

自然言語による 情報アクセス技術

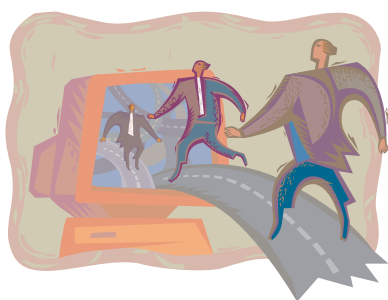
編集にあたって

福本 淳一

立命館大学 情報理工学部
fukumoto@media.ritsumeai.ac.jp

天野 真家

東芝リサーチコンサルティング（株）
shinya.amano@toshiba.co.jp



ブロードバンド・低額な定額料金制インターネットの普及、さらには携帯端末、ユビキタスコンピューティング環境などの普及により、文字通りいつでもどこでも大量の情報に対するアクセスが可能になっている。検索エンジンやポータルサイトを利用して、さまざまな情報を得ている方も多いのではないかと思う。たとえば、京都の河原町で美味しい地酒のある居酒屋について知りたい場合など、適当なキーワードを入力することで関連するさまざまな情報を手に入れることが可能である。

しかしながら、現状では本当に欲しい情報にたどり着くために、検索された情報を実際に見ながら情報を取捨選択する必要がある。また、求める情報が検索結果に存

在しない場合にはキーワードを削除、追加しながら検索を繰り返す必要がある。このように大量の情報に対してアクセスする技術にはいくつかの段階があり多くの研究がなされている。

まず、最初にインターネット上の大量の情報から、求める情報を含むものを選択する技術として情報検索技術がある。大量の情報から求める情報を選び出すためには、求める情報の内容を表すいくつかのキーワードを入力したり、自然言語文で入力する場合もある。いずれにしても内容を適切に表現したキーワードをいかに選択するかが、求める情報を含むページを可能な限り正確に得るための情報検索技術の重要なポイントである。

次に、選択された情報から求める情報のみを取り出す情報抽出技術がある。情報抽出技術では、重要な情報として考えられる人名や会社名、商品名や価格、サイズなどの情報を取り出すとともに、企業のトップの異動などのイベントや事柄に注目して得られた情報をまとめるといったことがなされている。情報抽出ではこれらの情報をパターンとして記述しておくことで求める情報を取り出すことが可能である。

また、得られた情報から重要な情報のみを選択し、要点の把握を支援するものとして自動要約技術がある。自動要約では、さまざまなテキストやWebなどの情報において重要な部分を判断したり、適当なクエリを与えることでクエリに特化した要約を与える。非常に簡単なものとして検索結果に対してそれぞれの要点を表示するということがインターネットの検索でもしばしば行われている。

さらに以上の技術を複合したものとして、大量の情報に対して質問文を与えることで質問の答えを得る質問応答技術もある。質問応答においては、自然言語文で記述された質問文を与えることで大量の情報から関連する情報を検索し、質問文の答えに相当するもののみを抽出することで答えを得るものである。答えの根拠や答えそのものが文書の要約となる場合もあるであろう。

以上のように大量の情報に対する情報アクセス技術として情報検索、情報抽出、自動要約、質問応答についてさまざまな研究がなされてきており、これまでも本学会や他の学会でもいくつかの解説や報告がなされている。本特集では、これらの分野の研究においてこれまでの研

究について概観するとともに、これらの技術が現在どこまで進んできているのか、どのような応用があるのかについて解説する。これにより情報アクセス技術に対してより多くの人に興味を持ってもらい、研究の一層の発展とともに、他の多くの分野においても情報アクセス技術が利用されることを望むものである。

第1編「情報抽出」では、インターネット検索技術としてWeb検索の技術動向と評価方法について解説する。Web情報は年々、増加の一途をたどっており、このように膨大で多様な情報に対してユーザが効率的に情報を検索するためにどのような技術が必要であるのかについて述べる。

第2編「Web検索の技術動向と評価手法」では、情報抽出技術についてこれまでの代表的な要素技術について解説するとともに、より多くのニーズに答えるための新たな展開についても解説する。従来までは人名、組織名、場所名などのある程度決まった形式の情報を対象にしていたが、多くの応用分野で使われることでさらに多様な情報を扱うことが求められており、応用も含めた最近の動向についても述べる。

第3編「テキスト自動要約」では、現在のテキスト自動要約技術が情報アクセス技術としてどの程度利用可能なのかについて実例を紹介しながら解説する。現在、自動要約においては新聞

記事だけでなくWebページ、メール、ニュースなどさまざまな情報に対して適用が行われており、それぞれどのような観点に基づいて要約がなされているのか、どのような要約技術があるのかについて述べる。

第4編「質問応答技術」では、質問応答技術について解説する。質問応答技術はユーザの求める情報を質問のかたちで受け取り、大量の知識源から答えとなる情報を返すものである。質問に対して的確に答えるためには、正しく質問を理解し、求める情報を的確に探し求めることが重要である。そのための技術について解説するとともに今度の研究動向についても述べる。

最後にご多忙の中、非常に限られた時間であるにもかかわらず快く執筆をお受けいただいた著者の方々に厚くお礼申し上げます。

(平成16年5月3日)

