

概念分類項目間の関係記述に基づく概念体系の試作

7S-8

三池誠司 岸本行生 松川智義 横田英司 高井貞治 天野真家

(株)日本電子化辞書研究所

1. はじめに

日本電子化辞書研究所では、自然言語処理を目的とした概念体系の開発を行っている[1]。本稿では、概念分類項目の関係記述[3]に基づいた概念体系の試作について報告する。

本試作では、2項関係抽象化支援システム[4]の階層関係抽出機能を用いた。この機能により構築される階層構造は、概念記述を演繹的に推論できることを保証する。また、推論可能な概念記述の削除により、概念記述のデータ量の高い圧縮率が期待できる。

2節で概念体系試作の方法について述べ、3節に試作の結果を示す。4節でノードと項目について考察する。

2. 概念体系試作の方法

次の手順で概念体系を試作した。(1)と(2)は、階層関係抽出機能により、自動的に実現した。(3)は、人手で行った。

(1) グループ化处理

概念分類項目間の概念記述を、こと概念に属する概念分類項目(軸の組み合わせ)と関係子の組(図1のマトリクスの行側)と、相手の概念分類項目(図1のマトリクスの列側)からなる2項データとみる。図1で、マトリクス中の1は、その組み合わせの記述が存在したことを示す。このデータから、グループ(組み合わせがすべて1となるもの)を抽出する。なお、マトリクスの列側の数字は、項目の識別番号である。

GROUP_70 (35, 21)	1 1 1 1 1 1 1 1 1 1
	1 1 1 1 1 1 1 1 1 1
	1 1 1 1 1 1 1 5 6 5 6
	1 1 1 1 1 1 1 1 1 1
	1 2 2 2 3 組 1 1 1
	人 2 1 2 3 1 線 1 2 3
	問 // // // // // // //
	仲 男 女 子 職 地 血 階
	問 問 供 業 名 緑 緑 級
	組 組 組 組 組 組 組 組
	続 続 続 続 続 続 続 続
所有: 状態 + a g t	1 1 1 1 1 1 1 1 1 1
機能(役割): 状態 + a g t	1 1 1 1 1 1 1 1 1 1
属性(階級): 順序 + a g t	1 1 1 1 1 1 1 1 1 1

図1. グループ化の例

ここで、グループの行側とグループの列側に対応する概念体系のノードを設定する。それぞれを行ノードと列ノード

Concept classification based on the relations  
between concept categories

Seiji MIIKE, Yukio KISHIMOTO, Tomoyoshi MATSUKAWA,

Eiji YOKOTA, Sadaharu TAKAI, and Shin-ya AMANO

Japan Electronic Dictionary Research Institute, Ltd.

ドとよぶ。項目を行ノードまたは列ノードとリンクし、行ノードと列ノードを記述に用いられた関係子で結ぶ。この模式を図2に示す。四角は項目を意味し、丸はノードを意味する。ノード内の数字は図2と3のノードの対応を示す。実線矢印が、項目とノード間のリンクを示し、点線矢印が関係子を示す。

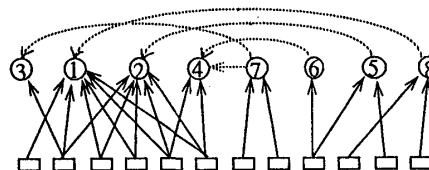


図2. 体系のノードの設定

(2) 階層化处理

各ノードに属する項目の包含関係により、含むものを上位ノード、含まれるものを下位ノードとして、階層化する。図3において、丸と丸を結ぶ矢印が、ノードの上位下位アークである。このとき、上位ノードから、下位ノードに属する項目を削除する。

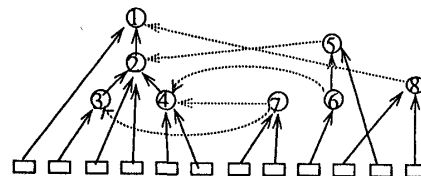


図3. ノードの階層化

(3) ノードの交わり関係に基づく下位ノードの付加  
共通の項目をもつノードが存在する場合、新たなノードを設定し、そこへ共通の項目を移す。

3. 概念体系の試作

得られた概念体系についての定量データを以下に示す。概念分類項目間の概念記述数は47,571であり、グループ化によって次の結果を得た。

グループ化された概念記述の数	44,184
グループ化されなかった概念記述の数	3,387
グループの数	78

階層化と下位ノードの付加により、次の結果を得た。なお、グループ化処理直後でのノードと項目のリンクの総数は5,651であり、上位ノードから下位ノードの項目を削除および移動することにより4,348に減少している。

ノード間の関係子アーク	78
ノード間の上位下位アーク	127
ノードと項目のリンクの総数	4,348
合計	4,553

概念記述のデータ量の圧縮率は、次のとおりである。

$$(47,571 - (4,553 + 3,387)) / 47,571 = 83.3\%$$

なお、リンクされている項目がすべて一致するノードは、行ノードで2個（異なり数1）、列ノードで25個（異なり数11）であった。重複が存在することは、そのノードに対応する上位概念のふるまいが多様であると解釈できる。もの概念にそのような上位概念が多いことがわかった。ノードの階層構造は、次のように示される。

	行ノード	列ノード
階層化されたノードの数	59	83
1ノードから出るアークの最大数	6	6
1ノードへ入るアークの最大数	5	9
階層の最大の深さ	4	9

表からわかるように、特に、行ノードではフラットな構造である。これは、こと概念では、上位概念の上位下位関係が比較的浅いことを意味している。

また、ノードにリンクされている項目の数は次のとおりである。

	行ノード	列ノード
属する項目の平均の数	3	20
属する項目の最大の数	38	185

#### 4. 考察

図4にノードの一部を示す。ノード1では、ノード4との関係が成立し、ノード2では、ノード4、5、6、7の関係、ノード3では、これらとノード8との関係が成立する。ノード1は、ノード1、2、3に属する項目に共通する属性をもつ上位概念であり、ノード2は、ノード2、3に属する項目に共通する属性をもつ上位概念である。

一方、概念分類におけるこれらの項目の分類作業上の上下関係は、図1の列側の項目の番号の桁数に示されている。これと、ノード1、2、3を比較すると、上位概念を導き出す目的で設定した概念分類項目の上下関係と、項目間の関係記述によって得た上位概念の上位下位関係とが一致しない部分が見える。分類項目上では上下関係に無い「人間」と「組織」が、共通の属性をもつ項目の集合の要素としてノード1に現れている。すなわち、ノード1は、特定のこと概念（ノード4）との関係からみて、概念分類上の上下関係になかった項目に属する概念のふるまいが等しいことを示している。

また一方、今後ノードの検証を行えば、例えば、図5のように、改良を行う余地がある。これは、上述のノード1に表されている上位概念を残し、新たに「組織」などの分類項目に対応する上位概念を設けることである。これによって、記述データの量は逆に増えることになるが、計算機利用や辞書のメンテナンスの点からこのような改良が有効

である可能性が予想される。この検証、改良が今後の課題である。



図4. 概念体系の一部

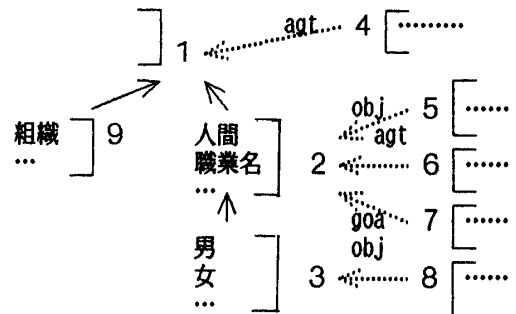


図5. 概念体系の改良の例

#### 5. 結論

概念分類項目間の関係記述を基に、2項関係抽象化支援システムの階層関係抽出機能を用いて、概念体系の試作を行った。この方法によれば、概念記述のデータ量の高い圧縮を実現する階層構造を構築することができる。また、考察で述べたように、内省によるノードの検証、改良を行う余地がある。

今後、2項関係抽象化支援システムの編集機能による方法を含め他の方法による体系構築の検討と、構築された概念体系の、実例から得た概念記述データによる検証、改良を行う。

#### 【謝辞】

本研究の機会を与えて頂いた横井所長、並びに貴重な意見・コメントを頂いた研究員の皆様に感謝いたします。

#### 【参考文献】

- [1] 高井他, 概念体系の開発思想, 情報処理学会第41回全国大会, 1990
- [2] 岸本他, 概念分類項目の設定, 情報処理学会第41回全国大会, 1990
- [3] 横田他, 概念分類項目間の関係記述, 情報処理学会第41回全国大会, 1990
- [4] 松川他, 2項関係抽象化支援システムの基本機能, 情報処理学会 第41回全国大会, 1990
- [5] 松川他, 記述量圧縮の観点からみた概念体系の構築, 情報処理学会NL-78, 1990