

高野 明彦†
Akihiko Takano

1. はじめに

近年、百科事典や新聞を始め企業内文書まで、あらゆる文書情報の電子化が進み、それらの有効活用なしに充実した知的生活は考えられなくなっている。個人がアクセス可能な情報量は増え続け、すでに15億ページを越えたともいわれるWWWコンテンツは、まさに地球を覆う情報の大海となった。しかし、そこから意味のある情報を汲み取るために我々が現在利用できる情報技術は、せいぜい書誌情報検索や全文検索であり、あるいは予め人手で分類されたディレクトリ情報の利用である。自分が直面している問題解決のヒントや、自分の思考を深めるのに役立つ示唆はなかなか得られない。21世紀に求められる情報技術は、膨大な情報を活用して個人あるいはグループの問題解決能力を高め、思考や議論を深めるものでなければならない。

本講演では、この課題に対して「連想の情報学」というアプローチを示し、その実現のために我々が研究開発しているDualNAVIおよびGETAという2つのシステムを紹介する。また、これらの研究的蓄積をベースに最近着手した「連想に基づく情報空間との対話技術」という5ヵ年プロジェクトの構想についても述べる。

2. 文字列検索から連想的情報アクセスへ

自分の脳に蓄えられた情報(知)に基づいて思考する人間を、外部の膨大な情報と創造的に相互作用(interact)させるためのカギは、人間の連想能力の活性化にあると考える。認知心理学における記憶検索やプライミングの研究が示すように、記憶内の連想的探索は関連知識の無意識的想起を引き起こし、それが人間の創造活動にとって重要な役割を果たしている。人間の創造性を高める新しい情報技術の基礎は、脳内の記憶の連想的探索・無意識的想起と、脳の外にある膨大な情報からの関連情報の探索・分析・提示との新しい結びつきに求められる。外部の膨大な情報が、ユーザの意図に基づく連想的探索可能な形で提供され、探索結果がユーザの無意識的想起を促す形でフィードバックされるならば、ある意味で「第二の脳」としての役割を果たせる。

このような新しい創造的相互作用に理論的・技術的基礎づけを与える「連想の情報学」の確立が求められる。その構築には情報科学・認知心理学を中心とする学際研究が必要となる。我々は、この重要な情報学の分野に、主に情報科学の手法によってアプローチして、膨大な情報に基づく連想過程に計算的な基礎づけを与え、この新しい研究分野の基盤となる情報処理技術を提供することを目指して研究を進めている。これまでに、文書連想検索を特徴とする

文書検索システムDualNAVIと100万件規模の文書DBに基づく連想計算を高速実行できるソフトウェア「連想計算エンジンGETA」を開発し、「連想の情報学」における各種実験システムの構築ベースを提供している。

3. 連想的文書検索システムDualNAVI

連想計算に基づく情報アクセスの有効性を示すため、我々は連想的文書検索システムDualNAVI(デュアル連想検索)を開発してきた(図1)。その特徴は、検索時にキーワードではなく文書(群)を入力して類似の文書を検索する文書連想検索と、検索結果として得られる文書群の内容的把握を助ける特徴語グラフ(自動要約機能)の有機的連携である。新聞、百科事典、学術論文誌など実用規模DBへの適用を通じて、原始的な形とはいえ、DualNAVIは創造的相互作用の場¹を提供することに成功し、「連想の情報学」の有効性を示した。

現在までに、DualNAVIシステムは、百科事典検索サービス「ネットで百科」で商用利用され、東大医科研の検索サービスBACEでも試用されている([NetEncy1999, BACE2000])。また、注目すべき学術情報検索システムとしてGoogleなどととも英国Nature誌に画面写真入りで紹介された[Butler2000]。

4. 汎用連想計算エンジンGETA²

DualNAVIの研究を通じて、連想計算(文書群同士、単語群同士、文書群と単語群間の類似性関連性計算)こそが最も基本的で重要な計算機構であると確信した我々は、大規模な文書集合について各種の連想計算を高精度かつ高速に処理できる汎用連想計算エンジンを開発した。

GETA(Generic Engine for Transposable Association)は、1000万件規模の文書コーパスを対象に、文書連想検索や文書群要約など創造的相互作用の実現に必要な連想計算を高精度かつ高速に処理するソフトウェアである。その性能の秘密は、インデックス情報の圧縮法、および分散処理のためのインデックス分割方式にある。

GETAは「連想の情報学」における各種実験システムの構築ベースとして、文書連想検索、文書クラスタリング、語彙連鎖抽出などの研究や実装に利用され始めている。GETAの最新版はオープンソース形態で無償公開されている[GETA2002]。

¹ 荒俣宏氏は「ロジェ・カイヨワが『対角線の科学』として提唱した創造的思考法をユーザにもたらしDualNAVIは、検索のアヴァンギャルドだ(平凡社月刊百科1999.12)」と評している。

² I P A独創的先進的情報技術PJ開発テーマ「汎用連想計算エンジンの開発と大規模文書分析への応用(研究代表者:高野明彦)」(平成11~13年度)

† 国立情報学研究所 <takano@acm.org>

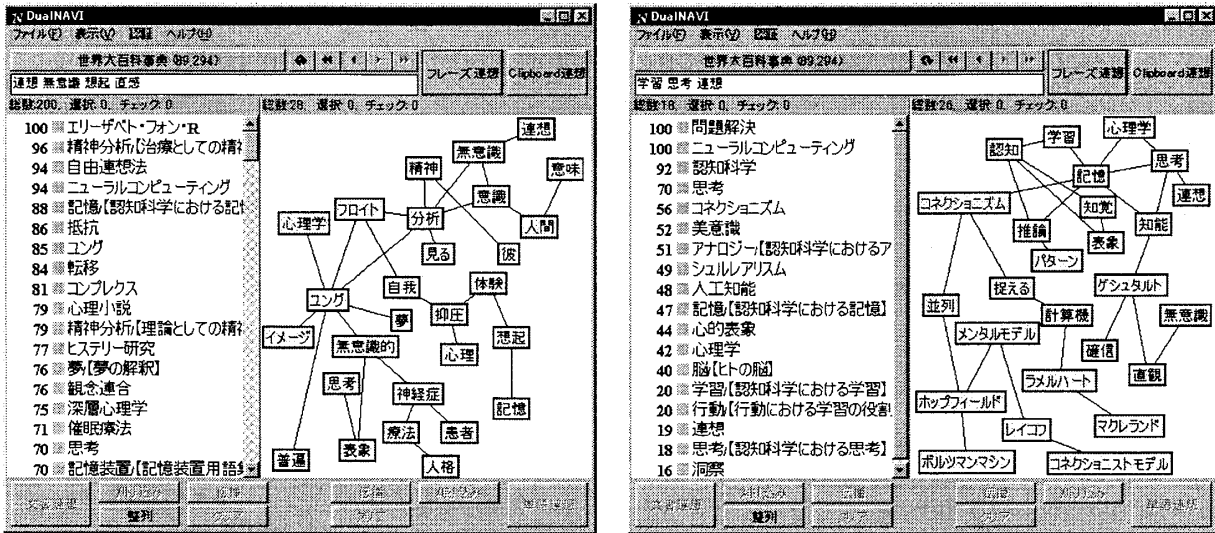


図1. DualNAVIによる百科事典検索例

5. 連想に基づく情報空間との対話技術

すでに述べたように、現在の電子化された情報空間は、高信頼な知識獲得手段としてはもちろん、創造性を刺激するメディアとしても極めて不十分であり、遠く書籍に及ばない。読書の有効性は、情報を理解しやすい構成・順序で摂取することにより、その情報についての系統的な文脈が頭の中に形成されるためだと考えられる。これから求められる情報技術は、情報空間においても読書に匹敵する効果的な情報体験の場を実現し、膨大な情報を活用して個人やグループの問題解決能力を高め、思考や議論を深めるものでなければならない。この課題の解決を目指して、昨年末より「連想に基づく情報空間との対話技術」という5カ年の研究プロジェクトを開始した¹。

本研究の目的は、この新分野に主に計算機科学の手法によりアプローチして、文書情報空間に奥行きと安心感を与える対話技法を構築し、電子化された情報空間での新しい「読書環境」を実現することである。そのため、膨大な情報に基づく連想過程の計算的な基礎づけや、情報空間の内容的概観をフィードバックして人の連想を効果的に刺激する対話手法の開発など、「連想の情報学」の基盤となる情報処理技術について追究する。特に、(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、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する。本技術により、ユーザは情報空間中で迷うことがなくなり、いつでも自分の知識で理解可能な記述レベルの説明にたどり着くことが保証され、情報空間に安心感がもたらされると期待できる。

¹JST 戦略的基礎研究推進事業 (CREST) の研究領域「高度メディア社会の生活情報技術」(領域リーダー:長尾 眞) の研究プロジェクト「連想に基づく情報空間との対話技術」(プロジェクトリーダー:高野明彦)(2001.12~2006.11)

参考文献

[BACE2000] Bio Association Central, 東大医科研ヒトゲノム解析センター, URL: <http://bace.ims.u-tokyo.ac.jp/>.

[Butler2000] D. Butler: Souped-up Search Engines, Nature, Vol.405, pp.112-115, 11May2000, URL: <http://www.nature.com/nature/webmatters/search/search.html>.

[GETA2002] 汎用連想計算エンジン (GETA) 公開サイト, URL: <http://geta.ex.nii.ac.jp/>.

[NetEncy1999] 「ネットでも百科」, 日立システムアンドサービス, URL: <http://ds.hbi.ne.jp/netencytop/>.

[Takano2000a] 高野, 他 10 名: 汎用連想計算エンジンの開発と大規模文書分析への応用, 第 19 回 IPA 技術発表会, 2000.10.

[Takano2000b] A. Takano, Y. Niwa, S. Nishioka, M. Iwayama, T. Hisamitsu, O. Imaichi, and H. Sakurai. *Information Access Based on Associative Calculation*. SOFSEM 2000, LNCS Vol. 1963, pp.187-201, Springer-Verlag, 2000.

[Takano2000c] A. Takano, Y. Niwa, S. Nishioka, M. Iwayama, T. Hisamitsu, O. Imaichi, and H. Sakurai. *Associative Information Access Using DualNAVI*. Kyoto International Conference on Digital Libraries (ICDL'00), pp.285-289, 2000.