

Reference Navigator: 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～

清田陽司[†] 田村悟之[‡] 中川裕志[†] 増田英孝[‡]

[†] 東京大学情報基盤センター 図書館電子化研究部門

kiyota@r.dl.itc.u-tokyo.ac.jp, nakagawa@dl.itc.u-tokyo.ac.jp

[‡] 東京電機大学大学院工学研究科 情報メディア学専攻

tamura@cdl.im.dendai.ac.jp, masuda@im.dendai.ac.jp

1 はじめに

インターネットが一般に普及する以前、情報探索を行うにあたって中心的な役割を果たしていたのは図書館であった。図書館には、膨大な資料を整理・組織化し、利用者の情報探索に役立てるためのさまざまなツールが存在する。代表的なツールとしては、日本十進分類法 (NDC)[1] や基本件名標目表 (BSH)[2] などの分類体系があげられる。

現在では図書館にかわって Web が情報探索の中心的な役割を果たしつつある。まず最初に Google などの Web サーチエンジンを利用することで、たいいていの検索質問に対しては何らかの Web ページが見つかるという状況になっている。しかし、Web 上の膨大な情報は十分に組織化されていないため、情報要求に適したページを見つけることは必ずしも簡単ではない。

近年、Web 技術を活用することで膨大な知識を多数の共同作業によって組織化しようという取り組みが盛んに行われるようになってきている。その代表例としては、Wikipedia¹ や、各種の Folksonomy サービス (Flickr² など) があげられる。その結果として、Web 上に一種の情報探索用オントロジーとよべるものが存在するという状況が生まれつつある。これらのオントロジーには、多数の人々の多様な観点が反映されているという大きな利点があるが、限界もある。これらのオントロジーの構築に関わる人々の中心的な動機は、「自分の興味を満たす情報を効率的に探せるようにしたい」というものであるから、「情報の消費者」としての観点が大きく反映される。しかし、「情報の生産者」、いいかえると「これまであまり知られていない情報を探し出し、新たな知見を生み出したい」という知的探究心をもった人々の観点は反映されにくい。なぜならば、そういった人々には自分にとって価値の高い情報資源の在処を秘密にしておきたいという動機がはた

くからである。結果として、Web 上に存在するオントロジーは浅い組織化にとどまりがちであり、探索を深く掘り下げていくという用途には必ずしも適していない。

一方で、Web の世界での変革に対応して、図書館の世界においても新たな潮流がみられる。図書館において情報探索の窓口の役割を果たしているレファレンスサービスでは、Wikipedia をレファレンスツールとして活用するという試みがなされている [3]。しかし、現在はレファレンスサービス自体が一般の人々にあまり認知されていないという状況である [4]。また、特定のトピックごとに図書館が提供できる関連資料をリスト化したパスファインダーとよばれる情報資源を Web 上で公開する取り組みも盛んに行われるようになって [5] が、図書館が人手をかけて構築する必要があるため、Web 上に存在するオントロジーに比較すればごく一部のトピックをカバーしているにすぎない。図書館はまだ Web の変革のスピードに追いついていないというのが現状である。

このように、Web による情報探索と図書館による情報探索にはそれぞれ問題点がある。そこで我々は、解決策として図書館の分類体系と Web で構築された各種オントロジーを対応付け、情報探索のインフラとして活用することを提案する。これが実現できれば、Web を出発点とした情報探索を、図書館を利用した深い探索に誘導していくことが可能になり、図書館の存在意義の再発見にもつながっていくと考えている。

本稿の構成を以下に述べる。まず 2 節において、図書館の分類体系 (ここでは NDC, BSH) とオントロジーとしての Wikipedia (とくに Wikipedia カテゴリ体系) を比較し、それを踏まえて 3 節で両者の対応付けと情報探索への統合的活用法を提案し、その有用性を述べる。4 節では最初の取り組みとして構築した異種オントロジーの統合ブラウジングツール「Reference Navigator」を紹介し、今後の研究計画を示す。5 節でまとめを述べる。

¹<http://ja.wikipedia.org/>

²<http://www.flickr.com/>

2 図書館の分類体系と Wikipedia の比較

本節では、情報探索ツールとしての図書館の分類体系と Wikipedia の比較を行い、それぞれの長所・短所を整理する。要点を表 1 の上方にまとめた。

図書館の分類体系

一般的に図書館情報学分野の専門家によって管理されている。例えば日本図書館協会³では、委員会を組織して NDC や BSH の改定作業を行っており、現在は NDC 第 9 版 (NDC9)、BSH 第 4 版 (BSH4) が最新版である。

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。また、特に BSH、LCSH(米国議会図書館件名標目表)、NDLSH(国立国会図書館件名標目表)などの各種件名標目表は、情報探索に広がりをもたせるためにきわめて有用なツールである。

反面で、改定の周期が長いと新しくできた概念には対応できない、一般的に上位概念は 1 つしかもてないため概念の多様な側面を反映しづらいという欠点をもっている。また、件名標目表は一般的に冊子体で提供されているため、一般利用者からは利用しづらいものである。LCSH などは Web から有料でアクセス可能であるが、一般利用者からは縁遠い存在であることに変わりはない。

Wikipedia

Web 上で Wiki システムによって共同編集されているオンライン百科事典であり、誰でもどこからでも自由に編集することができる。現在、200 以上の言語で作成されており、日本語版では 30 万項目以上の記事が存在する。

多数の人が編集に参加していることから、世界中のほとんどの概念を網羅していると思われる。また、ハイパーリンクを容易に張ることができるという Wiki の特徴を生かし、項目の組織化の取り組みも行われている。膨大な項目から効率的に一覧を生成するために「Wikipedia カテゴリ」というボトムアップ的な組織化の仕組みが導入されており、項目を多様な観点から分類することが可能となっている (Folksonomy に類似)。例えば、「価格」という項目には「マーケティング」「経済学」「市場」といった複数のカテゴリを付与

表 1: 図書館の分類体系と Wikipedia カテゴリの比較

	図書館の分類体系 (BSH4)	Wikipedia (カテゴリ)
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
親カテゴリ	1 つしかもてない	複数もてる
総カテゴリ数	11,184	15,532
カテゴリ名一致	1,363	

(2006 年 8 月現在)

することができる。また、カテゴリ自体にも上位概念 (親) のカテゴリを付与することが可能である。複数の親をもたせることができるのは、従来の分類体系にはない大きな特徴である。

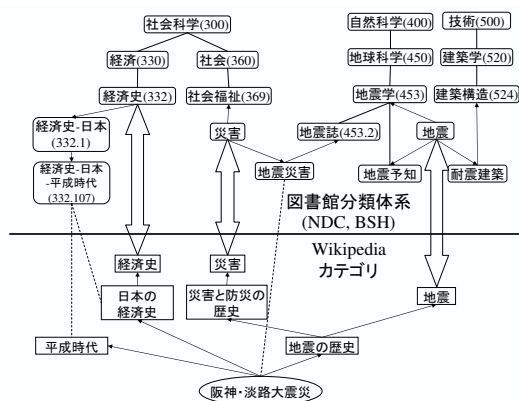
反面で、頻繁な編集が行われるため体系が安定していない。信頼性に欠ける情報も少なくないといった欠点がある。信頼性を確保するためにさまざまな取り組みがなされているが、対症療法的なものにとどまっている。また、1 節で述べた理由により浅い組織化にとどまっていることが多いという問題もある。

3 解決策: 両者の統合的活用

これまで述べてきたように、情報探索ツールとしての図書館の分類体系と Wikipedia には、それぞれ異なる利点と欠点があるため、情報探索を行う際には両者を適切に使い分けることが有用だと考えられる。では、どのような使い分けをすべきだろうか？本稿では、両者のもつ共通部分を対応付けることによって、情報探索の出発点として Wikipedia を用い、そこから概念を一般化することによって図書館の分類体系に導いていくという統合的活用法を提案する。

図 1 に情報探索への活用例を示す。まず、情報探索の出発点としての Wikipedia の活用について説明する。ここで、Wikipedia の記事「阪神・淡路大震災」には、カテゴリとして「日本の経済史」「地震の歴史」が付与されている。さらに、カテゴリ「日本の経済史」には上位カテゴリとして「経済史」が、カテゴリ「地震の歴史」には上位カテゴリとして「災害と防災の歴史」「地震」が付与されている。このように、Wikipedia の記事の一つとりあげてみると、関連するカテゴリ群をツリー構造として取り出せることがわかる。このツリー構造をたどることで、「阪神・淡路大震災」について調べる際の切り口が明確になっていく。例えば、「日本の経済史」「経済史」とたどっていくと、「阪神・淡路大震災が日本経済に与えた影響」について調べる

³<http://www.jla.or.jp/>



(図書館の分類体系において分類記号が付記されているものは NDC の項目、それ以外は BSH の項目である)

図 1: 図書館の分類体系と Wikipedia カテゴリの対応付け

ことが可能だとわかるし、「地震の歴史」「地震」とたどっていくと、「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)」について調べることが可能だとわかる。結果として、「阪神・淡路大震災」という探索主題を「経済史」「災害」「地震」に一般化して調べることが可能となる。

次に、Wikipedia カテゴリと図書館の分類体系(BSH, NDC)の対応付けと、図書館の分類体系の活用について説明する。Wikipedia カテゴリと図書館の分類体系の間には、後述するようにカテゴリ名が一致するものが存在する、図 1 では「経済史」「災害」「地震」が一致している。この一致を利用することにより、Wikipedia の活用によって一般化された探索主題を、図書館の分類体系によってさらに掘り下げていくことが可能となる、例えば「経済史」「経済史-日本」「経済史-日本-平成時代」とたどっていくと、NDC 分類記号 332.107 によって「日本経済事典」(日本経済新聞社、1996 年)などを探すことができるし、「地震」「耐震建築」とたどっていくと、件名「耐震建築」によって「地震とマンション」(西澤英和ほか、筑摩書房、2000 年)などを探すことができる。

ここで述べた統合的活用のアプローチは、以下の理由によってきわめて有用であると考えている。

オーバーラップが存在すること 表 1 の下方に示すように、図書館の分類体系(BSH4)と Wikipedia カテゴリの間には、カテゴリ名が一致するものが多数(Wikipedia の 15,532 カテゴリに対して 1,363 カテゴリ)存在するため、Wikipedia の項目から図書館の情報資源に誘導できる可能性は大きいと考えられる。また、類似したカテゴリ名も対応付



図 2: Reference Navigator

けることができれば、さらにその可能性は大きくなる。このことは図書館情報資源の有効活用につながる。

広範な概念を網羅できること 図書館の分類体系(とくに BSH)の「項目数が少なく、新しい概念を網羅できていない」という弱点を、Wikipedia カテゴリとの対応付けによって補うことができる。すなわち、Wikipedia に含まれる膨大な数の項目を情報探索の出発点とできるので、(潜在的には)世界中のあらゆる概念に対応可能である。

情報探索を収束させること Wikipedia カテゴリはボトムアップ的に構築されているため、浅い組織化にとどまっていることも多い。これをそのまま情報探索に利用すると、いわゆる Web サーフィンとなってしまう探索が発散してしまう危険性がある。図書館の分類体系と対応付けて利用することによって、情報探索が収束し、深い探索につながっていくと考えられる。

4 Reference Navigator

前節で述べた図書館の分類体系と Wikipedia の統合的利用の最初の試みとして、我々は異種オントロジーの統合ブラウジングツール「Reference Navigator」を試作した。本ツールのインタフェースを図 2 に示す。

本ツールは、Wikipedia、NDC9、BSH4 のオントロジー構造に加えて、Open Directory Project⁴で構築された Web ディレクトリ構造をリレーショナルデータベースに格納⁵しており、統合的に探索することが可能である。上のテキストボックスに探索主題(Wikipedia

⁴http://dmoz.org/

⁵NDC9、BSH4 の電子版の利用許諾を日本図書館協会より得ている。

の項目名)を入力すると、その上位カテゴリ名などがハイパーリンクとして表示される。ここでいずれかのカテゴリ名をクリックすると、さらにそのカテゴリの上位カテゴリ名、下位カテゴリ名などが表示される。同一のカテゴリ名が Wikipedia 以外にも含まれる場合には、それらの上位カテゴリ名、下位カテゴリ名も共に表示される。クリックを繰り返していくことで、情報探索を深めていき、図書館の情報資源を探すための手掛かりを得ることができる。図2では、カテゴリ「人工知能」の Wikipedia, BSH, Open Directory における上位・下位カテゴリ, Wikipedia において「人工知能」カテゴリに属する項目名、「人工知能」に対応する NDC 分類記号などが一覧表示されている。

現在は本ツールを東京大学総合図書館におけるレファレンスサービス業務などに試用している段階であるが、その結果を受けて研究を以下の方向に発展させていくことを計画している。

類似したカテゴリ名の対応付け 現在の図書館の分類体系と Wikipedia の対応付けは、カテゴリ名が文字列レベルで完全に一致したものととどまっている。自然言語処理技術を活用することで、類似したカテゴリ名などを精度良く対応付けることを目指す。例えば、図1の点線部分で示した関係(「阪神・淡路大震災」 \longleftrightarrow 「地震災害」, 「日本の経済史」+「平成時代」 \longleftrightarrow 「NDC332.107」)を対応付ける。

パスファインダーの自動生成 前述したように、いくつかの図書館で構築されているパスファインダーは情報探索にとって有用なツールであるが、人手で構築されているため網羅されている項目が著しく少ないという問題点がある。そこで、Wikipedia の項目に対して NDC 分類記号などを自動推定し、それが付与されている参考図書などを蔵書目録(OPAC)から取得することで、あらゆる概念についてのパスファインダーを自動生成することを目指す。

自動レファレンスサービスシステムの実現 京大レファレンスサービスシステム [6] やダイアログナビ [7] の方法論を発展させ、対話的に図書館利用者の情報探索を支援するシステムを構築することを目指す。具体的には、本稿で提案した統合オントロジーに加えて、情報探索に関するさまざまなメタ知識 [8] などを知識ベースとして利用するシステムを想定している。現在、東京大学附属図書館におけるレファレンスサービス業務分析の結果 [9] にもとづき、プロトタイプシステムを試作している。

5 おわりに

本稿では、図書館の分類体系と Wikipedia カテゴリという性質の異なるオントロジーを統合的に活用することで新たな情報探索のインフラを実現できる可能性を示した。これが実現すれば、Web を手掛かりとして図書館の利用によって探索を深めていくという新たな情報探索のパラダイムを人々に広め、社会全体の情報リテラシーの向上につなげることができると考えている。また、図書館の存在意義の再発見や、電子図書館の実現にも活かせるであろう。

参考文献

- [1] もり・きよし, 日本図書館協会分類委員会(編). 日本十進分類法 新訂9版. 日本図書館協会, 1995.
- [2] 日本図書館協会件名標目委員会(編). 基本件名標目表 (BSH) 第4版. 日本図書館協会, 1999.
- [3] 兼宗進. デジタル・レファレンス・ツールとしての wikipedia. 情報の科学と技術, Vol. 56, No. 3, pp. 103–107, 2006.
- [4] 斎藤文男. 図書館利用者にとってのレファレンス・サービス. 東京都図書館協会報, No. 81, 2001.
- [5] Laura B. Cohen and Julie M. Still. A comparison of research university and two-year college library web sites: content, functionality, and form,. *College and research libraries*, Vol. 60, No. 3, pp. 275–289, 1999.
- [6] 黒橋禎夫, 日笠亘. 京都大学附属図書館における自動レファレンス・サービス・システム. 情報管理, Vol. 47, No. 1, pp. 184–189, 2001.
- [7] 清田陽司, 黒橋禎夫, 木戸冬子. 大規模テキスト知識ベースに基づく自動質問応答 —ダイアログナビ—. 自然言語処理, Vol. 10, No. 4, pp. 145–175, 2003.
- [8] 長澤雅男. 情報と文献の探索 第3版. 丸善株式会社, 1994.
- [9] 國安結, 清田陽司, 綾部輝幸. 東京大学附属図書館におけるレファレンスサービス業務分析 - 自動レファレンスサービスシステムの実現を目指して -. 日本図書館情報学会 第54回研究大会発表要綱, pp. 101–104, 2006.