

「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、(1)連想を刺激して情報空間の内容的概観や情報の文脈（現在位置）の直観的把握を助ける「連想地図 (Associative Atlas of Cyberspace)」と、(2)記述レベルの異なる多層的な文書群を動的に関連づける「理解支援リンク」技術を開発して、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、異分野にまたがる参加研究者間の新しい連携の可能性を探るため、共通の研究対象として広い意味での情報技術関連書籍を選定して電子化を行った。特にXML化された7冊の辞典を対象にXML電子辞典利用環境の構築に着手した。編集意図の異なる辞典を連想的に結びつけることにより、斬新な電子リファレンス環境を提供できる見通しを得た。共用実験環境としてPCクラスタ装置を導入し、汎用連想計算エンジンGETAの並列処理版(第3版)をその上で稼働させた。現在、GETAの提供する連想計算は世界的に見ても最高水準にあるが、その実用性を実証するため、GETAを活用して連想検索機能を実装した。これは国立情報学研究所の新サービス「NII図書情報ナビゲータWebcat Plus (<http://webcatplus.nii.ac.jp>)」に採用され、250万件の図書情報を連想的に探索できるインタフェースとして注目された。今後は、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、それを利用する公開サービスの立ち上げを目指す。

2. 研究実施内容

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。専門性の異なる関連分野で活躍中の若手研究者を集め、各自の研究の成果を共通の対象データに適用することを通じて、狭い専門分野を越えた真に新しい研究連携の可能性を探るアプローチを取っている。具体的には、情報科学専門辞典1冊、大学講座シリーズ30冊、新書100冊という広い意味での情報技術関連

書籍を電子化し、共通の研究対象（共通コーパス）として設定している。

本年度は、共通コーパスを利用する最初のプロジェクトとして、XML化された7冊の辞典（岩波情報科学辞典、岩波ジュニア新書の小辞典6冊）を統合された一つの辞典として利用できるXML電子辞典システムの開発に着手した。情報、物理、化学、宇宙、生物の分野について、それぞれ異なる編集意図で作られた辞典を連想的に結びつけることにより、ユーザの興味や背景知識に応じた使いこなしを可能とする斬新な電子リファレンス環境構築への手がかりが得られた。本システムについては、次年度以降も拡張を続け、出版社が新しい電子リファレンスサービスを提供する際に利用可能な基盤ツールを目指して育てていきたい。



図1. NII 図書情報ナビゲータ Webcat Plus

研究プロジェクトに求心力を持たせるもう一つの試みとして、共同利用研究設備として、国立情報学研究所内にプロジェクト専用の研究用ドメインを確保し、そこに16ノード32CPU規模のPCクラスターを稼働させている。*DuaINAVI*やGETA、XML電子辞典システムなどの実用的基盤ツールが自由に使える共通実験環境が利用できる。

汎用連想計算エンジンGETA (<http://geta.ex.nii.ac.jp/>) の提供する大規模連想計算を利用して、連想検索エンジンを構築している。GETAの提供する連想計算は世界的に見ても最高水準にある。2002年10月には、NII 図書情報ナビゲータ Webcat Plus (<http://webcatplus.nii.ac.jp/>, 図1) という新サービスに我々の連想検索エンジンが

採用され、250万件の図書情報を連想的に探索できるインタフェースを提供して注目された。公開初日は1日で14万ページビューを記録するなど図書検索の人気サービスになりつつあるが、GETAの実用性を実証するショールームとして、今後も新機能を提供していきたい。GETAの新しい応用分野としては、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システムの試作を予定している。

本年度は、本研究課題の実質的な1年目に当たるが、GETAのオープンソース公開、Webcat Plusのサービス開始、XML電子辞典システムの試作など、研究成果に裏付けられた実用性の高い情報技術を発信できた。最新のGoogleの連想ランキング（「連想」で検索したページランキング）では、1位がWebcat Plus、4位がGETAとなっており、「連想の情報学」を標榜するに相応しい成果を上げつつある。

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（論文発表および特許出願）

（1）論文（原著論文）発表

なし

（2）特許出願

H14年度特許出願件数：0件（研究期間累積件数：0件）