

## 情報知識学会関西支部2009年度第1回（通算第10回）研究会報告

日時：7月18日（土）14:30～17:00

会場：大阪科学技術センター

論題：図書館分類体系と Wikipedia を統合した情報探索支援システムの開発

発表者：清田陽司氏（東京大学情報基盤センター）

共催：日本図書館研究会情報組織化研究グループ

出席：27名

内容：

発表者の自然言語処理技術研究をもとに、情報探索支援システム Littel Navigator（株式会社リッテル）が開発・運用されている。今回はシステムの機能・概要にとどまらず、基盤となる考え方に踏み込んだ発表があった。

### 1. Littel Navigator とは？

- ・本システムは、キーワードからピンポイントに資料・情報を検索するというよりは、「情報探索のヒント」を検索・提示することをコンセプトとしている。「パスファインダー自動生成システム」ともいえる。
- ・与えられたキーワードから、件名標目・書架分類・レファレンスブック・Web サイトなど様々なリソースを統合検索して結果を示す。また、Wikipedia と件名標目等の情報を組み合わせて関連概念を図示した「テーマグラフ」を自動生成し、曖昧な質問から具体的な探し方へのナビゲーションを提供する。OPAC や Google Scholar など、他の検索システムとのリンク機能も備えている。
- ・いくつかの大学図書館に導入実績がある。また、国立国会図書館が2009年5月に導入した「リサーチ・ナビ」にもコンポーネントとして導入されている。図書館以外に、マーケティング用途での利用実績もある。

### 2. 図書館の世界と Web の世界

- ・現在の Web 検索エンジンは事実を単純に問う質問には威力を発揮するが、テーマが曖昧なニーズには応えにくい。テーマを推薦（レコメンデーション）できる検索システムが求められる。
- ・適切なテーマ推薦を行うには、カバレッジの広さ、一定の組織化、情報の信頼性の担保、が必要である。Web は情報検索手段として、非常に広いカバレッジという利点を持つが、信頼性の面での判断基準を提供してくれない、フロー情報中心でストック情報は残りにくい、という欠点も持っている。
- ・また、Web 情報の組織化は、ソーシャルタギングなど一般利用者（情報消費者）による視点が重視され、ブラウジング指向に偏りがちである。情報散策（ブラウジング）には向いているが情報探索（知的探求）には十分に答えられない。
- ・Web のカバレッジの広さは、情報探索の足がかりとしては非常に有用である。一方で、図書館の持つ情報はストック性と信頼性に優れており、両者を統合できれば有効なテーマ推薦システムを構築しうる。

・両者を橋渡しするものとして、Wikipedia に注目した。Wikipedia は、Web 上に百科事典を構築するという明確なミッションを持ち、日本語版で60万件以上という膨大な項目数を擁している。また、「カテゴリ」という形で項目の組織化への取組みを行っていることは、他の Web サイトにはない大きな特色である。

### 3. 2つの分類パラダイム

- ・図書館的な分類体系を「オントロジー型」と呼ぶ。その特徴は、基本的には各概念が一つの上位概念を持つ、トップダウン的な構造である。
- ・個々の利用者が自身の視点でコンテンツにタグを付与する「フォークソノミー」が2005年ごろから普及してきた。Web 的な分類体系ととらえることができる。「フォークソノミー型」分類体系の特徴は、各概念が複数の上位概念を持つことを通常とみなす、ボトムアップ的な構造である。
- ・Wikipedia の「カテゴリ」は、フォークソノミー型のゆるやかな階層構造を有している。ただし、トップレベルに近い部分ではオントロジー型に近い形となっている。

### 4. 2つの体系の統合による分類自動導出

- ・図書館の分類体系を Wikipedia で拡張することを考える。図書館の件名表・分類表と Wikipedia（カテゴリ）には、概念名が一致するものが相当数ある。情報探索のスタート地点では Web 上の情報資源を主対象とするが、Wikipedia を結節点として図書館の分類体系へと誘導すれば、そこに結びつけられた図書館の多様な情報資源（入門書やレファレンスブックなど）に導くことができる。

### 5. Wikipedia についての考察とアイデア

- ・Wikipedia の全項目データが、XML 形式でダウンロードできる。膨大なデータであり、コーパスとして利用できる。また、表記揺れ辞書（リダイレクト等を用いる）、シソーラス（カテゴリを用いる）、固有表現辞書などとしても用いられている。
- ・Wikipedia は、誰でも編集できるが一定の秩序維持の仕組みを持つ。また、テンプレート等による半定型性、外部情報リソースとの連携、カテゴリによる項目の組織化なども特徴である。組織化については、フォークソノミー型とオントロジー型の中間の性格を持っている。
- ・Wikipedia のデータ処理を容易に行うため、データ解析ツール Wik-IE を作成して公開している。

### 6. おわりに

- ・今後の展開としては、LCSH を視野に入れた英語版など他言語版の開発、MeSH など分野別シソーラスの活用、シラバスとの連携など情報リテラシー教育の題材としての仕組み、などを考えている。
- ・「図書館か Web か」という二項対立ではなく、両者の役割を踏まえた見方が大事である。

参考：株式会社リッテル

<http://www.littel.co.jp/>