



たかの・あきひこ
国立情報学研究所 連想情報学
研究開発センター長・教授

高橋 “nature”2000年5月11日号に、「21世紀に注目される情報検索技術」として、Googleと並び連想検索(*1)が紹介されましたね。

高野 当時は日立製作所におり、そこで開発したDualNAVIという連想検索インターフェースが取り上げられました。そのエッセンスを抜き出したのが、汎用連想計算エンジンGETAです。これは、オープンソースとして公開していて、誰でも無償で自由に使うことができます。

高橋 何らかのきっかけがあったなら、連想検索はGoogleと双壁をなしていたと思いますか。

高野 当時も今も、連想はGoogleの先に行く技術だと信じています。

Googleのすごさは、ウェブの情報を膨大に集め、かつ混乱なくページランキングを行う仕組みや設備をすべて自力で作上げたことです。さらに、必要な資金をきちんと集め、ウェブ情報を信じられないスケールで収集する実行力もありました。

一方、僕たちは、連想検索の方がGoogleより勝っていると素朴に信じていましたし、今もそうです。情報がどんどん増えていけば、関連する情報も爆発的に増え、どんな情報が集まったか把握できなくなる。ランキングうんぬん以前に、どうやって内容を概観するかが課題になる。連想検索では、集めた情報から関連

検索から連想へ

ひらめきをもたらす情報技術

連想検索とは(*1)

単語や文章を入れて検索すると、それに近い内容の文書が提示されると同時に、内容のエッセンスを表す複数の単語が「関連キーワード」として示される。関連キーワードは、一つのテーマや言葉に対し、人間が連想するさまざまなイメージや概念に重なる。そこで、興味のある文書や関連キーワードを手がかりに次々と検索を進めていけば、漠然としたイメージから出発して望む情報を手に入れたり、思いもかけなかった領域に思考や調査の範囲を広げることができる。

連想検索を提供する情報サービス Webcat Plus (*2)

NIIの提供する全国大学図書館蔵書検索サービス。

新書マップ(*3)

新書・選書約1万冊が約1000個のテーマに分類されている。単語や文章を入れて検索すると「関連テーマ」と「関連キーワード」が示され、それをもとに連想検索を進め、適切な新書を選ぶことができる。

想-IMAGINE (*4)

「Webcat Plus」、「新書マップ」、「Book Townじんぼう(神保町古書店160店・31万冊)」、「文化遺産オンライン(博物館・美術館の収蔵品5000点)」、「ウィキペディア」など複数のデータベースを一気に連想検索できるサービス。

Beyond the Information Explosion

キーワードを抽出できるので、これを要約として情報の内容を概観できると考えたのです。

ですから、情報爆発といわれる今こそ、連想検索が真価を発揮する、言い換えると、Googleが直面している壁を先読みして開発した技術だと確信しています。技術のコアは完成しているので、どうすれば効果的で興味深いサービスを提供できるかと、Googleに負けぬ根性で取り組んでいるところですよ。

高橋 メーカーで基礎開発を行い、国立研究所で応用開発を進めるといって面白いケースですね。そもそも連想検索というテーマを、どうやって見出したのですか。

高野 始まりは1996年2月です。日立の基礎研究所にいたのですが、いくつかの小さなグループを1つの研究チームに再編し、そのリーダーになりました。そこには、文書間の近さを判断するための尺度の研究(情報検索)をする人と、bankなどの単語の多義性解消(自然言語処理)を文脈分析で追いかける人がいた。

美しいプログラムを書くことにしか興味のなかった僕には、両者は表裏一体のように見えた。内包する単語の頻度から文書同士の近さを評価するのと、単語の集まりである文脈から単語の意味を判断するのは、文書と単語というデュアルな空間の相

NII Interview Akihiko Takano + Mariko Takahashi



たかはし・まりこ
朝日新聞科学エディター

互反映だと…。そんな三者の触発から連想検索・連想計算の発想が生まれました。優れたソフトウェア設計者も加わって実現に至りました。

高橋 日立で開発した技術が、なぜオープンソースになったのですか。

高野 90年代も終わりになると、メーカーの基礎研究ブームが一段落し、やはり応用に繋がる研究をせねばという揺り戻しが来ました。僕たちも中央研究所に移ることになり、その時に考えたのです。連想計算のメンバーが、仕事の間が変わろうとも、たとえ日立を辞めることになっても、各自の研究を続けていける環境を作らねばならない。そうしないと研究者としての士気もモラルも落ちて、結局、何の成果も期待できない。

もちろんお金と人手をかけて開発したのは日立ですから、外から見ることでできるインターフェース部分は日立のもので、これに関する特許はきちんと出しました。その上で連想計算エンジンをオープンソースにすれば、分野自体が広がり、日立のメリットになると幹部を口説いて、中央研究所に移る直前の99年2月に基礎研所長の許可を得、情報処理振興事業協会(IPA)の「独創的情報技術育成事業」に申請し、採択されました。その研究成果が前述のGETAです。これは、企業時代の僕の唯一誇れるマネジメントですね(笑)。

NIIに移ったのは2001年1月で、NIIの全国大学図書館の蔵書800万件のデータベースなどが魅力でした。連想検索を導入すれば、どんなに豊かな世界が広がるかと…。今では、Googleで「連想」をキーワードとして引くとWebcat Plus(*2)がランキング1位、「新書」で検索すると新書マップ(*3)が1番に出ます。

想..IMAGINE

Googleの先を
行く技術だと
確信しています

高橋 世の中で活用される存在になったというわけですね。連想検索の次なる課題は何ですか。

高野 現在の連想技術の到達点は「想-IMAGINE」(*4)です。内容の信頼性が保証された情報源を並べて、

それらを連想的に柔らかく相互作用させるインターフェースは、バランスのよい視野を確保しながら電子情報空間を探索できる新技術として、キーワード検索を置き換える可能性を秘めています。今後は、言葉の壁を越えて連想が広がる「想」の多言語版を目指します。単なる翻訳による連想ではなく、例えば日本の地ビールのような街おこしのシンボルが、中国やアメリカなどの他文化では何になるのかがわかるような、「文化を超える連想」が目標です。

ヒット数を競う検索の時代はまもなく終わり、信頼に足る情報に基づいてユーザーにヒラメキをもたらす連想技術の時代が必ず来ると信じています。

◆インタビューの一言

連想検索は確かに面白い。次々と関連キーワードをたどっていくと、時のたつのも忘れてしまう。いわば「知的な遊園地」。ただ、ここで遊びすぎると肝心の文書を読む時間がとれないのではと心配だ。「もっといい資料があるはず」と、いつも落ち着かない恐れもある。目指すべきは普通の人のニーズにこたえるサービスだ。GoogleのCEOは朝日新聞に「最終的に目指すのは情報の意味を本当に理解できるようなサービス」と語っている。負けるな、NII!