

Web 質問応答システム: 自然言語による答の検索

森 辰則

横浜国立大学 環境情報研究院

〒 240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-7

mori@forest.eis.ynu.ac.jp

講演の概要

情報ネットワーク技術の発展により、多種・多様な情報がネットワーク等を通じて潜在的に入手可能な状況にある。しかし、それによってもたらされる情報爆発により、利用者が真に必要な情報に辿りつくことは容易ではない。情報アクセス技術は、このような状況の下、利用者の持つ情報要求を利用者自身から引出し、多量の情報の中において利用者が必要とする情報に導くための技術的枠組である。

テキストを対象にした情報アクセス技術において、重要なものの一つが質問応答技術である。質問応答技術とは、利用者が明確な情報要求を持っており、それを自ら自然言語の質問文の形で表現できる時に、その質問文の答となる表現を大量の文書コレクションの中から根拠情報とともに抽出するための技術である。過去に人工知能研究では「質問応答」として、限定した領域において組織化された知識や規則を用意し、それらに基づいて推論を行ない、利用者との対話を行なう仕組みが研究されていた。これと区別するために、我々が対象とするような、大量文書を情報源として用い、特定の領域に関する知識の整備を必要としない質問応答は、「領域非依存質問応答 (open domain question answering)」と呼ばれることもある。

検索要求に適合する文書を見つけ出す情報検索とは異なり、質問応答は、質問文に現われる「誰」、「何」、「なぜ」といった表現で指定される回答の種類を考慮した上で、文書中に現われる必要十分な文書部分を回答候補として切り出して利用者に提示してくれる点に特長がある。利用者は質問文という自然な形で情報要求を記述できるので、情報爆発時代における人にやさしい情報アクセス技術の一手法として位置付けることができる。

他の情報アクセス技術と同様に、質問応答は TREC や CLEF といった評価型ワークショップの中でその技術が磨かれてきている。特に、アジア圏の言語を対象とする研究者に対しては、国立情報学研究所 (NII) が主催する評価型ワークショップ NTCIR (NII Test Collection for IR Systems) が国際研究交流のフォーラムを提供してきた。

本講演では、上述の質問応答システムに関して概説を行なうとともに、Web 上に存在する膨大な量の文書情報を対象として我々が行なっている研究事例 (下記参考情報を参照) を紹介し、情報爆発時代における質問応答システムの役割について論じる。

参考情報

- MinerVA-N (ミネルバ-エヌ)
<http://www.forest.eis.ynu.ac.jp/cgi-bin/QA-www>
Web 文書を情報源とし、入力された質問文に対して、名前や数値・日時などの短い表現で回答をする質問応答デモシステム。
- MinerVA-D (ミネルバ-ディー)
<http://www.forest.eis.ynu.ac.jp/cgi-bin/QA-www-nf>
Web 文書を情報源とし、入力された質問文に対して、定義や理由、方法などの説明文章で回答をする質問応答デモシステム。