

# 拡張されたトピックマップデータベースの構築

栗原 優樹<sup>†</sup> 細谷岳志<sup>†</sup> 木村 昌臣<sup>†</sup>  
芝浦工業大学工学部情報工学科<sup>†</sup>

## 1. 研究背景と目的

近年、知識表現の方法の1つとして、知識を記号化し、それらの間及び関連する情報リソースを結び付けるトピックマップ技術が注目を集めている<sup>1)</sup>。トピックマップは主に任意の概念を指すトピック、トピック間の関係を表す関連、そしてトピックと Web 等の情報リソースとの関係を表す出現から構成されている。また、トピックマップを検索するための問い合わせ言語（以降、TMQL）は現在 ISO で提案段階にあり、最終的な標準化段階には至っていないが、この提案は tolog<sup>2)</sup> という問い合わせ言語をもとに行われている。トピックマップでは、トピックと関連の集合を知識とし、その構造はトピック間に関連を表す線を引くことでグラフ構造として表現される。トピックマップ情報を格納する既存のデータベースではトピックの隣接関係のみに着目した検索のみを実現しているため、特にトピック数が多い場合には直接隣接していないが関係性のあるトピック群などを取り出すことが出来ない。また、TMQL は検索のみが検討されており、更新、挿入、削除等の処理については触れていない。そこで本研究では、トピックマップを複雑ネットワークと見なしてこれにクラスタ解析を適用する拡張を行い、関連性が強いトピック群が形成するグループを抽出する機能及び挿入、削除機能を持つトピックマップデータベース（以降、TMDB）を構築する。本研究の TMDB とはトピックマップ情報及びネットワーク解析に必要な情報を保持し、必要に応じてこれらの情報の検索及びネットワーク解析を行うデータベースシステムを指す。

## 2. 構築する TMDB の提案

トピックマップのグラフ構造に着目し、クラスタ解析を行う際、トピックマップの情報やクラスタ解析に必要な情報を保持する必要がある。そのため、膨大となるデータを全てメモリ上に保持し、処理を行うことは処理が可能かどうかをメモリ容量に依存することになる。また、ユーザがクラスタ解析した結果を要求するたびに毎回解析する必要がある。そこでメモリ以外の記憶領域も活用するため、これらの情報を格納する TMDB の構築を行う。

### 2.1. システム概要と流れ

本研究のシステムは検索処理プログラム、OODB、解析プログラム、質問プログラムから成る（図1）。

なお、本研究でのクラスタ解析は統計ソフト R の spinglass.community 関数<sup>3)</sup> を利用するが、その際、隣

接行列が必要になる。隣接行列とはグラフを表現する為に用いる行列であり、頂点間に枝が張られていれば1、張られていなければ0の値を成分として持つ。本研究ではトピックを頂点、関連を枝としてトピックマップのトピック及び関連から隣接行列を作成する。

システムは以下の流れになる。まず、格納処理によってトピックマップ情報を格納する。格納されているトピックマップに対して解析処理によってクラスタ解析を行い、結果をデータベースへ格納する。さらに、質問プログラムが、ユーザより検索、挿入、削除のいずれかの処理の命令を受ける。この要求を構文解析し API を通して検索処理プログラムの対応するメソッドを呼び出す。検索処理プログラムは要求された命令をデータベースへ行い、結果を質問処理プログラムへ送る。

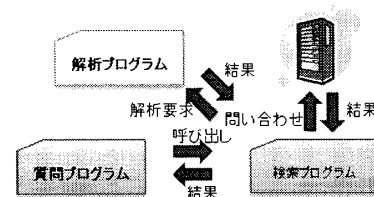


図 1 システム概要

### 2.2. TMDB で用いるデータベース

既存のトピックマップツールのデータベースの実現方法には主にリレーショナルデータベース（以降、RDB）を利用したものとおブジェクト指向データベース（以降、OODB）を利用したものがある。それぞれ比較したところ、RDB は普及率が高いため、データの再利用性が高い。しかし、トピック間の関連を調べる場合など、複数の表に跨った検索を行う場合には、各表のデータ行を掛け合わせた数に対して検索を行うので、トピックマップにおいては効率が悪い。一方、OODB では、データ間の繋がりに関する検索はオブジェクトの参照を辿るだけであるため効率が良い。そのため、本研究では TMDB に利用するデータベースとして OODB を用いる。なお、OODB にはオープンソースの db4o を利用する。

### 2.3. 質問処理プログラム

質問処理プログラムは、質問文を字句解析、構文解析し、検索処理プログラムに処理を行わせ、結果を受け取り、結果を質問文の条件に合うようデータを調整して出力する処理である。データの調整とは、射影やソートを行うことである。

字句解析を行うプログラムである Lexer と構文解析を行うプログラムである Parser の作成には、ANTLR というパーサ生成ツールを用いる。ANTLR を使用するには、問い合

The Construction Of The Extended Topic Map Database

<sup>†</sup>Yuuki Kuribara, Takeshi Hosoya, Masaomi Kimura

<sup>†</sup>Shibaura Institute of Technology

