

「高度メディア社会の生活情報技術」
平成13年度採択研究代表者

高野 明彦

(国立情報学研究所 教授)

「連想に基づく情報空間との対話技術」

1. 研究実施の概要

本研究の目的は、「連想の情報学」の基礎研究を行い、文書情報空間に奥行きと安心感を与える対話技術を構築することである。本研究は5年計画で実施する予定であり、最終的には、ユーザが理解度に応じて平易な入門的解説から高度な専門的記述までを自在に渡り歩く手段を提供し、電子化された情報空間での新しい「読書環境」の実現を目指す。研究成果は、多様な書籍コンテンツに適用して有効性や実用性を評価する。

本年度は、汎用連想計算エンジンGETAを活用した連想検索機能を発展させ、複数の異なるデータベース間での連想的探索を支援する技術を追究し研究を進めてきた。具体的には、2004年4月に文化庁と総務省とが公開した「文化遺産オンライン（試行版）」(<http://bunka.nii.ac.jp/>)のシステム構築において技術協力を行った。この試作版は、単にさまざまな文化遺産に関する情報を検索できるだけでなく、検索結果として得られる文化財の一覧から類似する文化財を連想的に検索することや、写真を見ながら気ままな連想で文化財サーフィンを楽しむことができる。さらに、文化財の詳細ページから関連図書を検索するWebcat Plusへのワープ機能が付いており、我々の提唱する連想的情報アクセス技術を利用したクロスDB連想検索の実現例となった。

GETAの連想計算による連想の品質は、原理的にコーパスの統計的性質の影響を受けやすく、人間には理解しがたい結果が返ることも珍しくない。この機械的連想の限界を打破するための1つの手法として、人間が知識や連想を駆使して作成した静的なコンテンツを利用する方法を考案して、まったく新しい読書ガイドサービスを構築した。具体的には、新書7000冊を人手で約1000個のテーマ（クラスター）に分類しておき、それらを動的に連想検索してユーザに提示する「新書マップ」（<http://shinshomap.info/>）というサービスを公開した。情報の信頼性を確保しつつ、ユーザの要求に柔軟に対応できる次世代ディレクトリサービス技術と考えている。

今後は、本研究プロジェクトの主要なターゲットである書籍コンテンツを対象に、新しい連想的情報アクセス環境（読書環境）を追究し、その応用を目指す。

2. 研究実施内容

本研究の成果を応用して、これまでに次の3つの公開情報サービスが構築された。それ

それぞれの概要について紹介する。いずれも全国紙やNHKニュースなどで大きく取り上げられ、毎日多くのページビューを稼ぐ人気サービスとなっている。

[NII図書情報検索サービスWebcat Plus]

2002年10月、国立情報学研究所は、江戸期以前から最新刊までの245万タイトルの和書情報が高速に検索できる図書情報検索サービス **Webcat Plus** (図1) を立ち上げた (<http://webcatplus.nii.ac.jp/>)。全国の1000館を超える大学や研究機関の図書館の収蔵情報と書籍の目次や概要が読めるポータルサイトとして、毎日3万ページビューを稼ぐ人気サイトとなっている。このサイトの特徴は、自由な長文や気に入った書籍から関連する書籍を検索できる連想検索機能だが、ここにGETAを活用して実現した我々の連想的情報アクセス技術が採用されている。検索結果を要約する関連ワードを提示して、より柔軟な関連性フィードバックを可能にしている。



図1. Webcat Plusと文化遺産オンライン (試行版)

[文化遺産オンライン試行版]

2004年4月、文化庁と総務省は、全国の美術館や博物館などが収蔵する文化遺産情報を、誰でも手軽にインターネットで検索、閲覧ができる「文化遺産オンライン」試行版 (図1, <http://bunka.nii.ac.jp/>) を公開した。試行版では、東京国立博物館や徳川美術館などが提供する約3000件のデータが検索できる。システム構築は国立情報学研究所が担当した。

このサイトのユニークな特徴は、検索結果として得られる文化財の一覧から気に入った文化財を選択して、それと類似の文化財を検索できる連想検索機能である。また、各文化財の詳細情報のページには、その作品と最も似ている作品5件の写真がいつも表示され、写真を見ながら気ままな連想で文化財サーフィンを楽しめる。ここにもGETAを活用して実現した我々の連想的情報アクセス技術が採用されている。

さらに、各文化財の詳細情報ページにあるWebcat Plusへのワープボタンにより、解説文等の情報を利用してWebcat Plusで関連書籍を検索することができる。これは我々の提唱する連想的情報アクセス技術によって初めて可能となるクロスDB連想検索のひとつの例となっている。電子情報空間との創造的な対話を、DBの垣根を越えて続けられる。

[新書マップ：テーマで探す新書ガイド]

2004年6月、連想検索を応用した新しい情報サービス「新書マップ」を立ち上げた(図2, <http://shinshomap.info/>)。現在流通している約7000冊の新書・選書が約1000のテーマに分類され、テーマごとに本のリストと読書案内を読むことができる。トップページで自由な長文やキーワードを与えて検索すると、それに関連する10テーマが選ばれて、星座図のように配置された「マップ」が現れる。円内のテーマノードをクリックすると本のリストが背表紙写真とともに表示される。円の周囲に配置された12個の関連語ノードをクリックすると、その語で再検索され、また新しいマップが返される。テーマのページは、拾われている本繋がりに関連テーマへリンクされている。本のリストはWebcat Plusにリンクされていて、目次や概要を読むこともできる。

基本は人手で丹精して作られた静的な1000個のテーマだが、それを連想計算でユーザの興味に応じて動的に集めてマップを作ることにより、無限の深みを演出できたと感じている。Flashを用いた動きのあるマップの見せ方と、書棚のアフォーダンスを想起させる背表紙写真の組み合わせにより、リアルな書棚では絶対できない「書棚ブラウザ」を可能にしている。情報の品質を保証しながら、連想機構を用いてユーザ要求への適応性を高める手法には普遍性があり、これこそが次世代ディレクトリサービス技術であると考えられる。



図2. 新書マップ：テーマで探す新書ガイド

情報の類似性に基づく検索法の有効性は多くの研究者により指摘されているが、それを1千万件規模の実データに適用している例は、我々の研究以外にはほとんど知られていない。PubMedの関連論文リンク (Related Articles) やGoogleの関連ページリンクは一見似ているが、いずれも個々のデータについて類似データをいくつか計算して静的に記憶しているものである。我々の連想計算のように、選択された複数のデータ (クラスタ) をクエリとして、それと類似するデータ群を動的に求めることはできない。検索者の発想や意図のコンテキストを的確に表現するには、データ群 (クラスタ) をクエリとする連想検索は必須であり、その意味で我々の連想検索技術は世界をリードしていると考えられる。

3. 研究実施体制

国立情報学研究所グループ

- ① 高野明彦 (国立情報学研究所ソフトウェア研究系 教授)
- ② 連想の情報学構築を担当

東京大学グループ

- ① 黒橋禎夫 (東京大学大学院情報理工学系研究科 助教授)
- ② 共起単語間の意味関係解析

東京工業大学グループ

- ① 岩山真 (東京工業大学精密工学研究所 助教授)
- ② 文書クラスタリング・パッセージ解析

日立中研グループ

- ① 丹羽芳樹 (日立製作所中央研究所 主任研究員)
- ② 連想計算実験環境の構築

4. 主な研究成果の発表

(1) 論文発表

- 高野明彦、「文化遺産オンライン」試験公開版システムの概要、文化庁月報、2004年7月号.
- 小池勇治、矢島匡人、高野明彦、絹川博之、Webサービスを利用した総合電子文書検索システム、FIT2004、Sept. 2004.
- 高野明彦、連想に基づく情報空間との対話技術、心理学ワールド27号、Oct. 2004.
- 高野明彦、西岡真吾、丹羽芳樹、連想に基づく情報アクセス技術—汎用連想計算エンジンGETAを用いて—、情報の科学と技術、Vol. 54, No. 12, pp. 634-639, Dec. 2004.