

図書館と Web の分類体系を統合的に活用した テーマグラフ可視化インタフェース

坂井 哲[†] 清田 陽司^{††} 増田 英孝[†] 中川 裕志^{††}

[†] 東京電機大学 未来科学部 ^{††} 東京大学 情報基盤センター

1 はじめに

情報検索には、ユーザの情報要求が明確であるとは限らないという問題がある。そのため情報検索システムの研究においては、ユーザの情報要求の曖昧性に対処する必要性が論じられている。

解決手段のひとつとして、図書館のレファレンスサービスが存在する。図書館員が利用者にさまざまな切り口からインタビューを行い、ヒントを示しながら徐々に曖昧な情報要求を具体化する。

本研究では、利用者の情報要求を具体化させるための一手段として、利用者が入力したキーワードに対する調べ方の切り口を可視化し、ヒントとなる情報資源に誘導するインタフェースを提供する。このインタフェースは、ユーザが入力したキーワード (e.g. 阪神・淡路大震災) とテーマ (e.g. 経済史、地震) のつながりを示すグラフ (テーマグラフ) をヒントとして表示する。ユーザにテーマグラフを見せることで、曖昧だった情報検索のテーマが明確になることが期待できる。キーワードが持つテーマがわかると、探索事項を一般化して調べることが可能となり、複数の観点から掘り下げて調べることができるようになる。またグラフ化することで、テーマとキーワードの関連、テーマ同士の関連を一目で理解することができる。

2 図書館分類体系と Wikipedia の統合利用

テーマグラフを構築するために利用する分類体系として、BSH (図書館件名標目表)、NDLSH (国立国会図書館件名標目表)、Wikipedia がある。図書館分類体系と Wikipedia にはそれぞれ利点と欠点がある。我々はお互いの欠点を相互に補完するために、これらを統合的に利用する手法を提案している [1] [2]。

各分類体系の比較を表 1 に示す。

BSH、NDLSH は件名標目表といい、資料の主題や形式をどんなキーワード (件名) で表すかについての

表 1: 図書館分類体系と Wikipedia の比較

	BSH,NDLSH	Wikipedia
アプローチ	トップダウン的	ボトムアップ的
長所	安定 深い組織化	多様な観点の反映 新しい概念への対応
短所	多様な観点の反映不可 新しい概念を含まず	不安定 浅い組織化
上位概念	原則として1つもつ	複数もてる

約束事である。図書やパンフレットなどの図書館資料には、件名が1つ、又は複数割り当てられている。件名を付与することにより、主題を表す統制キーワードによる図書館資料の検索が可能になる。BSH の構造から一部の件名を抜粋したものを図 1 に示す。専門家による慎重な作業を経てトップダウン的に構築されていることから、安定的に利用することができる、深い組織化がなされているという長所がある。しかし、語彙量が少なく、改定の周期が長いこと新しくできた概念には対応しにくいという欠点がある。

Wikipedia は、Web 上で共同編集されている多言語百科事典で、誰でもどこからでも自由に編集することができる。多数の人が編集に参加していることから、世界中の多くの概念を網羅している。各記事にはカテゴリが付与されていて、1つの記事・カテゴリに対し複数の上位概念 (カテゴリ) を与えることができる点が他の分類体系と大きく異なる。その結果、図 2 に示すように各記事に対して多様な観点がカテゴリのネットワークとして付与されている。しかし、専門家だけが編集するわけではないので、記事の信頼性、組織化の正確性を損なう場合がある。

これらの分類体系を統合的に利用することによって、広範な概念を網羅している Wikipedia を情報探索の出発点とし、そこから概念を一般化することによって図書館分類体系に導いていく。図書館分類体系における件名は、専門家による厳密な分類であるため、ある観点の集合とみなすことができる。これをテーマと呼ぶことにする。具体例を次節で示す。

3 テーマグラフ可視化インタフェース

実行例を図 3 に示す。ここでは、入力キーワードを「阪神・淡路大震災」としている。図 1 から「阪神・淡路大震災」は「経済史」、「地震」、「マスメディア」、

Induction and Visualization of Theme-graphs based on Library Classification Systems and Web Categorization Systems

[†] Satoshi Sakai

^{††} Yoji Kiyota

[†] Hidetaka Masuda

^{††} Hiroshi Nakagawa

School of Science and Technology for Future Life, Tokyo-Denki University ([†])

Information Technology Center, The University of Tokyo (^{††})

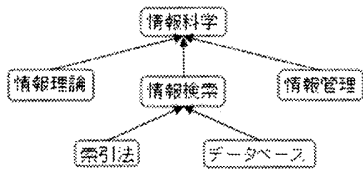


図 1: 件名標目表 (BSH) の「情報検索」の関係図

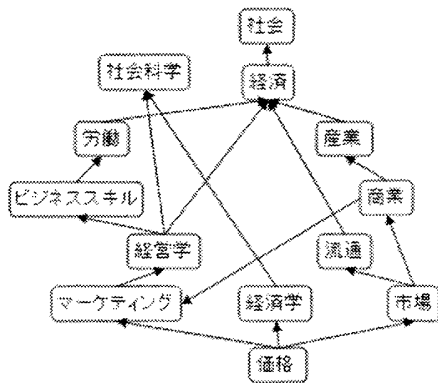


図 2: Wikipedia において記事「価格」に与えられているカテゴリのネットワーク

「平成時代」というテーマを持っていることが明確になり、調べる際の切り口ができる。例えば、「経済史」からは「阪神・淡路大震災が日本経済に与えた影響」について、「地震」からは「阪神・淡路大震災を引き起こした地震(すなわち 1995 年兵庫県南部地震)」について、それぞれ調べることが可能だとわかる。これをグラフ化することにより、概念間の関連を理解しやすくしたものが、テーマグラフである。

インタフェースを開発するにあたって、オープンソースのグラフ描画ソフトウェア Graphviz を使用した。Graphviz はグラフ構造を表示するのに適していて、画像ファイルを出力する。

テーマグラフを表示する手順を以下に示す。

1. ユーザーがキーワードを入力する。
2. 入力キーワードに一致する Wikipedia の記事・カテゴリを探索する。
3. 上位カテゴリの中で、まだ辿っていないものを探索する。
4. Wikipedia のカテゴリ名が BSH,NDLSH のカテゴリ名と一致すれば、それを 1 つのテーマとする。
5. テーマが存在した場合、上位カテゴリが存在しない場合は、1 つ前のカテゴリへと戻る。また、探索が爆発的に増えていくことから、最初の階層から 7 階層目まで達した場合には探索を打ち切る。
6. 3,4,5 を繰り返し、辿るべき上位カテゴリがな

れば探索を終了する。

以上の手順から作成した情報を Graphviz の入力として、テーマグラフをユーザに提示する。

またテーマについては、BSH,NDLSH にそれぞれ存在する場合とこれらの両方に存在する場合があるため、それぞれ色を分けて表示する。

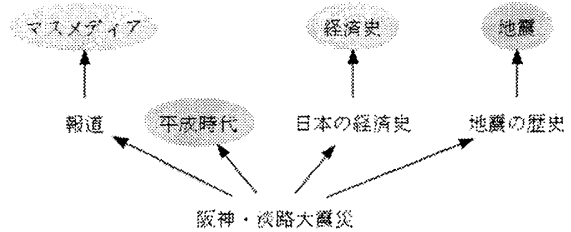


図 3: 「阪神・淡路大震災」のテーマグラフ
※「マスメディア」「経済史」は BSH、「平成時代」は NDLSH、「地震」は BSH 及び NDLSH の件名

4 おわりに

本稿では、情報検索におけるユーザの情報要求の曖昧性を解消するために、キーワードとテーマのつながりを示すグラフを提示することが有効であると考え、テーマグラフ可視化インタフェースを開発した。テーマグラフを構築するために、図書館分類体系と Wikipedia を統合的に利用した。

今後、実運用でインタフェースの有効性についての評価をしていく予定である。

謝辞

本研究は、文部科学省科学研究費補助金若手研究(B)(課題番号 18700134)の助成を受けて遂行されました。また、国立国会図書館書誌部から提供された「国立国会図書館件名標目表(NDLSH) 2006 年度版」を利用しました。

参考文献

- [1] 清田陽司, 田村悟之, 中川裕志, 増田英孝. ReferenceNavigator: 異種オントロジーの統合ブラウジングツール~図書館の分類体系と Wikipedia カテゴリの対応付け~. 言語処理学会第 13 回年次大会ワークショップ「言語的オントロジーの構築・連携・利用」論文集 (2007)
- [2] 田村 悟之, 清田 陽司, 増田 英孝, 中川 裕志. 図書館における自動レファレンスサービスシステムの実現に向けて ~Web 上の二次情報と図書館の一次情報の統合~. 情報処理学会 研究報告 情報学基礎 (SIG-FI)(2007)