

# 対話的情報探索支援のための統合分類体系の構築

清田 陽司 (Yoji Kiyota)

東京大学情報基盤センター 図書館電子化研究部門

E-mail: kiyota@r.dl.itc.u-tokyo.ac.jp

## 1. 対話と情報要求の階層

人が情報検索を行うプロセスは、情報要求が徐々に具体化するプロセスとしてとらえることができる。このプロセスをTaylorは以下の4階層に分類している[1]。

(Q1)直感的要求(visceral need):漠然と知識が欠けていることを認識しているが、どんな知識が欠けているかはわからない

(Q2)意識された要求(conscious need):どんな知識が欠けているかはほぼわかるが、どんな情報が必要かは明確に説明できない

(Q3)形式化された要求(formalized need):欠けている知識が明確にわかり論理的な質問ができる

(Q4)調整済みの要求(compromised need):検索を行う対象(質問を投げかける相手や検索システム)の知識構造や理解力を想定し、相手の制約を認識して検索を行うことができる

一方、情報検索を工学的アプローチによって研究する立場からは、さまざまな情報検索システムが提案されてきているが、ほとんどの研究はTaylorの階層におけるQ4を前提としており、Q1~Q3は対象外である。その理由としては、情報要求が具体化するプロセスを形式化するのが困難だということがあげられる。しかし、人が必要な情報を得られないのはそもそもQ1~Q3の階層から具体化できないのが原因となることが多い。

では、必要な情報を効率的に見つけられる人はどのようにして「情報要求の具体化」を行っているのだろうか？この命題を考える際に、文献探索のエキスパートである図書館員が行っている情報探索支援行動が参考になる。一般に、図書館のレファレンスサービスにおける質問回答の際、図書館員は利用者の調べたい主題が不明確な場合にさまざまな角度からインタビューを行い、徐々に利用者の情報要求を明確にしていく。さらに主題に関連するさまざまな情報資源に関する豊富な知識を活用することによって、利用者には適切な資料を提示することができる。すなわち、利用者の曖昧な情報要求(Q1~Q3のレベル)は、豊富な知識を持ったエキスパートとの対話プロセスによって具体化される。

コンピュータシステムによって図書館員の情報探索支援行動を模倣するには、多様な情報資源を活用するための仕組みを持たせる必要がある。世の中に存在する情報資源のなかでも重要な役割を果たしているのは、図書館情報資源とWeb情報資源である。しかし、

図書館情報資源には情報の新鮮度・カバレッジに問題があり、Web情報資源には情報の信頼性・専門性に問題がある。この問題を解決するには両者を統合して組織化・利用することによってお互いの欠点を補完することが必要である。本研究では、図書館において専門家によって厳密に構成されたトップダウン的な分類体系、Webにおいて多数の参加者によって構築されたボトムアップ的な分類体系を、両者の中間的な性質をもつ情報資源であるWikipediaを媒介として組み合わせることで、この問題を解決する。

## 2. 図書館分類体系とWikipediaの統合利用

効率的な情報探索を行うには、膨大な情報資源を組織化している各種の分類体系を活用することが不可欠である。図書館には、膨大な資料を組織化し、利用者の情報探索に役立てるためのツールとしての分類体系(日本十進分類法、基本件名標目表など)が存在している。一方近年では、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、効率的な情報探索に役立てようという取り組みがWikipediaなどで盛んに行われるようになってきている。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。情報探索ツールとしての図書館分類体系とWikipediaの特徴を整理すると、下記ようになる。

### 図書館分類体系

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(米国連邦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、情報探索に拡がりをもたせるために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、概念の多様な側面を反映しづらいという欠点をもつ。

### 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、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。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。複数の親をもたせているのは、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。

このように、図書館分類体系とWikipediaは対照的な特徴を有している。本研究では、両者のもつ共通部分に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している[2]。

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の記事を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。

### 3. 対話的情報探索支援システム Littel Navigator

上に述べた手法を図書館の現場において情報探索に利用できるように、対話型情報探索支援システム「Littel Navigator」を実装した。Littel Navigatorのユーザインタフェースを図2に示す。利用者は、調べたいことがらに関連していることばをキーワード入力欄に入力することで情報探索を開始できる。例えば、「阪神・淡路大震災」を入力すると、システムは「地震」「経済」「地震学」「災害」などのテーマを自動的に導出して表示する。さらに、それぞれのテーマが導出されたパス(経路)や、それぞれのテーマに関連する書架分類コード

(NDC)、有用な情報リソース、関連キーワードなどが表示される。

Littel Navigatorは、2007年5月より東京大学附属図書館などにおいて公開運用を開始している。また、Littel Navigatorの技術をベンチャー企業にライセンスし、国内外の大学図書館や公共図書館にサービスを提供していくという取り組みを行っている。この件に関しては日本経済新聞[3]、日経産業新聞[4]などで報道されている。

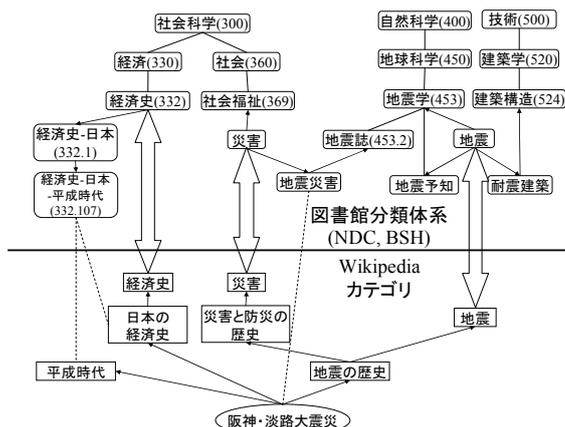


図 1: Wikipedia を利用した分類の自動導出

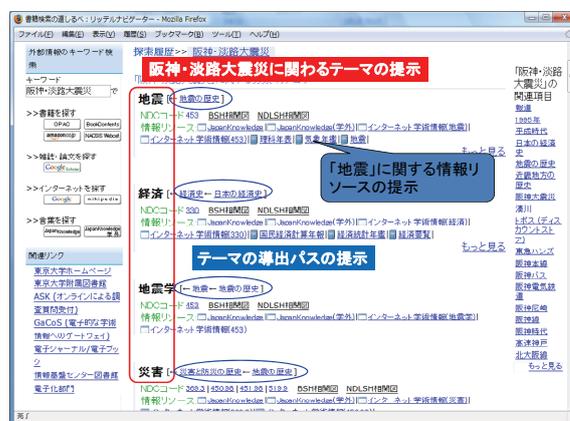


図 2: Littel Navigator のユーザインタフェース

## 文 献

- [1] R. S. Taylor, "Question-Negotiation and Information Seeking in Libraries", College and Research Libraries, vol.29, no.3, pp.178-194, 1968.
- [2] Y. Kiyota et al., Automated Subject Induction from Query Keywords through Wikipedia Categories and Subject Headings, Proc. LREC 2008, May 2008.
- [3] 日本経済新聞 2007年(平成19年)4月18日朝刊ベンチャー面: 図書館で本探し キーワード検索ーリッテルがシステム.
- [4] 日経産業新聞 2007年(平成19年)11月6日20面: 点検 大学発 VB(東大)リッテル 図書館の本分類表示 ~言語処理技術を活用~