

平成22年2月22日

報道発表資料

NTTレゾナント株式会社
日本電信電話株式会社

有名人の句の相関関係を可視化する「gooラボ ひとつなぎ」の実験を開始

～ブログなどの記事から人間関係を自然言語処理技術で自動抽出し相関図に～

NTTレゾナント株式会社（以下NTTレゾナント）と日本電信電話株式会社（以下NTT）は、本日より、ブログで話題になっている有名人の相関関係を相関図として一望できる「ひとつなぎ」をポータルサイト「goo」上の実証実験サイト「gooラボ」にて開始します。

本実験では、ブログ記事などから、「家族」や「共演」「友人」などの有名人同士の関係を自動抽出し、関係情報を集約・図形表示することにより、人物相関図を自動生成します。これにより、ユーザは、ブログなどで話題になっている句の人物関係を一望することが可能になります。

NTTレゾナントは、NTTサイバースペース研究所のテキスト意味抽出技術を採用し、新たな検索のUI（ユーザインタフェース）の可能性を追求することでユーザにとって楽しめる検索の形を提供します。

※ 本サービスはNTTとNTTレゾナントの共同実験として実施するもので、2010年2月22日から12月31日までを予定しています。

1. 「gooラボ ひとつなぎ」の特徴について

「gooラボ ひとつなぎ」は、NTTサイバースペース研究所が開発した「リッチインデクシング技術」の「関係抽出機能」および「人物同定機能」により、ブログ文書から人物関係を自動抽出することにより実現しています。

<画面イメージ>（※図は「豊臣秀吉」を検索した場合のイメージです。本実験ではタレント・スポーツ選手などの有名人の句な情報から人物相関図を自動生成することができます。）



(1) 関係抽出機能

日本語で書かれたテキストから「家族」や「共演」「友人」など"人"と"人"の間にどのような関係があるかを解析し、自動抽出する機能です。従来、テキスト中で人名と人名が10文字以内であれば、その間の単語をその人たちの関係と推定するなどの経験的な手段が用いられることが主流でしたが、本技術では文の構造や文脈情報、語彙的な情報を総合的に判断して関係を推定します。これにより、これまで以上に高精度で[Aさん、共演、Bさん]といった3つ組みの抽出が可能となりました。

(2) 人物同定機能

日本語テキスト中に記載された人名のニックネームからフルネームを推定する機能です。ブログ記事などに記載される有名人の名前はフルネームではなく、ニックネームを用いることが多いため、人間関係を抽出するだけでなく、人名が実際には誰のことを指しているかを推定することで、より高精度な人物相関図として集計することができます。

(3) リッチインデクシング技術 (参考:

<http://www.ntt.co.jp/journal/0806/>)

「リッチインデクシング技術」とは、自然な日本語で書かれたテキストから有用な知識を抽出・活用するためにNTTサイバースペース研究所が開発したテキスト意味抽出技術です。

リッチインデクシング技術では、主に固有表現^(*)を中心に、その指し示すものが実世界において何を意味するものかを推定します。また、それぞれの固有表現の間関係性や関連する評判等を抽出することにより、自然言語^(*)で書かれた文書の内容をデータベースに格納できる形で整理し

ます。これにより、各種の集計やマイニングが可能となります。

本技術は、「gooブログ検索」での利用をはじめ、「ブログ通信簿」「CODE検索」などの「gooラボ」サービスでも利用しています。

2. 利用方法

「goo」トップ (<http://www.goo.ne.jp/>) から「gooラボ」をクリックし、「ひとつなぎ」のデモサイトをクリック。検索ボックスに有名人の名前を入力、もしくは検索ランキングとして表示される人物名をクリック。

また、一部の有名人については、「goo」トップから人物名で検索して、「この人のつながりを見る」というリンクをクリックすることでも利用できます。

3. 各社の役割

(1) NTTレゾナント

本実験サービスに必要な、ブログクローリング技術、ならびに、検索インターフェース技術を提供します。本実験により、利用動向やユーザのご意見を基に、ビジネス性の評価・検討を行います。

(2) NTT

本実験サービスに必要な、リッチインデクシング技術を提供します。本実験により、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、今後の開発に反映します。

4. 今後の予定

NTTレゾナント及びNTTは、ポータルサイトの付加価値をさらに高めるために、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研究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) 【固有表現】

「誰が・いつ・どこで」などの疑問語に対する答えとなるような名詞的表現のことで、具体的には、人名・地名・組織名などの固有名詞や、日付・金額などの数量表現を指す。例えば、「日本電信電話株式会社」という表現は、「日本」、「電信」、「電話」といった単語が組み合わさって構成されているが、これ全体で組織名を表す固有表現となる。

(*2) 【自然言語】

プログラミング言語などの人工的な言語に対し、日本語や英語など人間が日常用いる言語のこと。

本件に関する問合せ先

NTTレゾナント 広報担当

Tel : 03-6703-6250

E-mail : pr@nttr.co.jp

NTTサイバーコミュニケーション総合研究所

(NTTサイバースペース研究所)

企画部広報担当

Tel : 046-859-2032

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

NTT ニュースリリース 

Copyright(c) 2010 日本電信電話株式会社