

まるで専門家のように、関連する情報を提示してくれる

「興味があるけれど、どんな資料を探したらいいかわからない」。こんなとき、既存のキーワード検索ではうまく調べられないことがある。しかし、未知の分野にくわしい人にきけば「こんなものもあるよ」と教えてもらうことができる。そんな専門家や図書館の司書のように、情報を提供する検索サービスが公開されている。キーワードではなく文章で検索する「連想検索技術」を紹介しよう。

Googleなどの検索サイトで検索ワードを入力すると、ウェブサイトがたくさん表示してくれる。そこには、通常「検索ワードを含むウェブサイト」だけがあがっている。そのため、はじめに最適な検索ワードを考え出すことが、検索サイトをうまく使うコツだ。

検索サイトのこのようなしくみは、「未知の分野」について調べる場合にはあまり向いていない。適切な検索ワードの見当がつきにくいからだ。

しかし、もし未知の分野についてくわしい専門家に相談できたら、さまざまな情報を教えてもらえる。専門家は、相談者がもちかけた、かたよった情報からでも、関連する情報を「連想」できるからだ。

重要なキーワードをいくつも選びだす

専門家のように「一を聞いて十を知る」検索エンジンは可能だろうか。国立情報学研究所の高野明彦教授は、ユーザーの入力した文章に関連するさまざまな情報を「連想」し、検索結果として返す連想検索エンジン(<http://imagine.bookmap.info/>)を公開している。

連想検索では、キーワード検索とはちがって検索ワードをしばらくこむ必要がない。興味をもった記事を見つけたら、その全文をコピーして丸ごと検索ボックスにはりつけばよい。すると、記事を受け取った検索エンジンは、文章を名詞や形容詞などに分解し、重要な単語を拾いだす。

このとき「重要かどうか」は、「単語が平均よりも多く出てくるかどうか」で判断している。たとえば「私」という単語はよく使われるので、入力された文章に多く含まれていてもキーワードになりにくいのだ。

キーワード群から文書を“連想”する

こうして選びだしたキーワード群をなるべく多く含む文書を探せば、内容の近い資料が集まる。また、集

まった文書全体からもう一度キーワードを抽出することで、元の文書になかった新しいキーワードを見つけることができる。まさに専門家が「連想」したときのように、さまざまな資料やキーワードが見つかるのだ。

このとき検索対象となるのは、図書館や博物館のデータベース、ウィキペディアのような百科事典サイトなど、文章の形式が決まっているウェブサイトだ。ブログのようにさまざまな内容が混在する自由に書かれたウェブサイトは、関連が薄い場合も表示されやすくなるので、連想検索に向かないという。

Newtonのホームページでは、最近の科学ニュースを「サイエンスニュース」で紹介している。その一つ、「朝型の人が夜に弱いわけ」という記事全文で連想検索を試してみた。たとえば1万冊以上の新書のデータベース「新書マップ」からは、睡眠や脳、体内時計に関わる入門書が多数表示された。また、連想検索された資料からは、「リズム」や「遺伝子」、「習慣」など、元の記事にはない新しいキーワードが提示された。

連想検索は、キーワード検索にくらべて広い範囲の情報が集まる分、見当がいいに見えるものも含まれる。しかし、連想検索で見つけた資料やキーワードで気になるものを基点に、さらに連想検索をくり返せば、欲しい情報に近づくことができる。こうは「検索を重ねる」ことなのだ。🍎

(担当：編集部 市田朝子)

協力

高野明彦 国立情報学研究所
連想情報学研究開発センター長・教授

キーワード検索と連想検索のちがいを

キーワード検索

自動車 不況 検索

入力した文字を含む情報だけを集める。

連想検索

自動車 ハイブリッド
不況 ガソリン

コンピューターが重要なキーワード群を選びだし、それらを多く含む文書を検索する。

スポットライトにたとえると？



一部だけ照らされ、全体が見えない。



複数のライトで家全体が照らされる。

左図は、キーワード検索と連想検索のちがいを示している。キーワード検索は照らす範囲が小さすぎるサーチライトに、また連想検索は、複数のライトで対象全体を照らすようすにたとえられる。