

知識ナビゲーションのための概念図の自動生成

村山 正司[†] 中村 裕一^{†‡} 大田 友一[†]

[†]筑波大学 機能工学系

〒 305-8573 つくば市天王台 1-1-1

E-Mail:murayama@image.esys.tsukuba.ac.jp

[‡] 科学技術振興事業団, さきがけ研究 2 1

あらまし: 文章は情報伝達のための最も基本的なメディアであるが、ボトムアップ型の理解プロセスを情報の受け手に強制する問題を持っている。本研究ではトップダウン型の理解を促進する図的メディアを文章と相補的に用いて情報の伝達・理解を容易にし、人間の知的活動支援を行なう「知識ナビゲーション」を目的とする。本稿では、そのアプローチとして、文章の意味的構造情報から概念図を半自動生成する手法について述べる。またその手法を計算機上に実装した結果を報告する。

キーワード: 知識ナビゲーション, 概念図生成, ハイパーメディア, 知識表現, タグ付言語

Automatic diagram generation for knowledge navigation

Masashi Murayama[†] Yuichi Nakamura^{†‡} Yuichi Ohta[†]

[†] Institute of Engineering Mechanics and Systems, University of Tsukuba

1-1-1 Tennoudai, Tsukuba, 305-8573, Japan

murayama@image.esys.tsukuba.ac.jp

[‡] PRESTO, Japan Science and Technology Corporation(JST)

Abstract: We often need diagrams for explanation, though a text is the most powerful and effective tools for communication by documents. For this purpose, we propose a novel scheme for diagram generation, in which the semantic structure of a text is effectively translated and linked to the text. This aims at “knowledge navigation” which supports human communication via documents.

In this paper, we first describe the characteristics of diagrammatic expressions, and propose several useful categories of expressions. Then, we show the correspondence between diagrams and texts, and our framework for automatic diagram generation. Our preliminary experiments shows the feasibility of “knowledge navigation” by diagrams.

key words: knowledge navigation, diagram generation, hypermedia, knowledge representation, tagged language

1 はじめに

知識ナビゲーションとは、大量の知識の収集・蓄積・流通・提示を計算機によって効率的に自動化することにより、理解支援・伝達支援を行うことである。そこでは複数の表現形態を用いて情報や知識を人間にとってわかりやすく表現することが必要となる。

ここで考えなければならない重要なこととして、複数のメディア間には理解形態の相違が存在することが挙げられる。トップダウン型の理解では、まず表現対象のコンテンツの概要を直観的に把握し、その後細部の理解へと進む。一方、ボトムアップ型の理解では、表現対象のコンテンツの細部から階層を上にとどる形で理解が進む。

現在最も広く用いられているメディアである文章はボトムアップ型理解を必要とするが、そこには次のようないくつかの問題点が挙げられている。

- 一次元的な表現制約を受ける。
- 物理的構造・論理的構造を直観的に表現しにくい。
- 重要な単語でも、表面的には他の単語と同レベルで表現されてしまい、概念の重要性を表現しにくい。

一方、図はトップダウン型理解を促進する。そのため、以下に挙げるようなメリットが生じる。

- 言語における一次元の表現制約にとらわれない。
- 焦点をあてるべき意味要素や重要点を簡単に強調できる。
- 種々の構造を簡単に提示できる。

しかし、図で表現できる知識や構造だけでは自然言語の持つ多様な意味を表現できない。

つまり単一メディアのみでは伝達効率が悪く、大量の知識を扱う際に問題となる。そこで、複数のメディアを利用することで、単一メディアのもつ問題を相互補完することが望まれる。また、メディア間にハイパーリンクを付加したハイパーメディアはさらに有効である [1][2][3]。

本稿では図的メディアと文章メディアによって構成されるハイパーメディアを構築する方法として、文章の意味構造から概念図を半自動的に生成する手法について述べる。その手法では、概念図と文章

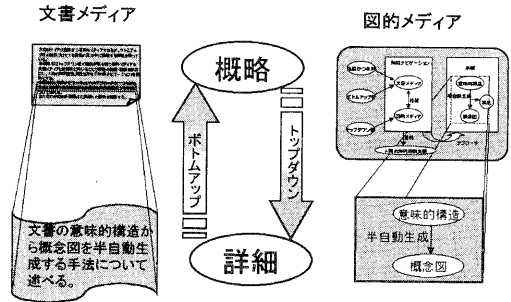


図 1: メディアの理解

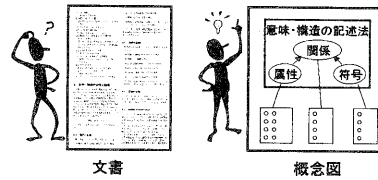


図 2: 概要把握の容易さ

を同時に提示するだけでなく、双方のメディアをハイパーリンクにより有機的に統合できる。これにより、概要把握と詳細理解の双方を容易にしたハイパーメディアの実現を目指す。

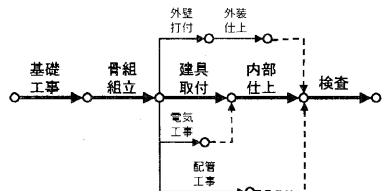
関連する研究には KJ 法支援ツール等の発想支援システムがある [4]。それらの目的はテキストオブジェクトの配置などによる人間の問題解決プロセスの支援である。それに対して、本稿の目的とするものは文書等の内容を効果的に伝えるためのコミュニケーション支援ツールであり、副次的に自分の発想を確認することによる発想支援を目指す。

以下本稿では、まず知識を表現するための概念図について述べる。そして文章の持つ構造と概念図を対応付けて、文章中の知識を概念図で表現する手法を提案する。また、本手法を計算機に実装して概念図生成実験を行う。

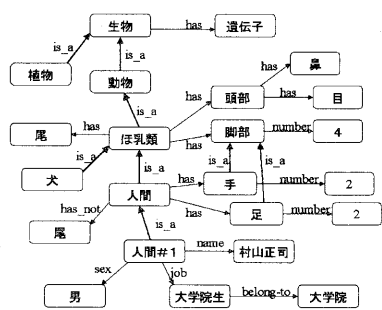
2 知識を表現するための概念図

2.1 図的メディアを用いた知識の提示

文書メディアと図的メディアを相補的に利用することにより、表現の自由度が増し、読み手の理解度が格段に向上する。しかし、図的メディアの質の良し悪しによって概略把握の容易さ、理解効果の向上の度合いが決まる [5]。



(a) 良い図の例 (概略が掴める)



(b) 悪い図の例 (焦点が判らない)

図 3: 図の例

良い図の例、つまり図的メディアの使用でわかりやすくなる例として、手順書などの図化などが挙げられる。例えば図 3(a) では重要な手順が強調され、クリティカルパスを把握することができる。

悪い図の例、つまり図的メディアの使用効果が小さい例として、まとまりのない単語の羅列や強調要素が存在しないような見にくい図化が挙げられる。図 3(b) のような意味ネットワークは、計算機内部の知識表現としては有用であるが、人間が見るための表現としては繁雑となる。種々の概念がまとまりなく混雑するため、送り手の伝えたい事柄の焦点がわかりにくいからである。

本研究では上記の問題を考慮し、良い図となるように文書の内容を図化する手法を提案する。そのために本研究では典型的な図形構造により表現できる意味的構造と文書で表現される内容との対応関係を調査した。その結果より、文書の意味的構造から図へ変換する変換手続きを設定した。

また、自動生成される図では人間の主観にそぐわない場合がある。そこで人手による編集操作を可能にし、またその編集操作の支援を行う編集ルールを同時に設定した。そうした人間とシステムのインタラクションを通じて、より良い概念図の生成を行う。

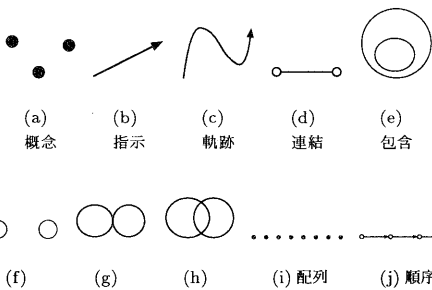


図 4: 図の規約的意味

2.2 意味構造を持つ概念図

図の構成には次に挙げる3つの機能が存在する [6]。

美的機能: 図をきれいに美しく整える。

認知的機能: 図を見やすくする。

記号的機能: 図で伝えたい意味内容が正しく伝わるように判りやすくする。

このうち、図的メディアを知識の提示に用いるためには、記号的機能が特に重要である。つまり、図には図形要素とその配置に関して規約的意味を持つものがあり、そこには言語構造に近い意味的構造が存在する。その例を図 2.2 に挙げる。

- 矢線は、方向を示し、また両端にあるものの関連性を表わす。(図 4(b))
- 線は、両端にあるものを連結する。(図 4(d))
- 線は、領域を分割することにより、概念の分割を表わす。(図 4(e)~図 4(h))
- 図上の領域包含は、包含・従属を意味する。(図 4(e))
- 図形要素が一列に並んでいれば、配列や順序、流れの意味を持つ。(図 4(i)、図 4(j))

以上のように、図で表わすと効果的に伝わる概念、概念間の構造がある。このうち階層的・包含的な構造については領域図により表現することができ、順序的・系列的な構造は連結図により表現することができる。

以下、本研究では、連結図と領域図を生成対象の概念図として扱う。



図 5: 概念図

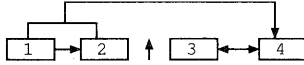


図 6: 「糸屋の娘」の文脈構造

3 文書と概念図の対応

3.1 文書中の意味的構造の分類

本研究で対象とする文書中の意味的構造には、大きく分けて文内構造、文脈・文章構造の2つがある。

文内構造は単語や語句間に存在する構造で、動詞の語彙的な意味構造や格構造、並列句などがある。例えば次のような構造がある。

- 概念の類似性や同一性
例) A、B、C等が挙げられる。
- 論理的・物理的な関係や構造
例) AとはBの一種である。
例) A、B、Cと順に行う。
例) AからBにCが送られる。

文脈・文章構造とは文や段落、文章間に存在する構造である [7]。例えば典型的な起承転結の構成を持つ次の文章の構造について述べる。

1. 大阪本町糸屋の娘。
2. 姉は十六、妹は十五。
3. 諸国大名は弓矢で殺す。
4. 糸屋の娘は目で殺す。

ここで第1文と第2文の間には展開の関係が存在し、第3文と第4文には対比の関係が存在する。また、第2文と第3文の間には話題転換の関係があり、第1文および第2文と第4文の間に展開の関係がある。このような構造を単純な図で説明すると図6のようになる。

以上に述べた意味的構造の代表的なものを挙げると、展開、推移、並列、対比、因果、補足、同格、修飾、階層などがある。しかしそれらを含む意味的

表 1: 4種類の関係

順序関係	順序、系列、軸などの関係
包含関係	階層や上下関係を表わす関係
同値関係	同値・同等、並列を表わす関係
修飾関係	説明や属性を付加する関係

表 2: 主な関係属性

時間	時間軸、時系列を示す属性
因果	原因・理由を示す属性
空間	実空間での物理的な属性
入出力	過程や原材料と生成物の関係を示す属性
話題	話題の流れ上にある属性
集合	組織や集合論上の関係を示す属性

構造全体は非常に多様であり、汎用的な表記と分類が望まれる。そこで本稿では、表1のように意味的構造全体を大きく4種類の「関係」に分類した。

さらにこれらの関係を表2のような「属性」パラメータを付加した。

また、対比などといった構造を表現するために、「符号」パラメータを使用する。通常の符号は「正」であるが、対比を表現する場合は「非」が使用される。

この枠組みにより、表3に示すように意味的構造を統一的に分類し表現できる。意味的構造の多様性には、「属性」パラメータの種類を増やすことで対応できる。

3.2 文書と概念図の意味的対応

意味を保存しつつ文書メディアから概念図を生成するためには、その二つのメディア間の意味的対応を保持する必要がある。本研究では、要素同士の対応と構造同士の対応をとる。

構成要素の対応

文書メディアと図的メディアの意味的な要素同士を対応させる。扱う構造により、次のように対応できる。

文内構造 単語・語句と図形要素が対応

文脈構造 文・段落と図形要素が対応

表 3: 主な意味的構造の分類と表現

意味的構造の表現	一般的表現	例文
順序：話題：正	展開	AだからBである。
同値：集合：正	並列	A、B、Cが挙げられる。
同値：話題：非	対比	Aである一方Bである。
順序：時間：正	推移	まずA、次にBを行う。
順序：因果：正	因果	AであるゆえにBである。
同値：話題：正	同格	AとはすなわちBである。
包含：集合：正	階層	AとはBの一部である。
順序：入出力：正	入出力	AからBにCが送られる。

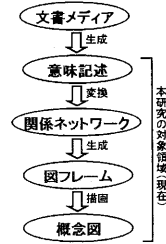


図 7: 概念図自動生成の流れ

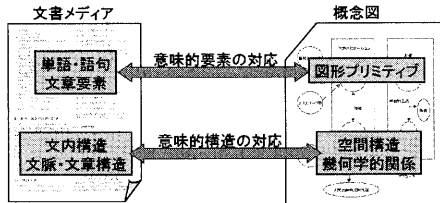


図 8: 文書と概念図の意味的対応

文書間構造 文書や文書群と図形要素が対応

構造の対応

次に意味的な構造同士を対応させ、それにより図のレイアウトを形成する。前述した文内構造と文脈・文章構造が、概念図における図形要素の配置規約などの空間構造に対応する。表1で挙げた関係が対応する空間構造は次のようなものがある。

順序関係: 連結図の持つ連結構造と対応する。

図形プリミティブを矢線で連結することでその順序と軸を示す。

包含関係: 領域図の持つ領域包含構造と対応する。

上位概念が下位概念を囲むような空間構造。

同値関係: 順序関係に準じる。

修飾関係: 順序関係に準じる。

現状では、文書メディアの意味的要素として単語・語句を、言語的構造として文内構造と文脈構造を考えており、その他の構成要素や構造間対応については今後の研究課題としている。

表 4: 意味記述の例

要素 (A,LABEL="XML")
要素 (B,LABEL="XML 宣言")
要素 (C,LABEL="DTD")
要素 (D,LABEL="XML インスタンス")
階層 (A,(B,C,D))

4 概念図の半自動生成手法

概念図の半自動生成手法の流れについて概説する。概念図の生成過程において複数のデータ構造を用意する。すなわち、入力データを図7のように多段階変換してゆくことにより、概念図を生成する。また、各データ構造間の変換は、各段階で変換ルール群を適用することで実現した。

4.1 概念図生成のためのデータ構造

概念図の生成過程において取り扱うべき構造が複数存在する。まず必要な構造は、入力する知識を表現するための構造である。

意味記述

文書メディアの意味構造を記述するために、宣言的記述を用いたデータ構造を用意した。これを本稿では意味記述と呼ぶ。具体的には、意味的要素および意味構造の宣言に接頭辞付パラメータリストをとって意味構造の記述を行う。ここでの接頭辞としては、3.1節で定義されたもの、及び表3に示されるような簡易表現を用いる。

実際に使用している意味記述の例を表4に示す。1行目から4行目が意味的要素の宣言である。A、B、C、Dという識別子を持つ意味的要素をテキストトラベルを付けて宣言している。5行目が意味構造

の記述であり、識別子 A で示される意味的要素が同じく B、C、D で示される要素の上位階層であるという宣言を行っている。また、この例では「階層」という簡易表現により意味構造を記述しているが、そのような場合はシステム内部で「包含:集合:正」という正規の形式に変換する。

図フレーム

システムの出力は概念図であるので、これを内部表現により記述する必要がある [8]。本研究では、概念図の図的構造・空間的構成を記述するデータ構造としてフレームを用いた。各フレームがそれぞれ一個の図形要素に対応する。ここでの図形要素を以後プリミティブと呼ぶ。フレームの構成要素であるスロットにプリミティブの属性情報が保存される。位置や大きさ、色や意味的属性などである。また、他のプリミティブとの空間的関係情報をも同時に保存する。

これを本稿では図フレームと呼ぶ。概念図は、この図フレームに記述された内容に基づいて生成される。

関係ネットワーク

意味構造とは文書メディアの意味的表現であり、図フレームは図的メディアの内部表現であるため、これらの間に意味的変換が必要となる。

本研究では複数の意味記述を統合するためのデータ構造として関係ネットワークを用いた。文書の意味記述における要素をノードとし、それら要素間の関係をリンクとして表現する意味ネットワークである。

このように意味記述から関係ネットワークを構成し、関係ネットワークから図フレームを構築するという過程を経ることで、概念図を無理なく生成できる。

4.2 データの変換

各データ構造の変換にはそれぞれの各段階で対応する変換ルール群を用いた。意味記述から関係ネットワークを構成する変換手続きは次のようになる。

1. 意味記述中の意味的要素に対応するノードを生成する。

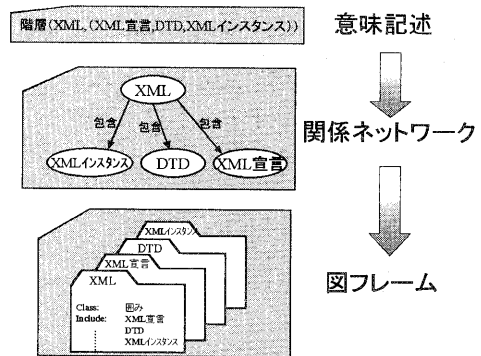


図 9: データの変換

2. 意味記述中の意味構造を入力し、対応するノード間リンクを生成する。

基本的にはこの二つの単純な手続きだが、ノード生成の際に意味記述での意味的属性を継承させたり、複数の記述を統合・圧縮してまとめるなどの処理も行われる。

次の段階である関係ネットワークから図フレームを構築する変換手続きは下のようになる。

1. 関係ネットワーク上の各ノードから、図フレームの原型を構成する。
2. 関係ネットワークのリンクに保存されているノード間の関係情報から、図フレーム間の空間的位置関係を生成する。その際に適用されるルールは次の通り。

順序関係: 概念要素同士を連結する連結要素は概念要素に接触していなければならない。一連の関係にある要素群は、初期状態では一直線上に並べられねばならない。

包含関係: 包含要素は被包含要素群を囲む閉曲線なくてはならない。

同値関係: 順序関係に準ずる

修飾関係: 順序関係に準ずる

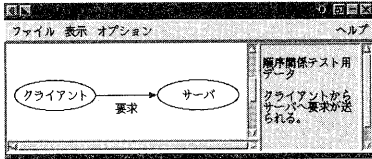
最後に、図フレームの座標値スロットに具体的な値を書き込む。そのために図フレームに記述されている空間的位置関係から具体的な配置座標を計算し、各フレームの座標値スロットに書き込む。次に座標値スロットに書き込まれた座標値に基づいて、プリミティブを概念図上に配置する。

表 5: 順序関係を示す意味記述

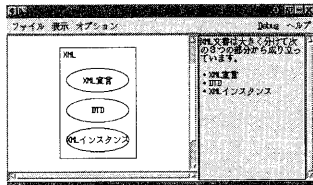
要素 (A,LABEL="クライアント")
要素 (B,LABEL="サーバ")
要素 (C,LABEL="要求")
(順序:入出力:正)(A,B,C)

表 6: 包含関係を示す意味記述

要素 (A,LABEL="XML")
要素 (B,LABEL="XML 宣言")
要素 (C,LABEL="DTD")
要素 (D,LABEL="XML インスタンス")
(包含:集合:正)(A,(B,C,D))



(a) 生成された連結図



(b) 生成された領域図

図 10: 基本的な概念図の生成実験

5 実験

図の生成

まず単一の意味構造よりなる概念図の生成例を示す。表 5 に示す順序関係の意味記述¹を入力すると、図 10(a) のように順序を表現する連結図を得る。また、同様にして表 6 を入力すると、図 10(b) のように階層を表現する領域図を得る。

次に表 7 のように意味構造が複合した概念図の生成を行った結果、図 11 にある図を得る。連結図の端子要素「サーバ」が領域図の囲み要素を兼ね、「クライアントからサーバに要求が送られる。サーバにはWWWサーバ、FTPサーバなどがある。」という意図が読み取れる。

¹この意味記述では、意味構造の要素宣言にパラメータを3つとっている。通常は一つ目のパラメータが関係の主体要素、2つ目のパラメータが客体要素を示すが、順序関係では3つ目のパラメータとして作用要素を指定できる。

表 7: 複合した意味構造を含む意味記述

要素 (A,LABEL="クライアント")
要素 (B,LABEL="サーバ")
要素 (C,LABEL="要求")
要素 (D,LABEL="WWW サーバ")
要素 (F,LABEL="FTP サーバ")
推移 (A,B,C)
階層 (B,(D,F))

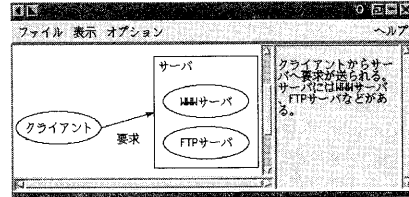


図 11: 複数の意味構造からの概念図の生成実験

は WWW サーバと FTP サーバがある。」という意図が読み取れる。

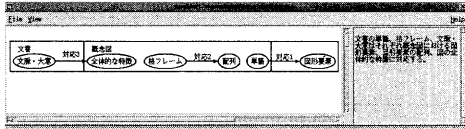
このように、複合した意味に特別に対応した描画ルールを整備しなくとも、複合した意味記述の内容を伝える概念図を生成できることが示された。これにより、比較的単純な描画ルールだけでも複雑な概念図を構成できることが考察できる。

しかし、図 11 では、「要求」とラベル付けされた矢線要素が斜めになっている。これは、連結図を構成するための描画ルールと領域図を構成するための描画ルールが競合したために水平にならなかったと考えられる。そのように、複数のルールが適用される場合にどのように整合性をとるかは今後の課題である。

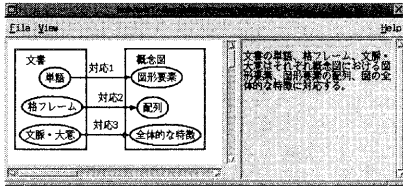
編集操作

編集操作の実験については、上記の実験とは異なる試作システム [9] により行った。

まず5つの複合した意味構造の記述を入力した際に、図 12(a) に示す図形が得られた。これは、「包含関係のないプリミティブが内部にあってはならない」というルールを実装していないこと、初期配置においてはプリミティブを横並びに配置するアルゴリズムのみ実装していることが原因である。このような図は満足できるものではないので、人手による編集操作を加える。図 12(a) 中の「単語」「各フレーム」「文脈・大意」というテキストによりラベル付



(a) 編集前



(b) 編集後

図 12: 編集実験

けされた3つのプリミティブをドラッグ&ドロップにより垂直に並ぶよう動かした。その過程では次のようなルールが適用された。

- 囲みプリミティブは常にその包含すべきプリミティブの領域を囲む位置・大きさを維持する
- 矢線プリミティブは接触すべきプリミティブと常に接触し続ける
- 順序関係の下位プリミティブは、上位プリミティブに連動して動く

その結果、動かしたプリミティブに関係する他のプリミティブが編集ルールにより変位・変形され、図 12(b) に示す図が得られた。

このように、少ない手順で比較的良好な図が得られる。これは、編集操作で図の意味的な構造を損なわないようにシステムが人間の補助を行った結果であると言える。

6 おわりに

知識ナビゲーションを目的とする概念図の生成手法について提案した。まず生成過程で処理すべき構造に適したデータ構造を多段階変換することによって概念図を半自動生成する手法について述べ、実装した試作システムを用いた概念図生成実験を行った。そして実験の結果、文章の構造を表現する概念図の生成が可能であることが実証された。

今後の課題としては、概念図生成の各段階に用いられるデータ構造変換ルールをより充実させることが必要である。また、関係ネットワーク構成段階において、推論などの知識処理を実装することで、より複雑かつ多様な意味構造を図化することができると考えられる。

参考文献

- [1] 中村裕一, 古川亮: 概念図理解を目的としたパターン情報と自然言語情報の統合, 情報処理学会論文誌, 第 36 卷, pp. 196-206 (1995).
- [2] 高橋美羽, 中村裕一, 大田友一: ハイパーメディア構築のための概念図と説明文の自動統合, 情報処理学会第 50 回 (前期) 全国大会 4N-8 (1995).
- [3] 恩田雅之, 中村裕一, 大田友一: 概念図と説明文を有機的に統合したハイパーメディアの構築, 情報処理学会第 50 回 (前期) 全国大会 4N-7 (1995).
- [4] 國藤進: 発想支援システムの研究開発動向とその課題, 人工知能学会誌, Vol. 8, No. 5, pp. 552-559 (1993).
- [5] 中村裕一, 村山正司, 大田友一: 図的メディアと言語メディアの統合による知識の解析と提示, 知能情報メディアシンポジウム予稿集 (1998).
- [6] 出原栄一, 吉田武生, 渥美浩章: 図の体系～図的思考とその表現～, 日科技連 (1986).
- [7] 永野賢: 文章論総説: 文章論的思考, 朝倉書店 (1986).
- [8] 佐藤亮一, 田中一敏, 門田充弘, 山下紘一: 概念図作成支援のための図の意味記述, 知識工学と人工知能 59-15, pp. 131-138 (1988).
- [9] 村山正司, 中村裕一, 大田友一: 知識ナビゲーションのための概念図の自動生成, 情報処理学会第 57 回 (後期) 全国大会 1L-2 (1998).