

意味関係記述のための概念構造モデル

藤原 譲、 劉 野、 頼 静絹

{fujiwara,liuye,rai}@dblab.is.tsukuba.ac.jp

筑波大学 電子・情報工学系

〒305 茨城県つくば市天王台1-1

要旨

言語の豊かつ多様な意味を理解するためには概念間の意味関係を十分に記述することが必要である。しかし、検索用のシソーラスでは表現形式が単純で、意味関係が十分記述、表現できない。そのため、意味関係の自動抽出及び概念構造の自己組織化の研究も不可欠である。本研究では自動抽出された断片的な意味関係と既存のシソーラスにおける意味関係との整合性を考慮した用語構造の統合について研究を行なった。統合後の意味関係においての重なり、多項関係、相対性、動的関係などに対応するために、均質化二部グラフモデルに基づき、概念構造を記述する。現在、超伝導などの先端科学技術分野を対象とした概念構造自己組織化システムを開発している。それに基づき、情報の蓄積、管理、検索における意味処理だけではなく、類推、帰納推論、仮説推論、発想、連想などの高度な応用が可能となる。

キーワード：意味関係解析、概念構造モデル、意味記述、意味理解、概念記憶構造、自己組織化

The Conceptual Structure Model for Description of Semantic Meaning

Yuzuru Fujiwara, Ye Liu and Jingjuan Lai

{fujiwara,liuye,rai}@dblab.is.tsukuba.ac.jp

Institute of Electronics and Information Science, University of Tsukuba

1-1-1 Tennodai, Tsukuba, Ibaraki 305, Japan

Abstract

Semantic understanding of language requires to describe sufficient semantic relationships among concepts to construct conceptual structures carrying various and rich semantic relationships. The conceptual structures include multi-hierarchical and various relationships, which are much more complex than usual thesauri. A homogenized bipartite model is used to describe the conceptual structures, which consist of diversified types of edges and nodes. The self-organized conceptual structure for superconducting materials has been developed by integrating fragmentary semantic relationships extracted automatically from technical terms and hierarchical relationships of root-thesaurus. The automatic integrating system includes broad term analysis and narrow term element analysis and consistency verification. The interface was developed to visualize the conceptual structure. The resulted conceptual structure may be used for advanced functions e.g. analogical reasoning, induction, abduction, creation and association as well as information retrieval and data management.

Keyword: analysis of semantic relationships, conceptual structure model, description of meaning, semantic understanding, conceptual memory structure, selforganization.

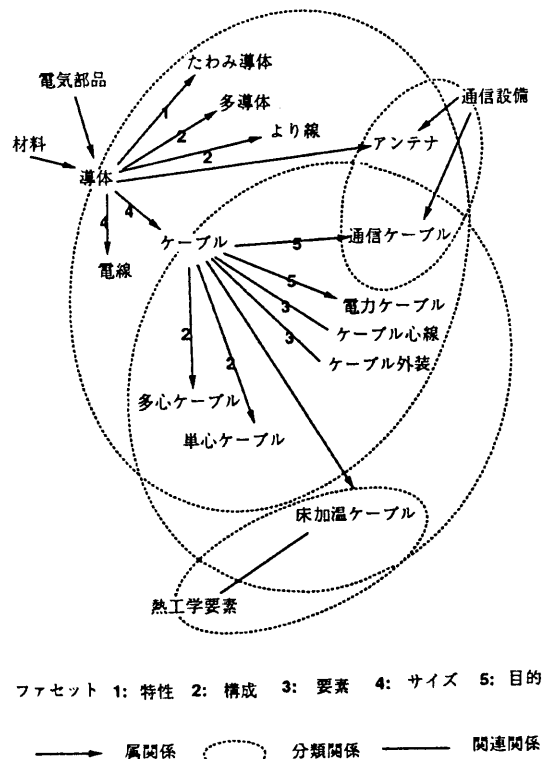
1 はじめに

思考支援のためには意味の理解・処理が必要であり、そのためには、多様かつ豊富な意味関係を含む概念構造の自己組織化が有効な手段である。従来のいわゆる「意味処理」は主として情報の検索などの応用にとどまり、意味関係としてはシソーラスにより階層関係、関連関係、同義関係のみを対象とした。一方、自己組織化型情報ベースシステムの開発には、情報の蓄積、管理、検索だけでなく、類推、帰納推論、仮説推論、発想、連想などの意味処理も対象としている。このような応用目的に対して、従来のシソーラスは意味関係の内容及び表現形式において様々な欠点が指摘されている。とくに、検索用シソーラスでは表現形式が単純で、意味関係の明確さ、詳細度が不充分である。そこで本研究は多様な目的に対応できる概念構造の構築を目的とする。とくに、概念構造は充分な意味関係を含み、その表現形式及び概念構造の自動的な構築方法についても述べる。

2 概念構造の構成

まず多様な目的に対応する概念構造の意味関係の内容及び表現形式について述べる。[1][2]
多様な階層関係

概念構造の主たるものの一つは階層関係である。専門用語の階層関係には、事物所属の視点か



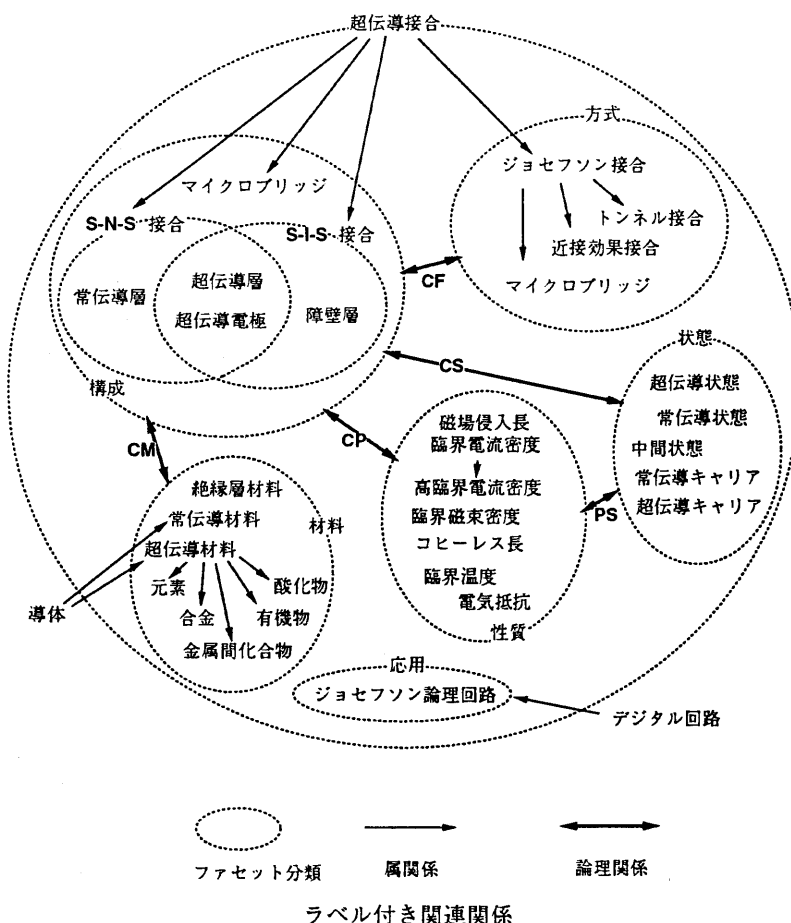
二つの階層関係を含む概念体系

ら形成する属階層関係と主題分野の視点から形成する分類階層関係があり、従来、それぞれの階層の関係によって互いに独立した概念体系が構成される。一方、意味関係をできるだけ全面的かつ明確に表示するために、二つの階層関係を含む概念体系が必要である [6]。図 1 ではファセット分類に基づいた階層関係と属関係に基づいた階層関係の例を示す。

従来のシソーラスが木構造の階層関係しか表現できないのに対して、この概念体系では複合階層が表現される。例えば、「アンテナ」が同時に「導体」と「通信設備」のサブファセットであるように、ある概念が複数のファセットに属することができる。また、「導体」が同時に「電気部品」と「材料」の下位語であるから、ある概念が複数の上位語を持つことが許される。このように、本概念構造が概念階層関係を充分に表示でき、同時に二つの階層関係の冗長さを解消できることがわかる。

関連関係

従来のシソーラスには階層関係以外の意味関係を全て単に関連関係として記述する。因果関係など論理関係の詳細な記述は類推、帰納推論、仮説推論などの応用にとって不可欠である。この



ため、概念構造には関連特徴を記述するラベル付きの関連関係を構築する。図 2 の例では過程 -

状態、実体-特徴などのラベル付け関連関係を示す。この中に概念構造の多項関係の表現例として、集合間の関連関係を示している。これらの関連関係を用いて論理関係の表現ができる。例えば、PS(property and state)によって、Value(SP (electric resistance, superconducting state), 0) という事実が表現できる。

均質化二部グラフモデルによる意味関係の記述

図1と図2の概念構造は部分的重なり、多項関係、相対性、動的関係などを含み、グラフの構造では対応できない[1-5]。そこで、ハイパーグラフ双対性を更に拡張した均質化二部グラフモデルに基づき、記述する[1][2]。モデルの定義は次の公式で述べる。

V、E、Lは頂点、辺、ラベルである。

$$E \subseteq 2^V \quad (1)$$

$$V = V \cup E \quad (2)$$

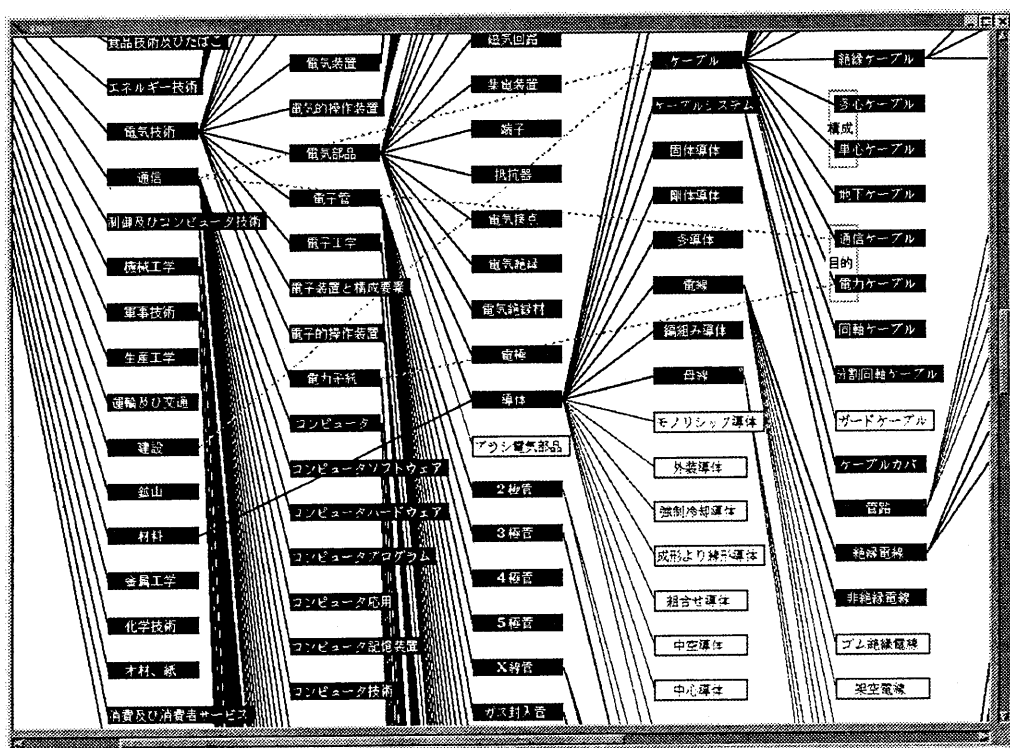
$$E = E \cup V \quad (3)$$

$$\sigma: L \rightarrow E \cup V \quad (4)$$

式(2)と(3)で再帰と内部構造が許される。

3 概念構造の自動構築実験

概念構造自己組織化システムの一つの機能として、ここでは超伝導分野を対象とした用語構造の統合を紹介する。この統合は、自動抽出された断片的な意味関係と既存のシソーラスにおける意味関係との整合性を考慮しながら行なわれている。



今度の実験はまず二つの階層関係を含む概念体系を作成する。一つの階層関係はルートシソーラス [7] のファセット分類階層関係を利用して作成し、もう一つの階層関係は以前に提案した SS-KWIC[8] という階層関係抽出方法に基づいて属階層関係として抽出された。さらに、この二つの階層関係を統合して、概念体系を構築する。システムは上位語分析、下位語要素分析、整合性調整などの機能を持つ。図3では概念体系の一つの結果例が示されている。

二つの階層関係を統合する時、体系階層間の干渉性が起こる [3][9]。このため、我々は整合性調整アルゴリズムも考案している。次にその一例を示す。「多重通信」について、SS-KWICによって

「通信」>「多重通信」(“>”は上下位関係を表す)

という関係が抽出される。また、ルートシソーラスの分類階層関係から

「通信」>「通信技術」>「通信伝送方式」>「多重通信」

という構造がある。同じ分野中の構造であるから、

「通信」>「通信技術」>「通信伝送方式」>「多重通信」

というより深い階層構造が保留され、冗長な関係「通信」>「多重通信」は削除される

一方、「経営科学」について、

「科学」>「経営科学」

という構造がSS-KWICによって、抽出されるが、

ルートシソーラス分類階層は

「管理科学」>「経営科学」という構造がある。

それぞれの関係の指す分野が異なるので、統合後に両方とも保留して、その間に関連関係を作る。

以上の手法を組合せて、自己組織的に概念構造を構築して、それを高分子のNMRスペクトル

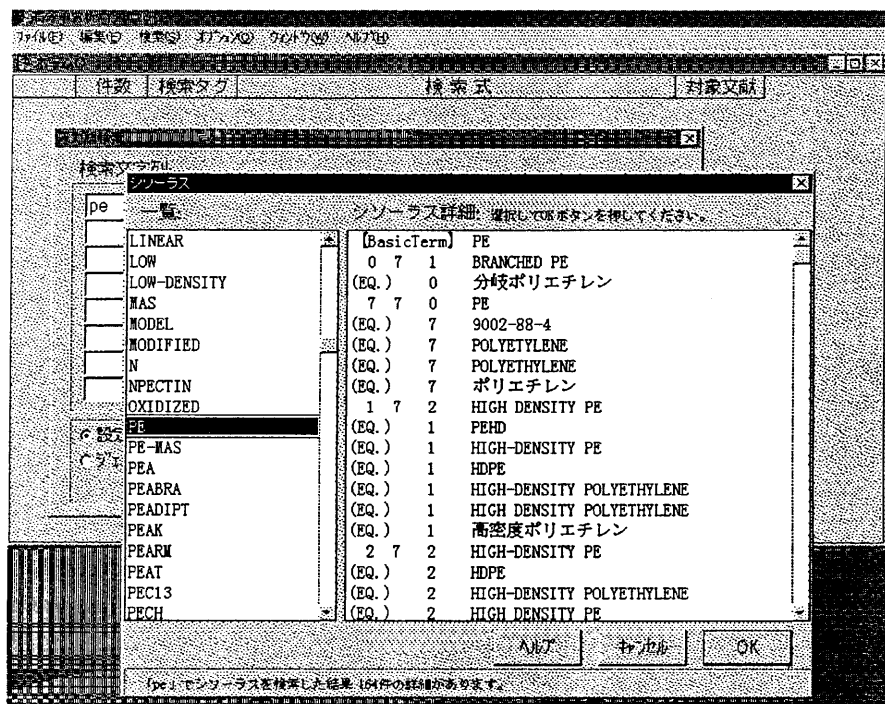


図4 高分子のNMRシステムにおける概念構造情報の例

の解析に応用するためのシステムの例を図4に示す。概念構造部分は主としてC-TRANで同値関係を抽出し、SS-KWICで階層関係と関連関係を抽出[8]したりので、多重継承を含む構造になっている。

4 おわりに

多様な目的に対応できる概念構造の構成及び自己組織的構築方法について述べた。概念構造の構成の中に、多様な階層関係、ラベル付き関連関係を組み入れたが、まだ内部構造、相対性などの複雑な問題に対処しなければならない。また、今後の課題として、論理関係を含む関連関係の構築や、類推、帰納推論、仮説推論、発想、連想などの意味処理への応用方法[1, 2, 10]などが挙げられる。

参考文献

- [1] Yuzuru Fujiwara: "The Model for Self Structured Semantic Relationships of Information and Its Advanced Utilization". Proceedings of The 47th FID Conference and Congress, Oct. 1994.
- [2] Yuzuru Fujiwara and Hironobu Gotoda: "Representation Model for Relativity of Concepts". International Forum on Information and Documentation, vol. 20, 1, PP 22-30, January 1995.
- [3] Fugmann, Robert: "Theoretische Grundlagen der Indexierungspraxis". 1992.
- [4] C.Berge. "Hypergraphs". North-Holland, 1989.
- [5] H.Boley. "Directed Recursive Labelnode Hypergraphs : A New Representation-Language". Artificial Intelligence, 9(1):49-85, 1997.
- [6] Y.Liu: "The Research and Test of Combination Schedule of Term Family with Classification in Thesaurus". Journal of Information Science, Vol.11, No2, PP 60-69, 1990. (in Chinese)
- [7] British Standard Institution. ROOT thesaurus, 3rd ed., MATCHING LIST. 1988. (Japanese version, 1990).
- [8] J.Lai, H.Chen and Yuzuru Fujiwara: "Extraction of Semantic Relationships Among Term - SS-KWIC". Proceedings of The 47th FID Conference and Congress, Oct. 1994.
- [9] Peter Ingwersen: "Information Retrieval Interaction". 1992.
- [10] Hironobu Gotoda, Jianghong An, and Yuzuru Fujiwara: "Analogical Reasoning of Organic Reactions Based on the Structured Compound-Reaction Diagram". International Conference on Forum and Applied Practical Reasoning, FAPR'96, PP 276-290, Germany, June 1996.