

## シソーラス評価用ソフトウェアの詳細設計

5P-3 甲斐 敏彦(1) 松田 政明(2) 小川 清(3)

大同工業大学(1,2) 名古屋市工業研究所(3)

## 1 はじめに

ソフトウェアの開発過程における用語の体系的な整理と、ソフトウェア開発過程における用語の体系的な理解に用いるシソーラスを管理するためのソフトを制作し、シソーラスのどのような形式でデータベースのフィールドに持つかを、実際に必要な機能を開発し評価することにより詳細設計し直すことにより実際の仕様を検討した。

## 2 基本仕様

任意の語との上下関係を規定すると用語の集合の中でループが発生する。また、階層構造が深くなり過ぎる可能性がある。目的としている

---

Design of software for evaluating thesaurus

Toshikiko Kai, Masaaki Matsuda, Kiyoshi Ogawa

Daido Institute of Technology, Nagoya Municipal Industrial Research Institute.

2-21 Daido-cho, Minami, Nagoya, Aichi 457 Japan

3-4-41 Rokuban, Atsuta, Nagoya, Aichi 456 Japan

シソーラスの構造は、特別な場合を除いて、5階層程度で収めることにした。プログラム上は中間作業におけるバッファを含めるために9階層までの登録が実現できるようにした。

任意の体系のシソーラスを評価するためのソフトウェアは、ループ構造、階層の深さの恣意性により、用語の理解にとっては必ずしも有効ではない。

しかし、現実社会を描写する上では、中間過程において、ループ構造が生じるものも処理する必要がある。

今回は、同じ用語を別々に登録することと、同一の用語は最終的には1カ所のみ正規化することにより、この問題を回避した。

## 3 階層構造化の設計

## 3.1 下位語

ある用語の下方への関係を情報として持つ方法を検討する。上位語は特定の一つの用語をフィールドに持っているの、ある用語が持っている下位語は特定の一つではない。つまり、ある用語が、位置情報として下位語のフィールドを持つことは、予想されるだけのフィールドを持たせる必要がある。これはデータベースとし

ては好ましくない。

### 3・2 上位語

情報と情報を一定の規則で関係を規定し、その関係によって目的の情報を検索する方法を検討した。

始めに上方への位置情報を規定する。任意の一つの上方への関係を上位語と表現すると、ある用語に特定の上位語のフィールドを設定すれば、関係を辿ることにより最も上位の用語のたどり着くことになる。そこで一番高いところにある用語を特定の用語の集まりの代表として見出し語と表現する。

### 3・3 分類コード

階層構造を表現する方法を記述する。

分類コードとして、順に大分類コード、中分類コード、小分類コード、細分類コードと4階層でデータを持つことを考える。各2桁をコードとして持ち、細分類コードは8桁となる。つまり8桁のコードの集まりを設定し、左から順に2桁ずつ増やしていく方法である。例として、12345678というコードは、大分類コード=12、中分類コード=1234、小分類コード=123456、細分類コード=12345678となる。

この方法では、各分類コードが99件以下の登録であることが前提になる。

### 3・4 次の用語の設定

階層構造を再構成するには、位置情報を持たせることによって、用語を登録する時点で本来の格納場所を確保でき、位置情報を用いてソートができる。

その方法は、リスト構造を用い、次の用語を示すポインタを持つことにある。前述の分類コードを新たに付加する方法に比べて、設定するデータを少なくできる。

## 4 フィールドの設計

ここまでで必要となるフィールドは、用語名、その用語を表すコード、その用語の階層、上位語のコード、次の用語を示すコードである。また、用語体系の評価をするためでもある。今回は、下位語の数、同意語の数を表示するフィールドを設けた。そのほかの評価方法に必要なフィールドは、随時追加できるような設計とした。

## 5 プログラミング

プロトタイプは、データベースソフトを用いて行い、簡単なプログラミングを付加した。用語の体系化が目的であるので、当面はこのプロトタイプで十分である。

## 6 今後の課題

評価用のソフトウェアの枠組みができあがったので、各種評価指標の計算部分を組み込むことにより、用語体系の評価を行っていくことが今後の課題である。