

平成17年2月10日

(報道発表資料)

日本電信電話株式会社
NTTレゾナント株式会社

gooラボで提供中の日本語自然文検索実験「Web Answers」に新機能を追加

～専門用語・人物の説明や評判を探し出す百科事典のような検索機能を実現～

日本電信電話株式会社(以下NTT、本社：東京都千代田区、代表取締役社長：和田 紀夫)とNTTレゾナント株式会社(以下NTTレゾナント、本社：東京都千代田区、代表取締役社長：資宗 克行)は、NTTレゾナントが運営するポータルサイト「goo」(*1)の「gooラボ」(*2)上で展開中の日本語自然文検索サービス実験「Web Answers」(*3)に、新たに専門用語・人物の説明や商品・サービスの評価・評判を探し出せる、新機能の検証・評価のための共同実験を、2005年2月10日～2005年5月9日までの予定で行います。

従来の「Web Answers」において回答可能な質問は、だれ(人名)・どこ(地名)などの名称に関する質問や、いつ(時間)・いくら(金額)などの数量に関する質問などに限定されてきました。今回の機能追加により、専門用語や人物に関する説明を含む特定のWebページを検索し、回答候補を抽出する「定義検索機能」と、商品やサービスに関する意見や評価を含むWebページを検索し、回答候補を抽出する「評判検索機能」を実現しています。[\(別紙\)](#)

これにより、自然な日本語の質問により、Webを百科事典のように使ったり、Web上に散在する事物に対する様々な意見や評価を効率よく調べることができます。また、Web上の膨大な情報を知識源としているため、従来の百科事典などでは難しかった、広範な分野の質問や最新の話題に関する質問に回答することが可能になります。

1. 背景と目的

インターネットは知識の宝庫と言われますが、その知識は無数のWebページに無秩序に分散しています。また昨今、様々なネットユーザが掲示板やブログなどにおいて多様な意見を表明するようになり、その結果、商品購入の際の情報収集や、マーケティング調査、企業のリスク管理などのために、これらの意見を効率的に検索したいという要求が高まりを見せております。この膨大なインターネット上の情報から、まさに“知の集約”を実現するWeb検索サービスが、注目されています。

一方、百科事典、専門用語辞典、人名録などは、系統的かつ網羅的に情報を掲載していますが、比較的大きなニーズがある分野でしか編纂され

ず、現代の多様化した人々の興味を満足することは難しく、また、これらは必ずしも最新の情報を反映しているとは言えません。

この度、提供する新機能は、従来の“名称”および“数量”に関する質問に加えて、用語の定義、人や会社のプロフィール、検索対象物に関する評判情報など、回答可能な質問の範囲を大幅に拡大し、より多くのユーザの検索ニーズに応えるサービスを実現すると共に、一度の検索でユーザが求めている情報を提供可能とすることで大幅に利便性を向上させています。NTTサイバーソリューション研究所とNTTレゾナントは、一般のインターネットユーザを対象として「gooラボ」にて実験を行うことにより、同機能の技術検証とビジネス性の評価を行います。

2. 各社の役割

(1) NTT

日本語自然文検索サービスのコンセプト提案を行うとともに、本検索サービスに必要な高度テキスト解析機能を提供します。これにより、商用環境における同機能の技術的検証を行うとともに、精度向上のためのデータを得ることにより、今後の開発に反映させます。

(2) NTTレゾナント

実験終了後の商用サービス導入検討に向け、「goo」の利用度向上やサービス効果の検証を通してビジネス性の評価を行います。また、NTTグループのインターネット事業分野を代表するインターネットポータルとして、日本人が日本語を利用して検索する際に最も利便性が高く、“知の集約・集積”を実現する検索サービスの提供を目指します。

3. 技術のポイント

従来のテキスト解析技術に加え、以下の技術を新たに開発したことにより定義検索・評判検索を可能にするとともに、更なる検索の高速化と回答精度の向上を図りました。

(1) 定義表現、評価表現抽出とWebページランキング技術

用語の定義および人物のプロフィールに関する質問や、評判に関する質問に対して回答となり得る表現を抽出する技術です。

インターネット上の技術文書や事典・用語集などのテキストデータを言語学的に分析して、例えば、「MPEGとは、映像データの圧縮方式の一つである」における「～とは～である」のような用語の説明(人物のプロフィールも同様)を記述する表現(定義表現)に関する言語知識データベースを構築し、このデータベースと定義らしさを判定する統計的な手法を組み合わせることにより、定義表現抽出を実現しました。

また、インターネット上の掲示板やブログなどのテキストデータを言語学的に分析して、例えば、「絶品(名詞)」「しんみり(副詞)」「分かり易い(形容詞)」「役立つ(動詞)」など、事物に対する意見や評価を記述する表現(評価表現)に関する言語知識データベースを構築し、このデータ

ベースと評判らしさを判定する統計的な手法を組み合わせることにより、評価表現抽出を実現しました。

更にユーザの質問文に対する回答として適切かどうかを、回答候補が出現する文脈に基づいて統計的に判定し、ランキングすることにより、回答を含んでいると考えられるWebページを上位に表示することが可能となりました。

(2) 回答抽出処理の高速化

質問に対する答えを高速に抽出する技術です。探索アルゴリズムの適性を図ったことにより、従来の抽出技術の約2倍の高速化を達成しました。これにより、ユーザからの一つの質問に対して複数回のインターネット検索を瞬時に実行することができ、より精度の高い回答を抽出することが可能となりました。

4. 今後の予定

ブロードバンド時代におけるポータルサイトの付加価値をさらに高めるために、NTTでは、インターネット検索サービスの更なる高機能化のための開発を進めていきます。また、NTTレゾナントでは、本実験で得られたデータをもとに、「goo」でのサービス化も視野に入れビジネス性の検討を図る予定です。

【用語説明】

(*1) 【goo】 <http://www.goo.ne.jp/>

1か月あたり約1,460万人のユーザからのアクセスを有する、日本を代表するインターネットポータルサイト。サービスの中核である「検索サービス」では、Webページの検索サービスのみならず、辞書や地図、その他各種実用情報などの多彩且つ膨大なデータベースの検索サービスを提供しています。また、併せて「コンテンツ」「パーソナル便利ツール」「コミュニティサービス（約480万会員）」等の幅広いサービスも提供しています。

(*2) 【gooラボ】 <http://labs.goo.ne.jp/>

NTT研究所で開発された新技術を活用し、先進的なインターネットサービスの可能性を社会に提示することを目的とする実験サイト。

(*3) 【「Web Answers」の共同実験】

<http://www.ntt.co.jp/news/news04/0402/040205b.html>

2004年2月5日より「gooラボ」で共同実験を行っているサービス。

・ [別紙 Web上の膨大な知識を百科事典のように使う自然文インターフェース](#)

本件に関する問合せ先

NTTサイバーコミュニケーション総合研究所

企画部広報担当 定方・山下

Tel : 046-859-2032

E-mail : ckoho@lab.ntt.co.jp

NTTレゾナント ポータル事業本部

Tel : 03-5224-5500

E-mail : pr@nttr.co.jp

NTT ニュースリリース 

Copyright(c) 2005 日本電信電話株式会社