築ツール群

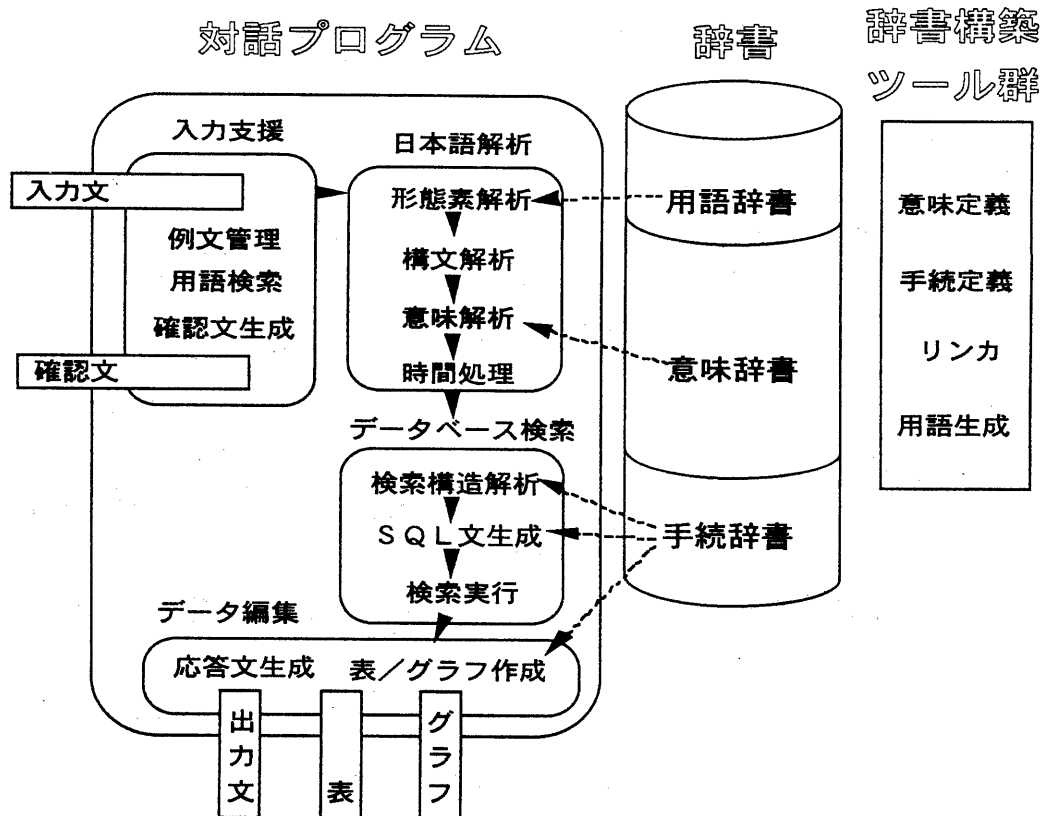


図3 Anyoneのシステム構成

対話プログラムは、図3にしめすように、大きくは「日本語解析部」「データベース検索部」「データ編集部」「入力支援部」からなる。図には示さないが「制御部」が、各部分の制御をおこなう。

辞書は、「用語辞書」「意味辞書」「手続辞書」の3つの部分からなっている。「用語辞書」は単語の形態素情報を記述している。「意味辞書」は、データの構造情報を仮想スキーマとして含む単語間の関連付け情報(概念スキーマ)を持っている。「手続辞書」は、実際の検索手続を記述するのに必要なコンピュータ上での名前を持っている。具体的には、データのテーブル名/カラム名(DBスキーマ)、表示形式を指定するためのマクロ名、データベースの検索演算の情報をもっている。

辞書構築ツール群は、「意味定義部」「手続定義部」「リンカ」「用語生成部」の各部分からなっている。「意味定義部」は概念スキーマを記述するツール群で「手続定義部」は、データ名等を記述するツールである。「リンカ」は「意味」の「手続」の関連付けをおこなう。

3.2 対話処理の動作

まず対話プログラムの標準的な処理フローの概略を示し、後で各フローの詳細を示す。

入力された日本語検索文を、まず日本語解析し、概念スキーマをノードとする意味ネットワークに変換する。必要に応じて確認文を生成し、ユーザに提示する。そしてSQL文を生成し、データベース検索をおこなう。結果は応答文または、表あるいはグラフで表示する。

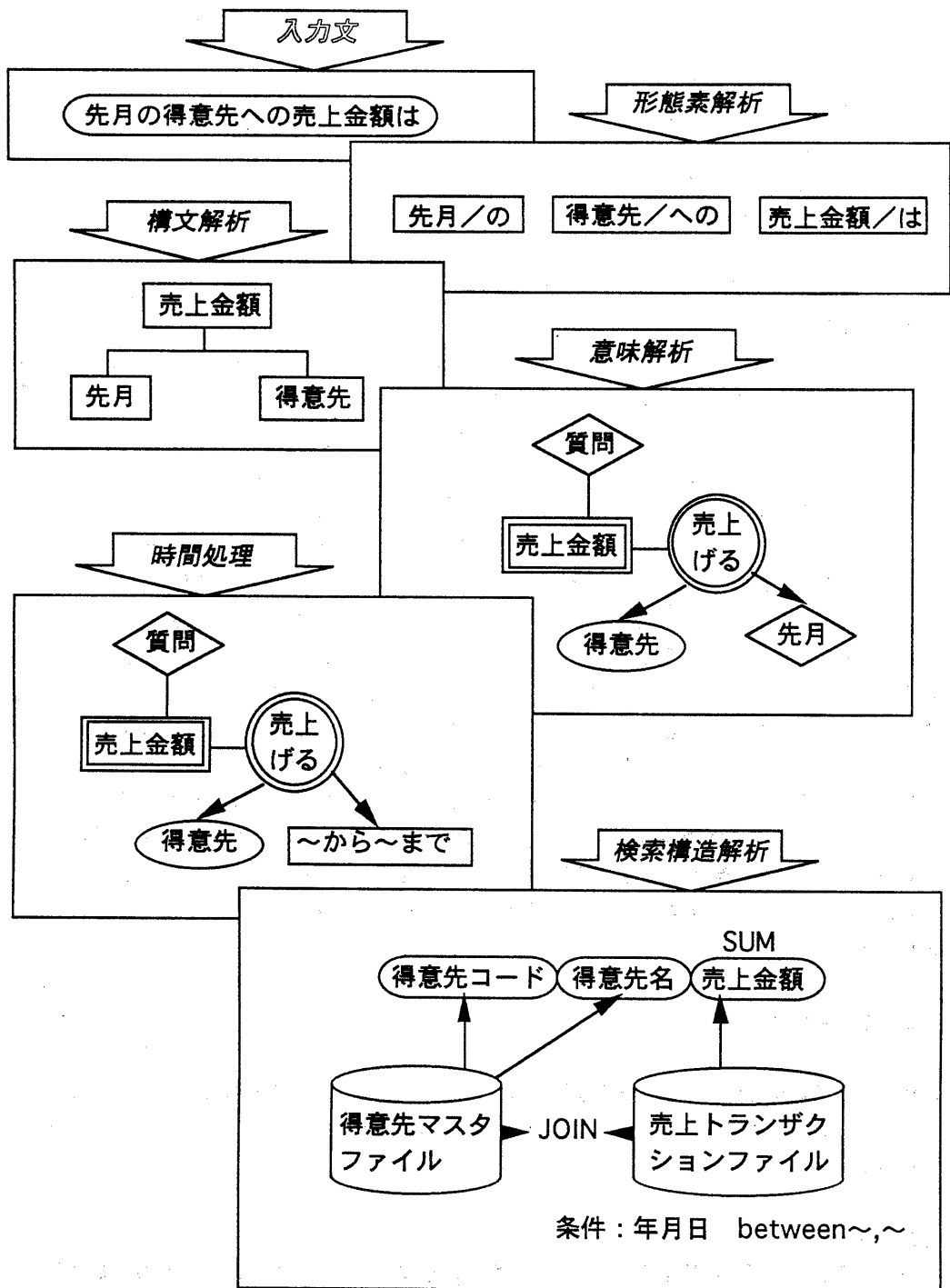


図4 対話プログラムの処理フロー

次に、図4に、各フローの詳細を示す。例として、検索文「先月の得意先への売上金額は」を用いる。

日本語解析部は、形態素解析、構文解析、意味解析、および時間処理からなる。形態素解析は、日本語検索文を文節単位に区切る。例では、「先月の」「得意先への」「売上金額は」という3文節の列になる。構文解析は、文節列の係り受けを判定し構文木に組み上げる。例では、「先月の」および「得意先への」の2つの文節が「売上金額は」に係るという構文木を出力する。意味解析は、主要な形態素に対応する概念スキーマを選択し、その制約条件を用いて構文木を概念スキーマ・ネットワークに変換する。例では、「先月概念スキーマ」「得意先マスタ概念スキーマ」「売上概念スキーマ」を選択し、概念スキーマ間の関係する項目間をネットワーク結合する。時間処理は、日本語の期間表現を具体的な期間条件に展開したり、省略された場合は既定期間条件を設定する。

データベース検索部は、検索構造解析、SQL文生成、検索実行からなる。検索構造解析は、概念スキーマ・ネットワークを解析して、検索する項目を選び出し、手続辞書に記述されたDBの実スキーマからなる検索木を作成する。例では、検索結果は「得意先コード」「得意先名」「売上金額」であり、《売上トランザクションファイル》に「先月の」に対応する検索条件をつけて、DB上の結合をはった検索木を生成する。SQL文生成は、検索木からSQL文を生成する。検索実行は、検索するデータベースシステムにあわせて、検索の実行を制御し、検索結果ファイルを受け取る。

データ編集部は、応答文生成と表/グラフ作成からなる。応答文生成は、検索結果が1件1項目の場合に、「[検索主題]は、[検索結果]です。」という形式の日本語文を生成する。表/グラフ作成は、検索構造解析結果から、検索結果の検索キーの数を判定して、検索結果ファイルの編集表示をおこなう。検索結果を表で表示するかグラフで表示するかは、ユーザが検索終了後に指示する。特に、検索文中に特定の表示形式を指示する言葉があったときは、その形式で表示する。例では、検索キーは1つなので、グラフ化する場合は「得意先名」をX軸に「売上金額」をY軸に設定する。

入力支援部は確認文生成、例文管理、用語検索からなる。確認文生成は、時間処理後の概念スキーマ・ネットワークを日本語表現に戻し、ユーザに提示する。例では15日締めの場合、「<売上日>1993年6月16日から1993年7月15日までの得意先への売上金額は？」となる。ユーザは日本語解析の結果をチェックしたり、修正できる。例文管理は例文の提示/編集、用語検索は使用可能な用語の提示をおこなう。

3. 3 処理可能な日本語

対話プログラムが処理できる制限日本語について述べる。ニーズ調査結果を満たす最小限の表現を(1)~(4)のように設定した。さらに、表現上の揺れに対応するために(5)~(7)などの強化をおこなっている。

(1)基本表現

「~の~の~は」が基本的な表現である。検索主題を文末の「~は」で示し、複数の条件を「~の」で修飾する。複数の条件の語順は原則として自由である。基本表現によって、検索主題に関して様々な分析視点を条件として組み合わせたデータが入手できる。なお、検索主題が計算可能なデータで検索キーが設定されていない場合は、結果は合計値となる。

(2)比較表現

「～の～が～より大きい(小さい)～は」という相対比較および、「～の～が大きい(小さい)～は」という絶対比較がある。相対比較は「～が」で表現した比較対象の検索データにしきい値を設定する表現であり、データの降順(昇順)にソートする。絶対比較は、全データのソートになる。

(3)期間表現

「～から～まで」という具体期間表現と、「過去～カ年(月,日)」および「上(下)期」や「春夏秋冬」などの相対期間表現がある。開始終了年月日が完全でない場合は、完全な具体期間表現に展開補完する。

(4)検索キー表現

「～の」でも暗示できるが、「～別の」で検索キーが明示でき、表/グラフの軸を指定できる。

(5)複合語

DB項目名が「当年月別粗利金額」といった複合語の場合に、これを「月別の～の当年の粗利金額」のように要素語に分離して表現できる。

(6)埋め込み文

活用語連体形を用いて「～した～」などの埋め込み文を1段だけ表現できる。

(7)文末表現

検索主題「～」のみの体言止めや、「～を見たい」「～を表示せよ」などの表現ができる。

以上の制限日本語に誘導するための支援機能として以下を用意している。

- (a)業務別にニーズの高い例文を実装できる 例文集
- (b)使用可能な語彙を調べられる用語検索
- (c)検索文の問題箇所と修正案を可能な限り提示するエラーメッセージ
- (d)対応する概念スキーマが本質的に曖昧な形態素の場合に、候補を列挙しユーザ選択を促す問い返し
- (e)同義語やユーザ固有の概念スキーマが追加できる辞書構築ツール

3.4 表示形式

検索結果の表示形式は、表1のようなものを準備し、関連づけられた言葉で指示すると、自動的に選択されるようにした。表示形式に関連づけられた言葉がないときは、標準形式を選択する。

また、この表示形式は、ユーザ向けに開放された辞書構築ツールで追加することができる。

表1 表示形式

形式	～は	～別の～は	～別の～別の～は
標準	棒グラフ	棒グラフ	3Dグラフ
売上状況	——	重ね合わせグラフ	——
A B C分析	——	A B C分析	——
月別前年比較	重ね合わせグラフ	選択後グラフ化	——
月別目標達成状況	重ね合わせグラフ	選択後グラフ化	——

4. おわりに

日本語の文型の単純化と、データの構造情報の知識を辞書として実装することによって、日本語文によるデータベース検索システムを実現した。名詞句を検索操作を組み立てる部品としてイメージできるため、検索文が考えやすいという評価を、社内の試使用ユーザから得ることができた。日本語解の速度も当初の目的を達成し、実時間での応答が可能になった。

今後の課題としては、適用分野の拡大、辞書作成の簡略化、日本語解析のより以上の高速化、簡単な文脈処理の実現などがあげられる。

また、今回、販売管理のパッケージの情報系処理の日本語対話システム化をおこなう上で、最も障害となったのは、パッケージが事務用言語で通常のファイル処理を中心として作成されているため、基幹データが正規化された関係データベースとして整備されていないという点だった。同一のデータが数カ所に存在するため、データの構造を再構成するために多くの時間が費やされた。今後、データ中心設計の技法が普及し、リポジトリを利用できるようになることを期待したい。

Anyoneは、現在、開発を完了し発売を開始している。今後、エンドユーザによる実使用の評価を得て、改良をすすめていきたい。

参考文献

1. 中所：エンドユーザコンピューティングソフトウェア危機回避のシナリオ、情報処理、Vol32, No8, 950-960(1991)
2. 長尾：自然言語処理技術の最近の動向、情報処理、Vol33, No7, 741-745(1992)