

日本語対話システムAnyoneにおける

5M-4

辞書構築

清水 英弘† 板橋 美子† 大塚 仁司‡

†三菱電機(株) コンピュータ製作所、 ‡三菱電機(株) 情報システム研究所

1. はじめに

日本語対話システムAnyoneはオフコン(MELCOM80)において、日本語を入力しデータベースを検索するシステムである。入力となる日本語を解析する際に、システムが参照する知識は、辞書として実装される。本論文では、辞書の構造および辞書の構築法(生成法)について述べる。本システムでは、利用分野(適用データベース)毎に容易に辞書の構築が行えるよう、辞書を分類し専用の構築ツールを用意した。特に、単語レベルでの概念の記述を必要としない構築(用語辞書の自動生成)を可能とした。

2. 辞書の構成

本システムにおける辞書は、解析のレベルにより以下の3種に分けられる。

- (1) 用語辞書 … 単語の形態素・構文情報を保持し、日本語解析の形態素解析(文献3)で参照する。
- (2) 意味辞書 … 言葉の意味的な情報(概念スキーマ:文献4)を保持し、日本語解析の意味解析および時間処理で参照する。
- (3) 手続辞書 … 検索のためのデータベース名や、グラフ表示のための情報などを保持し、検索構造解析やグラフ表示で参照する。

それぞれの辞書は、構築のレベルにより以下の3つの部分に分けられる。

- (a) 基本部 … 適用するデータベースや個人に依存しない、共通の部分。
- (b) データベース依存部 … 適用するデータベースに依存する部分。
- (c) ユーザ定義部 … 個別のユーザによって変更できる部分。

以上のように分類することによって、様々な分野のデータベースや個々のユーザに対応した辞書を構築することができる。

3. データベース依存部

本システムでは、入力された日本語文中の単語間の意味的な関連づけができるか否かを、それら各単語を組み合わせた検索ができるか否かという問題として扱う。その解析を行なうために、対象となるデータベースのデータ構造や言葉の意味的な情報を、意味辞書として実装している。辞書を構築する際、最も重要となるのは意味辞書であり、この部分を換えることにより、異なる分野のデータベースを検索することができるようになる。

Dictionary builder in "Anyone" a Japanese language interactive system

†Hidehiro SHIMIZU, †Yoshiko ITABASHI, ‡Hitoshi OHGASHI

†Computer Works, MITSUBISHI Electric Corporation.

‡Computer & Information Systems Lab., MITSUBISHI Electric Corporation.

データベースのデータ構造は概念スキーマネットワークとして意味辞書内に再構成してある。この概念スキーマネットワークでは、データベースのデータを、“マスタ”、“できごと”、“状態”、“集計”の4つのクラスに分けネットワークのノードとする。また、データベースのジョイン関係を基にネットワークの結合子を張る。

4. ユーザ定義部

個々のユーザによって異なる言葉や、検索ニーズに柔軟に対応できるように、上記4つのクラス以外に次の2種類について、意味辞書と手続辞書においてユーザが自由にマクロ定義できることとした。

- (1) 制約語 … “マスタ”のデータに対して条件をつけて検索するための言葉。
- (2) 連結語 … “できごと”や“集計”のデータを組み合わせて同時に検索し、特定のグラフを表示させるための言葉。

この他に、同義語の定義をユーザが追加できるものとした。この定義の中に送り仮名の揺れを定義しておく、関係するすべての単語に対してそれぞれの同義語を自動的に定義する処理を加えてある。また、ユーザの固有名詞データからある文字列(“株式会社”など)を削除して同義語として登録するための定義も、ユーザが追加・変更できるものとした。

5. 辞書構築の概要

以上のように分類、整理した辞書を構築するために、種々の構築用ツールを用意している。

意味辞書は、概念スキーマを表現する専用の記述言語により、データの4つのクラス毎にソースを作成し、それを専用のパーザを用いてコード化する。手続辞書も同様に、ソースを作成し、専用のパーザによりコード化する。そしてそれぞれのオブジェクトを専用のリンクにより結合し、辞書を生成する。この時、意味辞書の概念スキーマの情報から、検索のための用語(“できごと”に関する動詞やデータベースの項目名など)を自動的に用語辞書に登録する。さらに実際のデータベースの名称データから、固有名詞を自動的に用語辞書に登録する。よって、単語レベルでの概念記述を独立して作成する必要がない。

6. おわりに

以上、日本語対話システムAnyoneにおける辞書およびその構築について述べた。辞書を利用と作成両方の面から分類、整理し、容易に他分野や個人別に辞書を構築できるようにした。特に、単語レベルの概念記述を必要としない構築を可能とした。

参考文献

- 1) 板橋美子ほか：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，情報処理学会情報システム研究会資料，45-2(1993)。
- 2) 永松靖朗ほか：日本語対話システムAnyoneにおけるユーザーインタフェース，情報処理学会第47回全国大会，5M-01(1993)。
- 3) 大塚仁司ほか：日本語対話システムAnyoneにおける形態素・構文解析，情報処理学会第47回全国大会，5M-02(1993)。
- 4) 高梨郁子ほか：日本語対話システムAnyoneにおける意味解析，情報処理学会第47回全国大会，5M-03(1993)。