

■大山敬三 コンテンツ科学研究系 教授

【タイトル】

日本発の検索エンジン技術を目指して

【本文】

“Web 検索エンジン”と聞いて、真っ先に思い起こすのは Google と Yahoo! でしょうか。いずれも米国生まれで、今や世界中で使われています。日本には実質的に Web 検索エンジンはありません。これには著作権の問題など日本固有の制約が大きく関係しています。しかし、だからといって現代の情報インフラである検索エンジンを海外に頼り切っていていいわけがありません。だから、私はこの分野に興味のある人たちと協力して、独自の新しい検索エンジン技術をなんとしても開発したいと考えています。

検索エンジンの進歩を加速する

“Web 検索エンジン”とは、インターネット上の情報をキーワードによって検索できる Web サイトのことで、大量の情報の中から欲しいものだけを選び出すことができます。

こうした検索エンジンが役に立つのは、単に検索によって選び出された情報を漫然と表示するのではなく、利用者にとってより重要だと判断される情報から順番に並べられるところです。一般に、正確で有用な情報は、ほかのページから参照されるケースが多くなります。つまり、リンクが多く張られているページほど重要度が高いのです。この考え方によって、検索結果に順番を付けるわけです。

ところが、インターネットに公開される情報はますます増え、それともなつて情報の質は低下しています。今、広く使われている技術だけで、信頼できる情報を選び出すことは難しくなってきました。そのため、検索エンジンは常に進歩していかなくてはなりません。

NII の役割の 1 つに、大学や民間などの研究者と協力して研究を推進していくことがあります。その一環として、2001 年から 2006 年まで、私は、情報アクセス技術の評価するための国際的なプロジェクト「NTCIR」で Web 検索エンジンの評価にかかわってきました。具体的には、いくつもの大学や研究機関が開発した「検索システム」に対して、検索対象となる 1 億ほどの Web ページのデータベースを用意し、ユーザーが知りたいと思う情報要求を検索課題として数十から数百種類提示します。そして、各検索システムが出力した各課題に対する検索結果について、ユーザーが求める情報に近いかという観点で「適合性」を判定して正解データを作ります。これを用いて多数の検索課題に対する各システムの成績を比較することによって、さまざまな検索技術の有用性の評価が可能となるのです。このデータベース、課題、正解データの三点セットはその後も多く研究者に使っていただき、検索エンジン技術の研究推進に貢献できたと考えています。

「怪しさランキング」をつくる

新しい検索エンジン技術の研究として、私自身が現在取り組んでいるのは、ネットワーク上の情報の安全・安心をどうしたら実現できるのかというテーマです。さまざまなアプローチが行われていますが、私は、怪しい情報に注意を喚起するための「怪しさランキング」が有効なのではないかと考えています。それには Web 上の情報のいろいろな側面をとらえる必要がありますが、例えば怪しいページは互いにリンクを張る傾向が強いという特性を利用するわけです。この技術開発の難しさは、Web 上の情報は多数意見が必ずしも正しいとは限らないという点です。現在、何をもちて“怪しい”と判断すればいいかを探っているところです。

最終的にはこうした取り組みの中から、日本発の Web 検索エンジンを生み出したいと考えています。

(取材・構成 池田亜希子)