

文書概覧のための概念図の自動生成手法

Automatic Diagram Generation for Document Outline View

村山 正司[†] 中村 裕一[†] 大田 友一[†]
Masashi Murayama[†] Yuichi Nakamura[†] Yuichi Ohta[†]

[†] 筑波大学 機能工学系

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

はじめに

大量の文書データに簡単にアクセスできる環境が整い

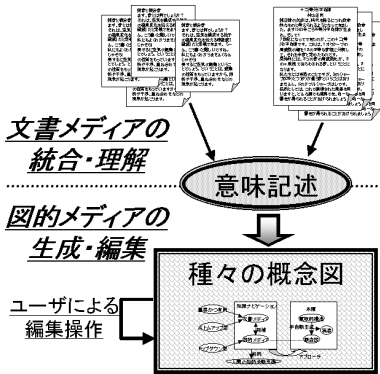


図 1: 文書の可視化による概覧

つつある。しかし、必要な情報を短時間で把握することは、未だに難しい問題である。本稿では文書メディアの概覧性を高めるため、図 1 に示すように、文書内容を表現する概念図を自動生成する手法を紹介する。

文章の意味的構造の可視化

本研究 [1] は文書内容を図的メディアによって説明することを目的としている。そのようなメディア変換には二つの大きな問題がある。文書から意味的構造を抽出する問題と、意味的構造から適切な概念図へ変換する問題である。我々はそれを次のようにして解決した。

A: 文書から意味的構造を抽出するメディア理解の問題

入力文書として、タグ付き文書を用いる。本研究では、文書の言語的構造をタグによって明示することのできる GDA タグセット [2] を利用する。文書中の意味的構造には主に文内構造及び文脈・文書構造の二つがあるが、これらは GDA の関係子を用いて記述できる。

B: 意味的構造から概念図を生成するメディア生成の問題

図的メディアでは空間構造によって情報を表現するが、その空間構造が伝達効果に大きな影響を与える。そのため、典型的な概念図により表現できる意味的構造と、文書が持つ意味的構造との間に存在する対応関係を調べ、図の生成ルールとして利用した。だが、自動生成された図が必ずしも良い図になるとは限らない。そのため、人手による編集をシステムが支援することとした。それには、図の意味的構造を保持し、図の編集操作に対して意味的構造を崩さないような補助処理を行う。

概念図の生成手法

GDA の関係記述を 4 種類に大きく分類する。これらは、次の表に示すように連結図・領域図にそれぞれ対応

している。この対応関係を用いることで意味的構造を概念図で表現する。

対応する図	文章の意味構造	GDA 関係子
連結図	順序関係	src, gol, cau 等
	同値関係	eq
	修飾関係	cnd, bas, agt 等
領域図	包含関係	sup, sub

GDA の記述から中間記述を多段階変換してゆき、概念図の生成を行う。その流れは次のようになる。

1. GDA 文書の入力・意味記述への変換
GDA 文書中の関係子を 4 種類の関係記述に変換
2. 関係ネットワーク構成
意味要素間の関係を意味ネットワークの形式に変換
3. 図フレーム構成
意味要素に対応するプリミティブ (基本的な図形要素) を一つのフレームに割り当て、それらの間の幾何学的関係を記述
4. 描画ルール適用による図の描画
 - (a) プリミティブの視覚的属性 (形状・色等) 決定
 - (b) プリミティブの生成 (初期配置)
 - (c) プリミティブの配置・レイアウト

このような手法で概念図を生成した例を図 2 に示す。サンプルデータとして用いた文書 (図左側) は簡単なものであるが、図からその内容が概ね判別できる。

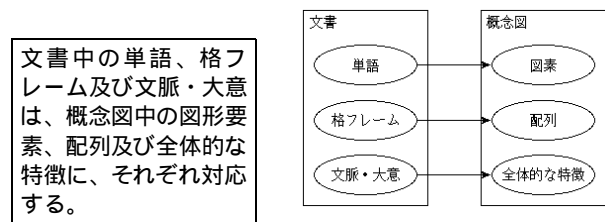


図 2: 概念図生成結果

おわりに

我々の提案する手法で、比較的簡単な文書に対して概念図を生成できることを検証した。今後は、より多様な文書の入力、多様な概念図の生成について検討し、本手法の有用性を確認する予定である。

参考文献

- [1] 村山正司, 中村裕一, 大田友一: 概念図の自動生成によるタグ付文書の可視化, TL 99-25, 電子情報通信学会研究報告 (1999).
- [2] 橋田浩一: GDA: 意味的修飾に基づく多用途の知的コンテンツ, 人工知能学会誌, Vol.13, No.4, pp.528-535 (1998)