

■安達 淳 コンテンツ科学研究系・教授

【タイトル】

混沌とした情報の海に羅針盤を立てる

【本文】

今、私たちを取り巻く情報の量は指数関数的に増えています。この状況を「情報爆発」と呼んでいます。一方、言語の獲得以降、人間の持つ知識の量は爆発的に増えてきたものの、情報処理能力は限られたままで、2 万年ほど前からほとんど変わっていないと思います。そんな人間が、どうすれば広大な海のような情報の世界を見通し、うまく航海ができるようになるのでしょうか。この問題について研究しています。扱っているのはテキスト情報(文字情報)が中心です。その中にある知識や知恵を、目的に応じて簡単かつ適切に取り出せるようにしたいと思っています。

情報の海から信頼できる知を引き出す

「情報爆発」といっても、まったくオリジナルな情報がどんどん生み出されるわけではなく、ある情報の一部を抜き出したり、複数の情報の部分どうしを組み合わせたようなものがほとんどです。その中には、論理的なプロセスは同じで結論だけが正反対というものや、組み合わせ方が不適切な情報などがたくさんあります。このように情報の海は混沌としており、その中から本当に信頼できる情報、有用な知識や知恵をどう選び出すかが大きな課題です。

これを解決するためのモデルはすでにいくつかあります。よく知られているのは多数決による解決や、権威のある人・機関が発信するランク付けを評価基準にするといった方法です。しかし、薬害問題などでもわかるように、権威のある人や機関・国が保証したとしても、確かとはいえません。「自己責任による自己判断」という方法もありますが、適切な情報を入手し整理する手段がなければ、的確な判断を下すことはできません。このような判断の役に立つ検索技術を開発することが、私たちの使命だと考えています。

目的や状況に合わせた柔軟な検索を可能にする

信頼できる検索技術を実現する 1 つのアプローチとして、テキストの中から、ヒト、モノ、イベントなどの項目を抽出するだけではなく、抽出した複数の情報が同じものなのか、違うものなのかを判断する機能をあげられます。これは年金問題で有名になった「名寄せ」、つまり同姓同名のヒトの区別や、同音異義語をどう扱うかといった処理です。この件に関する私のグループの研究成果は NII の学術基盤推進部が提供している論文検索サービスにすでに反映されています。

さらに進めば、計算機が与えられたキーワードだけで検索をするのではなく、ユーザーが置かれている状況をそれまでの操作履歴などから推測し、本当に求められている情報を探せるようにできないか、と考えています。

もっと先の話としては、イメージはわかっているけれど言葉でうまく説明できないものを探すシステムを作れないかと思っています。なぞなぞのような問いに答えられるシステムといえいいでしょうか。これは教育を始めさまざまなシーンで役に立ちます。例えば、子どもが街中で見た「自動車」のことを調べようとしたとしましょう。ところが「自動車」という名前を知らない、今の検索エンジンではすぐには探せません。でも子どもの頭の中には自動車のイメージがあるわけです。私が作りたいと思うのは、子どもたちが、そのイメージを、例えば「タイヤが 4 つあって動くもの」と入力すれば、「自動車」と答えてくれるシステムです。キーワードがわからないということは、大人が海外に行ったときや、高齢者も経験することです。そういう、うまく言葉で説明できないものを探すシステムを作るというのが、中期的には重要だと思っています。

(取材・構成 齋藤 淳)