

シソーラス構築支援ツール：THEMSYSの開発

7N-6

田口 信一* 深瀬 仁* 今崎 光勇** 上野 晴樹*

*東京電機大学

**株式会社メイテック

1. はじめに

データベースは、高度情報化社会において非常に重要な位置にあり、様々な分野に導入されている。このデータベースの索引、検索するのに利用されるものとしてシソーラスがある。しかし、シソーラスを構築する上で幾つか問題点があるため、標準化されたシソーラスを構築することが困難である。よって、本稿では、主題分野について深く認識のある専門家が、シソーラスを構築する過程においてユーザをサポートするシステム、“THEMSYS”の開発について述べる。

2. シソーラス構築の際の問題点

シソーラス構築においては、JIS-X-0901-1991、ISO-2788-1986により定義されているが[1]、シソーラスを構築する際の問題点としてa)多大な時間、労力が掛かる。b)人手により構築されるため、一貫性に欠ける。c)データベースの成長に合わせた維持、管理がとて困難。d)シソーラスの専門家とシソーラスの網羅する分野の専門家を必要とする。などがあげられる。

よって、我々はユーザの負担を軽減し、ファセット分類を用いて効率よく及び統一された手順において、シソーラスを構築するツールの開発を目的とする。

3. シソーラスの構造及びモデリング

シソーラスを構築する際には必ず主題分野が存在し、シソーラスの適用範囲を明確にする必要があり、それから用語を収集する。収集された用語をファセット分類により分類をする。ファセット分類を用いて組織化された体系的シソーラスの編成は、用語が表している基本となるカテゴリの種類によって、用語をクラス及び集合に組織化していく。この時に階層の組織化の論理的な基準を示すため及び異なる種類に関連している概念の用語をまとめて示すために、ノードラベルを用いて表示する。ノードラベルの前後の用語は、概念的に関係の深い用語である。ファセット分類により組織化された用語の間には、階層関係（上位語-下位語関係、またはBT-NT関係）、連想関係（関連語関係、またはRT関係）、同義語関係（優先語-非優先語、またはUSE-UF関係）の関係が存在する。

また、シソーラスの階層構造を汎用フレームシステムZEROを用いて表現すると、図-1のようになる。

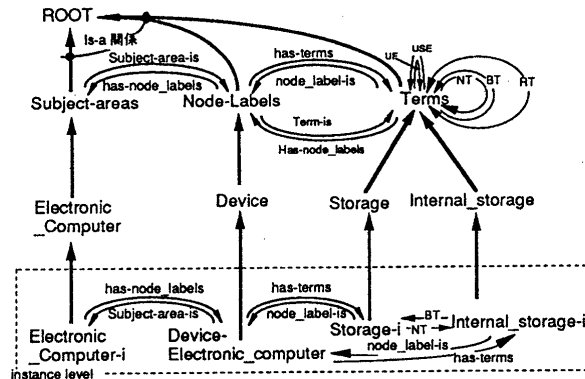


図-1 電子計算機技術における階層構造

また、主題分野、ノードラベル、用語の各フレームに含まれる主要なスロットを表-1、表-2、表-3に示す。ノードラベルフレームのCategoryスロットは、ノードラベルのカテゴリ値を記述するスロットであり、用語フレームのPropertyスロットは用語を説明する値を記述する。また各要素間の関係を示すスロットとして、has-node_labels, node_label-is, has-terms, term-isスロットがある。また用語フレームには階層関係を示すスロットとして、BT, NT, 連想関係を示すスロットとして、RT, 同義語関係を示すスロットとして、UF, USEの各スロットが設定されている。

表-1 SUBJECT_AREAS-FRAMEの主要スロット

属性名	型	属性値
A-KIND-OF	FRAME	抽象-具体関係の親フレームへのポインタ
DDESCENDANTS	FLIST	抽象-具体関係の子フレームへのポインタの集合
HAS-NODE-LABELS	FLIST	主題-ノードラベル関係の、主題フレームからファセットフレームへのポインタの集合
NT	FLIST	主題-用語関係の、主題フレームから下位用語フレームへのポインタの集合
RT	FLIST	主題-用語関係の、連想関係を持つ用語フレームへのポインタの集合

表-2 NODE_LABELS-FRAMEの主要スロット

属性名	型	属性値
A-KIND-OF	FRAME	抽象-具体関係の親フレームへのポインタ
DDESCENDANTS	FLIST	抽象-具体関係の子フレームへのポインタの集合
SUBJECT_AREA-IS	FLIST	主題-ノードラベル関係の、ノードラベルフレームから主題フレームへのポインタ
HAS-TERMS	FLIST	ノードラベル-用語関係の、ノードラベルフレームから用語フレームへのポインタの集合
TERM-IS	FLIST	ノードラベル-用語関係の、ノードラベルフレームから用語フレームへのポインタの集合
CATEGORY	STRING	ノードラベルがどの基本カテゴリに属するかを示すスロット

Development of Thesaurus Management System:THEMSYS

Shinichi TAGUCHI*, Hitoshi FUKASE*, Mitsuyu IMAZAKI**, Haruki UENO*

*Tokyo Denki University **Meitec Corporation

表-3 TERMS-FRAMEの主要スロット

属性名	型	属性値
A-KIND-OF	FRAME	抽象-具体関係の親フレームへのポインタ
DDESCENDANTS	FLIST	抽象-具体関係の子フレームへのポインタの集合
NODE-LABEL-IS	FLIST	ノードラベル-用語関係の、用語フレームからノードラベルフレームへのポインタ
HAS-NODE-LABELS	FLIST	用語-ノードラベル関係の、用語フレームからノードラベルフレームへのポインタの集合
SN	STRING	用語についての、スコープノート
UF	FLIST	非優先語(用語)フレームへのポインタの集合
USE	FRAME	優先語(用語)フレームへのポインタ
BT	FLIST	上位語(用語)フレームへのポインタの集合
NT	FLIST	下位語(用語)フレームへのポインタの集合
RT	FLIST	連想関係にある用語フレームへのポインタの集合
PROPERTY	STRING LIST	その用語に対する属性(キーワード)の集合

4. シソーラス構築の手順

4. 1 シソーラス構築の手順

シソーラスを構築する手順としては

- a) 主題分野の選定, 及び用語収集
- b) ノードラベル設定
- c) 用語の分類, 関係付け
- d) シソーラスの表示

である[2][3].

我々が支援するのは, c)の, 用語をノードラベルにより分類し, 用語間の関係付けを行う部分である. すでに, 主題分野の網羅すべき領域について, 用語を収集して, いくつかの基本カテゴリを発見した上で, 基本カテゴリに基づきノードラベルを設定しているとする. 用語の分類については, 用語の概念を把握した上でその概念を的確に表現しているノードラベルが存在するかどうかを判断する. そして分類した上で同じノードラベルの中に階層関係となる用語があるかどうかを調べる.

4. 2 分類, 及び階層付けの手段

分類, 階層付けは, ノードラベル, 用語を設定するとき同時に定義したカテゴリ値, 及び属性値により行なう. 図-2に用語を分類する方法を示す. ノードラベルに分類済みの用語の属性値により候補を探す.

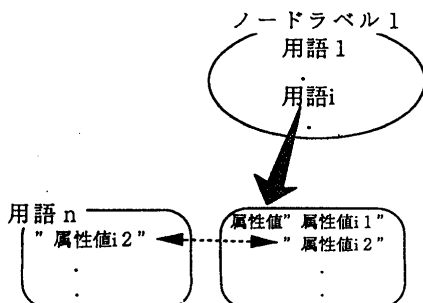


図-2 ノードラベルの候補の見分け方

また, 上位語, 下位語を探す手段として図-3に示す. 各用語フレームの属性値を参照し関係のある用語を探す. 用語を体系化した数が増すに従い, 過去の事例が増えるために, 用語に対するノードラベルを絞り込むことが容易となる. 同様に, 階層化された用語が増えるに従い, 概念の同じまたは近い用語の集合ができるので, 見つけるのが容易になる.

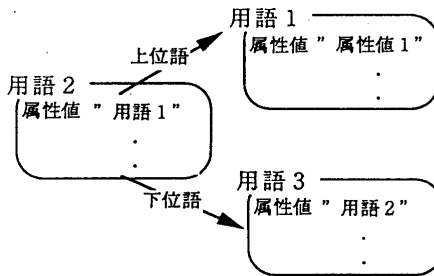


図-3 上位語, 下位語の見分け方

5. 具体例

主題分野 "電子計算機技術" のいくつかあるノードラベルの一つ "デバイス" に用語 "記憶装置" が分類されていると, このノードラベルには用語 "記憶装置" の属性値 "記憶装置" "装置" が割り当てられるため, 用語 "内部記憶装置" を分類するときには属性値 "内部記憶" "記憶装置" の一つ "記憶装置" によりノードラベルが "デバイス" であることが推測されユーザに示される. また, 分類された後 "デバイス" 内において用語 "記憶装置" が用語 "内部記憶装置" の属性値の一つと一致するため, 用語 "記憶装置" の下位語として "内部記憶装置" が推測され示される.

6. まとめ

本稿では, シソーラスを構築する過程を支援するシステムの開発, 及びその方法について述べた. 本システムを用いることにより, シソーラスを構築する過程に於てユーザの負担を軽減することが可能である. 今後の課題として, より細部への分類について研究を進めていく予定である.

参考文献

- [1] シソーラスの構成及びその作成方法 X 0901-1991, 日本工業規格
- [2] 内藤, 中倉 良夫, 影浦 峯他訳: シソーラス構築法, 丸善株式会社
- [3] 川村 敬・訳: シソーラスの構築と維持-生き残りに必要な知的作業の指針, 情報管理学会誌, Vol133, No.6, sep.1990