

特集：リサーチ・ナビ—調べものに役立つWebサービス—
【リサーチ・ナビとは】

リサーチ・ナビ検索システムの技術

清田 陽司

1. はじめに

Webサーチエンジンの普及は、一般の人々の「情報探し」に対するもの見方に大きな影響を与えている。仕事や買い物、旅行、食事、読書、料理といった日々の行動をするにあたって「まずサーチエンジンで探す」という習慣は人々の間で広く共有されている。これにともない、「検索キーワードさえ適切に選べばサーチエンジンは答えを与えてくれる」とか、「サーチエンジンで何もみつからなければそれ以上探しても仕方ない」という誤解も生じかねない状況が生まれている。

しかし、私たちが日々抱く疑問の中には、検索キーワードを入力するだけでは答えを見いだせないものも数多く存在する。「関東大震災が発生したのはいつか?」「〇〇ってどんな病気?」という疑問に対しては、サーチエンジンで簡単に答えにたどりつくことができる。一方で、大学の学生の「関東大震災についてレポートを書かなきゃいけないんだけど、いったいどんな資料から調べたらいいの?」という疑問や、難病を抱えた患者の「自分の病気について最先端の治療を行っている病院を探す方法は?」という疑問に対しては、サーチエンジンは明解な答えは与えてくれない。

「レポートの書くための資料の探し方を知りたい」「病院を探す方法を知りたい」といった疑問を抱くのは、情報探しのテーマがあいまいな場合が多い。このような場合、情報探しのテーマを利用者に推薦することが求められる。情報探しのテーマを推薦するためには、サービス提供側は以下の3つの条件を満たしていることが必要だと考えられる (図1)。

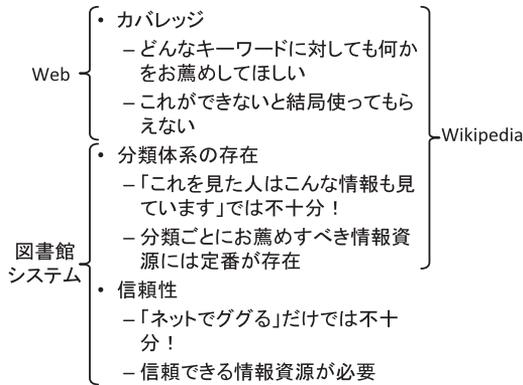


図1 テーマ推薦の要件

1. カバレッジ

あらゆる分野の質問に対して何らかの情報資源を推薦することが求められる。この条件を満たしていない限り、結局はサービスを利用してもらえない。

2. 分類体系の存在

お薦めすべき情報資源には、分類ごとに「定番」とよべるものが存在する。例えば、図書館においては入門書（新書など）や参考図書がそれに該当する。「定番」となる情報資源を利用者に提示するためには、あらかじめそれらの情報資源を整理しておく分類体系が必要である。

3. 信頼性

どんな情報資源を推薦したとしても、利用者側が「このお薦めは信頼できる」という感情を抱かなければ結局は利用されない。「ネットでググる」ことで得られる情報資源を超える信頼性の根拠を利用者に示すことが重要となる。

日常の情報探しにおいて最も広く用いられているWeb検索エンジンは、上記の3つの条件のうち「カバレッジ」は満たしているが、「分類体系の存在」「信頼性」においては不十分である。1990年代のWebではYahoo!などのディレクトリサービスがこれらの条件を満たす役割を担っていたが、Web空間の情報量が爆発している現代においては人手によるディレクトリの作成が追いついていない状況である。

一方、図書館の代表的なWeb情報サービスであるOPACでは、「分類体系

の存在」「信頼性」の条件は満たしているが、「カバレッジ」については不十分である。思いついたキーワードをOPACに入力しても、何も情報が得られない場合はかなり多い。

オンライン百科事典Wikipediaは、この3つの条件を考慮したとき興味深い位置に存在する。誰でも編集が可能であることから「信頼性」については課題があるが、「カバレッジ」「分類体系の存在」の条件は満たしていることから、Webと図書館システムの間のギャップを埋める架け橋として利用できる可能性がある。

国立国会図書館のWebサイトにて2009年5月から一般公開されているリサーチ・ナビの検索システムの開発においては、上記の考え方にもとづき、Wikipediaを情報資源の一つとして活用する方針をとった。具体的には、Wikipediaと国立国会図書館件名標目表（NDLSH）を統合的に使い、入力されたキーワードが図書館分類体系の中でどう位置づけられるのかを示すテーマグラフと呼ばれる図を自動的に生成する機能を実装している。また、リサーチ・ナビ検索エンジンが対象となる情報資源は国立国会図書館が公開しているものが中心であるが、ユーザインタフェースにはWebで汎用的に利用されている技術を積極的に活用している。主なものを以下に挙げる。

- タブデザインの採用

限りのある画面領域を有効に活用するため、画面内にタブパネルを配置している。ユーザは5つのタブ（「すべて」「調べ方」「本・サイト」「キーワード」「百科事典」）を選択することで表示内容を切り替えることができる。検索結果は、ユーザの用途に沿ってそれぞれのタブ内に分類して表示される（図4）。また、「すべて」タブには検索結果のサマリを表示させており、ユーザに対するナビゲーションの役割を果たすようにしている。

- Web APIの活用とマッシュアップ

リサーチ・ナビ検索システムが対象とする主なコンテンツは、PORTA（国立国会図書館デジタルアーカイブポータル）のAPIを経由して得ている。これにより、国立国会図書館が所有する多様なコンテンツへの対応を少ない工数で実現することができた。また、PORTA APIと書籍販売サイトが提供するWeb APIを連携させ、検索にヒットした書籍の書影を表示したりレビューを参照したりする機能も実装しており、内部職員向けに試験運用している。

- Ajax技術によるサーバとの非同期通信

リサーチ・ナビ検索システムはWeb APIの利用によって多様なコンテンツへの対応を実現しているが、コンテンツによっては検索にかかる時間にはばらつきがあるため、検索キーワードによってはレスポンスの悪化を招く可能性がある。そこで、ユーザインタフェースにAjax技術を採用し、Webブラウザからコンテンツ毎にサーバへの非同期通信を行わせるシステム構成とした。これにより、検索結果が得られたコンテンツから順次画面に表示させることが可能となった。

- ブラウザ側での検索履歴保持
ユーザによる試行錯誤を支援するため、ユーザが過去に入力した検索キーワードをブラウザに保持しておき、プルダウンリストによって簡単に呼び出せる機能を実装した。
- CMS（コンテンツ管理システム）とのシームレスな連携
コンテンツ管理システム Movable Type によって運用されているリサーチ・ナビ本体とハイパーリンクにより連携している。

本稿では、まず2節でリサーチ・ナビ検索システムの開発の背景にある考え方について論じ、つづいて3節においてシステムの全体構成と技術的なポイントについて詳しく解説する。4節では、リサーチ・ナビ検索システム独自の特徴であるテーマグラフのアルゴリズムについて述べる。また、開発にあたって工夫した点について5節で述べる。最後に6節においてまとめを述べる。

2. リサーチ・ナビ検索システム開発の背景

本節では、Web情報資源と図書館情報資源の特徴を比較し、利用者へのテーマ推薦を実現するための解決策について論じる。

2.1 Web情報資源と図書館情報資源のアクセス手段の現状

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、国立国会図書館分類表（NDLC）、日本十進分類法（NDC）、国立国会図書館件名標目表（NDLSH）や基本件名標目表（BSH）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。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人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。結果として、Web上に存在するオントロジーはブラウジングを便利にするための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、図書館の世界で必要な情報にアクセスするための代表的な手段としてはOPACが存在するが、入力したキーワードに対して何もヒットしないケースはかなり多い。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¹⁾が、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²⁾。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行われている³⁾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。

2. 2 タクソノミーとフォークソノミー

図書館で利用されている分類体系には、NDLC（国立国会図書館分類表）やNDC（日本十進分類法）に代表される図書分類法と、NDLC、LCSH（米国議会図書館件名標目表）、BSHなどの件名標目表が存在する。これらの分類体系は、多くの場合図2左のような構造をもっている。すなわち、頂点となる項目から枝分かれしていく木構造となっている。木構造においては、頂点以外の項目はただ1つの親をもっている。また、親項目がさらに親項目をもつことから、多数の階層を構成している。木構造をなす分類体系のことを、本稿ではタクソノミー（Taxonomy）と呼ぶこととする。

一方で、ここ数年Web技術を活用することで膨大な知識を多人数の共同作

タクソノミー

- トップダウン的な分類
- ひとつの上位概念
- 多数の階層

フォークソノミー

- ボトムアップ的な分類
- 複数の上位概念
- 項目とタグの2階層

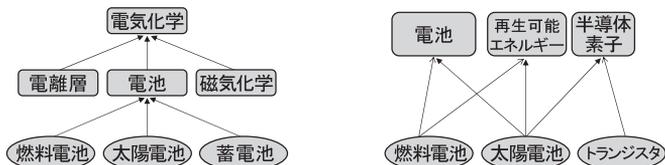


図2 タクソノミーとフォークソノミー

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。その代表例としてフォークソノミー (Folksonomy) と呼ばれるサービスが存在する。例えばFlickr⁴⁾、delicious⁵⁾、YouTube⁶⁾、はてな⁷⁾などが挙げられる。フォークソノミーとはサービスの一般利用者がコンテンツ (Web ページ、画像、動画など) に自身の視点でタグを付与できる仕組みであり、一般的には図2右のような構造をしている。フォークソノミーの構造においては、各コンテンツは複数の親を持っている。また、フォークソノミーは項目 (コンテンツ) とタグの2階層のみから構成されている。

従来、大量の情報を分類する手法は、タクソノミー的な分類体系によるものがほとんどであった。タクソノミー的な分類体系は分類の専門家によってトップダウン的に管理されており、一貫性のある分類がなされるという長所を持っている。ただし、分類の多様な観点を反映しにくい、語彙数が少ないという短所もある。フォークソノミーは、タクソノミーと対照的な特徴をもっている。多数の参加者がそれぞれの視点からボトムアップ的に分類を作っていくことから、多様な観点からの分類がなされる、多くの語彙をカバーしているという長所をもっているが、一貫性のない分類がなされてしまうという短所もある。

このように、大量の情報・資料を分類する方法論としては、タクソノミーとフォークソノミーの2つの考え方がある。従来はタクソノミーとフォークソノミーの2つの世界はお互いに独立して存在しており、両者の連携はあまり考慮されてこなかった。

Wikipediaの分類体系である「Wikipediaカテゴリ」は、タクソノミーとフ

オークソミーの両者の特徴を持ち合わせている（図3）。例えばWikipedia日本語版では、「価格」という項目に「マーケティング」「経済学」「市場」などのカテゴリが付与されているが、これらのカテゴリはWikipediaの編集時に自由に付与することができ、フォークソミーの特徴を有している。一方で、カテゴリにさらに親カテゴリが付与されて多数の階層をなしている点はタクソミーの特徴を有している。また、最上位層に属する「社会」「社会科学」「経済」「労働」「産業」などのカテゴリは1つもしくは少数の親カテゴリをもっており、図書分類体系ときわめて類似した構造をしている。さらにWikipediaは日本語版ですでに60万項目を超える記事が存在しており、カバレッジの面からも優れている。Wikipediaの分類体系構造はタクソミーとフォークソミーの2つの世界の橋渡しとなる可能性を持ち合わせている。

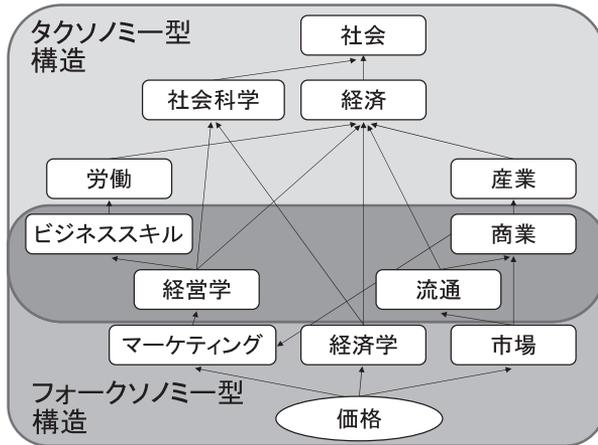


図3 Wikipediaカテゴリの構造

2. 3 解決策

これまで述べてきたように、利用者へのテーマ推薦を実現する上で、Web情報資源と図書館情報資源にはそれぞれ課題がある。Web情報資源には、多人数による知識の共有を可能とする技術をベースとした多様で便利なアクセス手段が提供されている反面、テーマ推薦を実現するという観点では十分な組織化がなされていない。一方、図書館情報資源にはテーマ推薦に適した厳密な分類体系が存在するが、Webのインタフェースに慣れ親しんだ人々にと

ってはアクセス手段が十分とはいえない。この課題を解決するために、リサーチ・ナビ検索システムの設計においては、Webで汎用的に利用されている技術を可能な限り取り入れつつ、図書館の分類体系をシステムに取り込むためにWikipediaを媒体として活用する方針をとった。

3. システムの構成

本節では、まずリサーチ・ナビ検索システムのユーザインタフェースを示した上で、ユーザインタフェースに表示されるそれぞれのコンテンツが導き出される仕組みを解説する。

3. 1 ユーザインタフェースのデザイン

リサーチ・ナビ検索システムのユーザインタフェース (UI) を図4に示す。UIの部品配置は一般的なWebサービスに準拠しており、上部に検索ボックス、下部に検索結果表示エリアを配置している。それぞれの部品の役割を以下に示す。

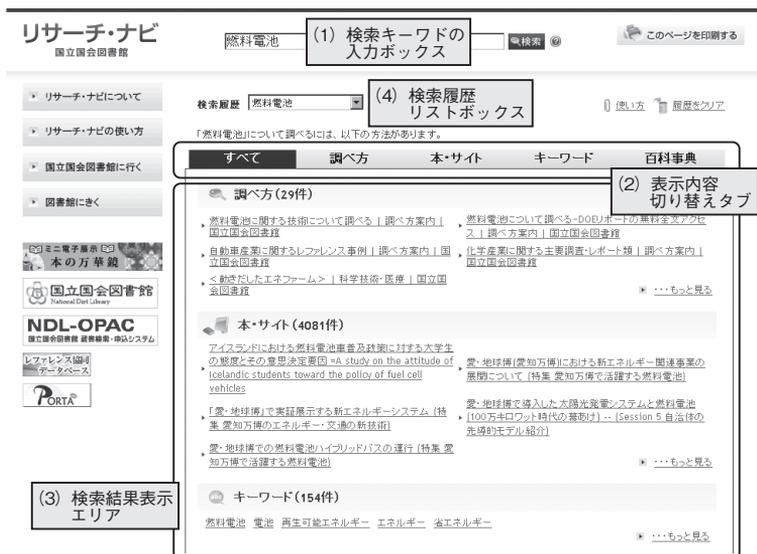


図4 リサーチ・ナビのユーザインタフェース (「すべて」タブ)

(1) 検索キーワードの入力ボックス

ユーザがこのボックスにキーワードを入力すると、検索サーバにおいて検索処理が行われる。その結果は(3)の検索結果表示エリアに表示される。

(2) 表示内容切り替えタブ

ユーザの目的に応じて検索結果表示エリアの内容を切り替えることができるようにしている。現在、「すべて」「調べ方」「本・サイト」「キーワード」「百科事典」の5種類のタブを設けている。それぞれのタブの役割については次節で説明する。

(3) 検索結果表示エリア

検索キーワードと選択されたタブに応じて、検索結果アイテムのハイパーリンクのリストを見やすく整理して表示する。ハイパーリンクをクリックすると、当該アイテムを表示するウィンドウが開く。ハイパーリンクにはキーワード入力と同様の役割を果たすものもある(たとえば「燃料電池」で検索した場合に「キーワード」タブ内に関連語として表示される「再生可能エネルギー」をクリックすると、「再生可能エネルギー」で検索した結果の表示画面に遷移する)。

エリアの広さは検索結果アイテム数によって縦方向に伸縮する。アイテムが多数の場合には複数のページに分割してページ遷移ボタンを下部に表示し、表示エリアが過度に縦長にならないようにしている。

(4) 検索履歴リストボックス

ユーザが以前探したキーワードがプルダウンリスト内に格納されている。任意のキーワードを選択すると、画面状態をそのキーワードによる検索結果に復帰させることができる。

3.2 各タブのコンテンツ

リサーチ・ナビ検索システムの画面には、図4に示すように5つのタブ(「すべて」「調べ方」「本・サイト」「キーワード」「百科事典」)が設けられている。以下では、「燃料電池」というキーワードで検索を行った場合を例として、それぞれのタブに表示されるコンテンツについて説明する。

● 「すべて」タブ(図4・図5)

他の4つのタブに表示されている検索結果の抜粋が表示される。タブ上方には、「調べ方」「本・サイト」「キーワード」「百科事典」タブの検索結果のうち、それぞれ5件が表示されている。抜粋の表示により、ユー

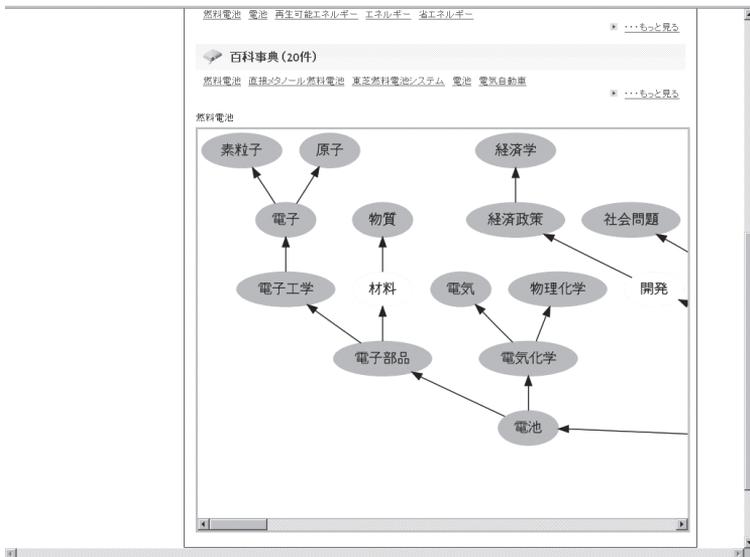


図5 「すべて」タブの表示コンテンツ（下）

ずはどのタブに自分の求めている情報が存在するのかを類推することができる。5件より多くの検索結果を確認したい場合には、「もっと見る」リンクをクリックすることで、対応するタブに遷移することができる。また、タブ下方にはテーマグラフ（4.2節で後述）が表示されており、「燃料電池」というキーワードと図書館分類の関連を俯瞰することが可能である。

- 「調べ方」タブ（図6）

タブ上方には、リサーチ・ナビCMS（Movable Type）に掲載されているコンテンツの検索結果が「調べ方ガイド」として表示される。ここでは、「燃料電池」について調べる手掛かりを簡潔にまとめたページがいくつかヒットしている。

タブ下方にはレファレンス協同データベースに登録されているレファレンス事例からのPORTAによる検索結果が表示される。

- 「本・サイト」タブ（図7）

タブ上方には、「燃料電池」をメタデータに含む国立国会図書館の蔵書の検索結果がPORTA APIを通じて表示される。また、館内の職員向けに

国立国会図書館に行く

図書館にきく

電子書籍の
本の万華鏡

国立国会図書館
National Diet Library

NDL-OPAC
国立国会図書館 検索検索・申込システム

レファレンス協同
データベース

PORTA

すべて 調べ方 **本・サイト** キーワード 百科事典

🔍 調べ方 (29件)

① 調べ方ガイド (25件)

- 燃料電池に関する技術について調べろ | 調べ方案内 | 燃料電池について調べろ-DOEレポートの無料全文アクセス | 調べ方案内 | 国立国会図書館
- 自動車産業に関するレファレンス 事例 | 調べ方案内 | 化学産業に関する主要資料レポート調べ方案内 | 国立国会図書館
- くわだたしエネファーム | 科学技術・医療 | 国立国会図書館
- ナノテクノロジー産業について調べろには | 調べ方案内 | 電気自動車(EV electric vehicle)について調べろ | 調べ方案内 | 国立国会図書館
- ナノテクノロジー産業に関する主要統計資料 | 調べ方案内 | 国立国会図書館
- 調べ方案内 | 産業技術・工学 | 科学技術・医療 | 国立国会図書館

前へ 1 2 3 次へ

② 調べもの事例: マニュアル (4件)

以下の論文を前題としてよめか? Author(s): Yao, S.-C.; Fedder, G. S.; Tang, X.; Hsieh, C.-C.; Alyousef, Y. M.; Wadlmer, M. L.; Jaron, C. H. Title: Thermo-fluids considerations in the development of a silicon-based micro-scale direct methanol fuel cell Page: 171-182 Conference: 2004 Jun - Bled, Slovenia Name: International thermal science seminar; Thermal sciences 2004 Publisher: Reston; ASME, 2004 Standard No: ISBN: 9619139305 (casadi) Source Ed: Bergles, A. E.

レファレンス協同データベース

2020年6月までの「低炭素自動車の世界生産台数予測データ」が掲載されている資料があります。期にハイブリッド車・電気自動車・燃料電池車の単独データを除いています。

レファレンス協同データベース

- コージェネレーション(熱電併給)の市場動向を調べています。レファレンス協同データベース
- 電気自動車などのエコカーに関する調査レファレンス協同データベース

前へ 1 次へ

図6 「調べ方」タブの表示コンテンツ

国立国会図書館に行く

図書館にきく

電子書籍の
本の万華鏡

国立国会図書館
National Diet Library

NDL-OPAC
国立国会図書館 検索検索・申込システム

レファレンス協同
データベース

PORTA

すべて 調べ方 **本・サイト** キーワード 百科事典

📖 本 (4081件)

国立国会図書館にある本 (4081件) →アジアの事はアジア言語OPACでさがしてください

本の情報からさがす (3710件)

アイスランドにおける燃料電池普及と政策に対する学生の態度とその意思決定要因 -A study on the attitude of Icelandic students toward the policy of fuel cell vehicles
NDL雑誌記事索引

「愛・地球博」で実証展示する新エネルギーシステム [特集 愛知万博のエネルギー - 交通の新技術]
NDL雑誌記事索引

「愛・地球博」での燃料電池ハイブリッドバスの運行 [特集 愛知万博で活躍する燃料電池]
NDL雑誌記事索引

- 「愛・地球博」における燃料電池の実証展示
NDL雑誌記事索引
- 「愛・地球博」ワンダーサーカス電力屋 - 活躍する新エネルギーシステム [特集 愛知万博で活躍する燃料電池]
NDL雑誌記事索引

内容情報からさがす (371件)

- 燃料電池自動車のすべて / 世界の潮流 / 2005.7 国立国会図書館の蔵書の目次
- 燃料電池市場の現状と将来展望 / 2006.4 国立国会図書館の蔵書の目次

前へ 1 2 3 4 5 次へ

図7 「本・サイト」タブの表示コンテンツ

は書籍販売サイトのAPIも併せて利用し、ISBNで結びつく書籍の書影や読者レビューを表示する機能も試験的に実装している。

タブ下方には「燃料電池」を目次中に含む書籍が目次データベースより検索・表示される。

● 「キーワード」タブ (図8・図9)

タブ上方には、分類自動導出アルゴリズム (4.1節にて後述) によって導出された件名 (NDLSHに登録されているもの) がスコア順に表示される。図の例では、「電池」「再生可能エネルギー」「エネルギー政策」など、燃料電池に深く関係するトピックが抽出されている。それぞれの件名には、対応する日本十進分類法新訂第9版 (NDC9) のコード (「燃料電池」であれば「572.1」) と関連件名が併せて表示される。NDC9のコードをクリックすると、PORTAによる検索画面に遷移する。

タブ下方には、テーマグラフと関連キーワードのタグクラウドが表示される。タグクラウドには、Wikipediaにおいて検索キーワードと関連度の高いキーワードが表示される。文字の大きさは入力キーワードとのベクトル空間におけるコサイン類似度を反映させており、関連度の高いキーワードが大きく表示される。任意のタグをクリックすると、タグのキーワードを新たな検索キーワードとして再検索が行われる。



図8 「キーワード」タブの表示コンテンツ (上)

● 「百科事典」タブ (図10)

Wikipediaの本文中に検索キーワードが出現する項目が検索結果として表示される。

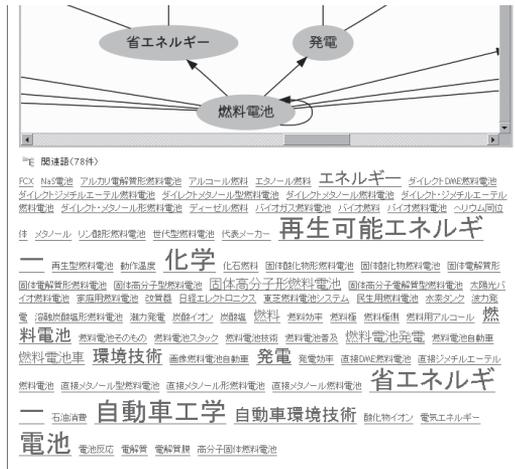


図9 「キーワード」タブの表示コンテンツ (下)



図10 「百科事典」タブの表示コンテンツ

3. 3 リサーチ・ナビ検索システムの構成

リサーチ・ナビ検索システムは、図11に示すように、ユーザインタフェースとしてのWebブラウザとバックエンドサーバから構成されている。

3. 3. 1 フレームワーク

リサーチ・ナビ検索システムは、Google Web Toolkitをアプリケーションフレームワークとして採用している。Google Web Toolkitは、Ajaxアプリケーションの効率的な開発に特化したツールキットであり、Java言語によりWebブラウザとバックエンドサーバの双方の機能を統合的に開発することができる。Java言語によって記述されたWebブラウザ側のロジックは、コンパイル時に自動的にJavaScript言語に変換されるため、特別なプラグイン等のインストールを必要とすることなく一般的なWebブラウザで動作する。また、システム開発時にはJava言語によるJavaScriptのエミュレーション機能が利用できるため、デバッグ等が容易である。

システムの設計においては、Ajaxアプリケーションの大きな特徴である非同期通信を積極的に活用する方針をとった。図11に示すように、本システ

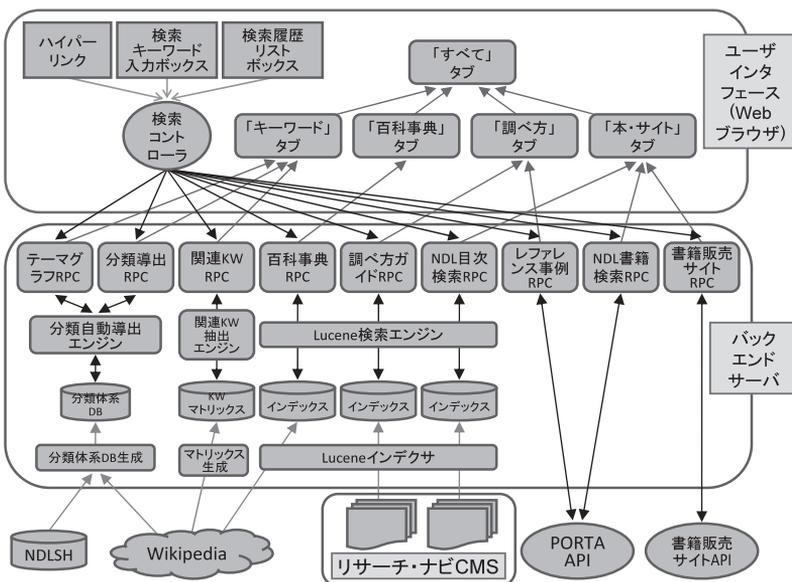


図11 リサーチ・ナビ検索システムの構成

ムは外部のWeb API (PORTA、書籍販売サイト) を含む多数の情報資源に同時にアクセスする必要がある。もし、全ての情報資源の検索処理が完了した後に検索結果画面をレンダリングする構成をとると、ユーザが検索キーワードを入力してから結果が表示されるまで長い間待たされてしまうことがある。Webブラウザ側から直接各々の情報資源への問い合わせを非同期通信により行うことで、検索処理が完了した情報資源から順次画面に表示させ、ユーザのストレスを最小限に抑えることが可能となった。

3. 3. 2 ユーザインタフェース

リサーチ・ナビ検索システムのユーザインタフェースは汎用のWebブラウザによって動作する。リサーチ・ナビのWebサイト (<http://rnavi.ndl.go.jp/>) にアクセスし、検索ボックスにキーワードを入力することで、リサーチ・ナビ検索システムのユーザインタフェースを構成するHTMLファイル、スタイルシートファイル、画像ファイル、JavaScriptファイルがWebブラウザに自動的に読み込まれる。読み込み後、JavaScriptが自動実行され、ユーザインタフェース内に検索キーワード入力ボックス、検索履歴リストボックス、タブなどの部品を配置する。

図11中の検索コントローラは、JavaScriptによって実装されたモジュールであり、検索の実行において重要な役割を果たす。ユーザが検索要求処理 (検索キーワード入力ボックスへの入力、ハイパーリンクのクリック、検索履歴リストボックスからのキーワード選択のいずれか) を行うと、検索コントローラが起動し、各々の情報資源に対応するRPCメソッド (次節にて述べる) への非同期通信を実行する。非同期通信のコールバックメソッドには、各タブへの検索結果の埋め込み処理が実装されている。これにより、検索処理が終了した情報資源から順次画面に反映されるようになっている。

画面中に配置されているタブのうち、「すべて」タブには特別な役割を持たせている。「すべて」以外の4つのタブ (「調べ方」「本・サイト」「キーワード」「百科事典」) に表示された検索結果のうち上位にヒットしたものが、抜粋として「すべて」タブに転送される。

3. 3. 3 バックエンドサーバ

リサーチ・ナビ検索システムのバックエンドサーバは、Javaサーブレットとして実装され、Webコンテナ (Apache Tomcat) 内で動作している。Javaサーブレットには各々の情報資源に対応したRPCメソッドが実装されてい

る。RPC (Remote Procedure Call) とは、別のコンピュータ上にあるメソッドを遠隔実行することを可能とする技術を指す。各々のRPCメソッドは、ユーザインタフェース側から簡易な手続きにより呼び出すことができる。

各々のRPCメソッドの処理は、大まかに以下の4種類に分類される。

1. 外部のWeb APIを直接呼び出すもの

「NDL書籍検索RPC」と「レファレンス事例RPC」は、国立国会図書館PORTAのAPIを直接呼び出している。また、書籍販売サイトのAPIを呼び出すRPCは、館内の職員向けのバージョンにおいて試験的に利用されている。

2. バックエンドサーバ内部に置かれた検索インデックスを利用するもの

リサーチ・ナビCMS (Movable Type) に蓄積されているコンテンツ、およびWikipediaの全文データに関しては、全文検索ソフトウェアLuceneによって実装されたインデクサと検索モジュールを利用する。RPCは、定期的にクローリングされたデータから生成されるインデックスを参照して検索キーワードにヒットする文書群を抽出し、ユーザインタフェース側に渡す。

3. 関連キーワード抽出エンジンを利用するもの

関連キーワード抽出エンジンは、Wikipediaから生成されたキーワード = 文書マトリックス (行列) を参照して関連キーワード群 (検索キーワードと近い出現パターンをもつもの) を導出する。導出される関連キーワードには、入力キーワードとのベクトル空間におけるコサイン類似度を反映した関連度スコアが付与される。

4. 分類自動導出エンジンを利用するもの

分類導出RPCは、4. 1節において後述する分類自動導出アルゴリズムによって導出される件名のリストをユーザインタフェース側に渡す。また、テーマグラフRPCは、同じアルゴリズムによって生成されるテーマグラフの画像ファイルを生成してユーザインタフェース側に渡す。

4. Wikipediaと件名標目表を統合的に活用した分類自動導出

本節では、図11に示したリサーチ・ナビ検索システムの構成図における「分類自動導出エンジン」のアルゴリズムの概要について簡単に解説する。また、分類自動導出エンジンが出力するテーマグラフと呼ばれるグラフの活用法について説明する。

4. 1 アルゴリズムの概要

前述したように、Wikipediaは「テーマ推薦の要件」と「カテゴリの構造」の観点から見たときに、きわめてユニークな特徴をもっている。この特徴をうまく用いて、情報探索の出発点としてWikipediaを利用し、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方法を提案している。

図12に導出アルゴリズムの概要を示す。まず、Wikipediaカテゴリの構造について説明する。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。

次に、Wikipediaカテゴリと図書館の分類体系の対応付けについて説明する。Wikipediaカテゴリと図書館の件名の間には、カテゴリ名が一致するものが存在する。図12では、「経済史」「災害」「地震」が一致している。

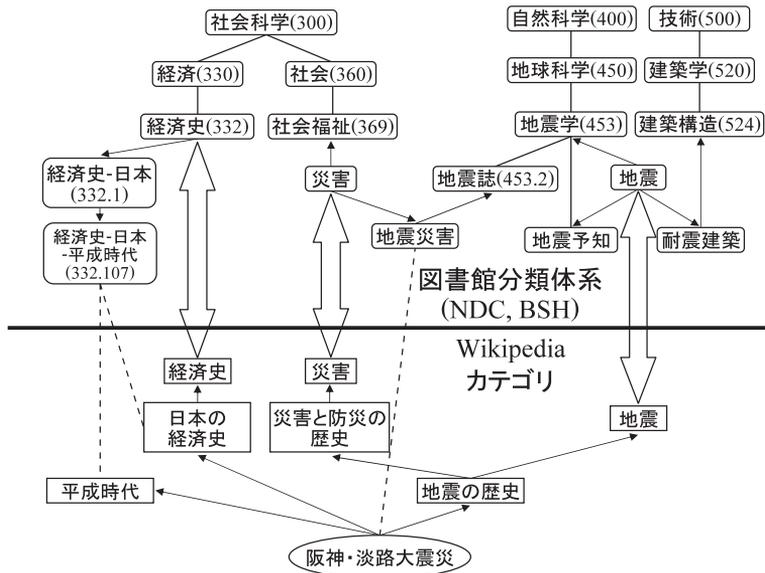


図12 導入アルゴリズムの概要

よって、「阪神・淡路大震災」につながるカテゴリが構成する有向グラフの構造を再帰アルゴリズムによってたどることで、「阪神・淡路大震災」に関連する分類を自動的に導出することができる。リサーチ・ナビでは、グラフのエッジに対する重みスコアをノード間の文字列類似度によって定義し、ビームサーチによって重みスコアが相対的に大きい件名を絞り込むアルゴリズムを採用している。図8の例では、「燃料電池」というキーワードに対して、「燃料電池」「電池」「再生可能エネルギー」「エネルギー」「省エネルギー」「発電」などの件名がスコアの大きな件名として導出されている。

4. 2 テーマグラフによる分類の俯瞰

上記に示したように、Wikipediaカテゴリ構造と図書館件名標目表を活用することで、任意のWikipediaの項目に対して、関連する件名を自動的に導出することができる。リサーチ・ナビ検索システムでは、この導出の過程をグラフ構造として自動的に描画する機能を実装している。このグラフ構造は、与えられたWikipediaの項目がどのようなテーマを有しているのかを示していることから、テーマグラフと名付けている。

図13に、「燃料電池」という検索キーワードから導き出されたテーマグラフの例を示す。テーマグラフの内容を考察していくことで、「燃料電池」という概念がどのようなテーマと関連を持っているのかを知ることができる。例えば、「再生可能エネルギー」「循環型社会」「環境問題」などの件名と「燃料電池」との関連性を考察すると、「燃料電池は環境問題解決の切り札として注目されている」という背景を見いだすことが可能である。

PORTAではさまざまな有用な情報資源が利用可能となっているが、全ての情報資源を検索対象とすると、ユーザの目的に合致しないコンテンツも大量にヒットしてしまうことが多い。そこで、リサーチ・ナビの開発目的である「利用者への調べ案内」に合致する情報資源である「レファレンス協同データベース」などを重視してヒットさせるようにした。

- Wikipediaの位置づけ

Wikipediaは調べものためのツールとして有用性をもっている反面、信頼性においてさまざまな課題が指摘されているため、利用者に最終的な情報資源として提示することは必ずしも望ましいとはいえない。このため、Wikipediaはあくまでも多様な情報資源の一つであるという位置づけを利用者に示すデザインとした。具体的には、Wikipediaへのハイパーリンクは「百科事典」タブ内に集約し、利用者が過度にWikipediaに誘導されないように配慮した。

- リサーチ・ナビCMSとのシームレスな統合

リサーチ・ナビ独自のコンテンツはCMS（Movable Type）によって管理・公開されているが、リサーチ・ナビCMSとリサーチ・ナビ検索システムに共通の画面デザインを採用することで、利用者が両者の違いを意識することなく検索システムを利用することができるように配慮した。

6. おわりに

リサーチ・ナビ検索システムの開発においては、汎用的に利用されているWeb技術を活用することによって、国立国会図書館がもつさまざまな情報資源への新たなアクセス手段を実現することを目標とした。また、Wikipediaを「情報の架け橋」として活用し、さまざまなキーワードから関連する件名を自動的に導出する機能を実装することで、情報探しのためのヒントを利用者に与えるようにした。

今後は、利用者からのフィードバックやアクセスログの分析を行うことによってシステムの有用性を評価するとともに、さらに使い勝手の向上を目指す予定である。

謝辞 Wikipediaと件名標目表を統合的に活用した分類自動導出の手法の研究に関しては、文部科学省科学研究補助金若手研究（B）「Web上の情報資源と図書館情報資源を統合的に利用する情報探索支援システム」（課題番号18700134）の支援を受けた。また、本研究は東京大学情報基盤センターの中川裕志教授、東京電機大学未来科学部の増田英孝准教授・坂井哲氏の協力を得て遂行された。

リサーチ・ナビ検索システムの開発の過程では、国立国会図書館主題情報部参考企画課の方々には多くの有益なアドバイスをいただいた。また、開発にあたっては情報サービス関連各種企業の方々には尽力していただいた。

参考文献

- 1) 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, 2006, pp. 103-107.
- 2) 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, pp. 1-5.
- 3) Cohen, Laura B. and Still, 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, 1999, pp.275-289.
- 4) <http://www.flickr.com/>
- 5) <http://delicious.com/>
- 6) <http://www.youtube.com/>
- 7) <http://www.hatena.ne.jp/>

(きよた ようじ 東京大学情報基盤センター学術情報研究部門助教)