

株式会社富士通研究所

本社住所	〒211-8588 川崎市中原区上小田中 4-1-1
URL	http://jp.fujitsu.com/group/labs/
展示名	特許意味検索の プロトタイプシステム
お問合せ先	担当部署 メディア処理システム研究所 TEL 044-874-2489 FAX 044-754-2923 E-mail ushioda@jp.fujitsu.com

■アピールポイント

文の意味構造を活用した、自然文入力による特許ランキング検索のプロトタイプシステムを開発しました。構文解析および意味解析を経て得られた、語と語の間の意味的繋がりを検索の類似度評価に取り入れることにより、従来のキーワードベースの検索システムに比べ、よりユーザの意図に沿った表現を多く含む特許を検索することが可能になります。

【産業日本語との関連】

産業日本語の代表例として取り上げられている特許文書について、知的生産性の向上のための日本語意味処理の効果の検証を行っています。検証対象の1つとして、特許検索への適用例を御紹介します。

【詳細】

インターネット検索はじめ私達が通常目にする自由入力の検索システムでは、ユーザが指定した1つあるいは複数個のキーワードを入力として、それらのキーワードを含む(あるいは多く含む)文書を検索するタイプが主流です。しかし特許検索のように技術的に込み入った内容の文書を検索する場合、個々のキーワードがお互いどのように意味的に繋がっているかをユーザが指定できた方がユーザの意図に近い文書が検索できると考えられます。たとえば「塩素ガス、劣化、低減」と言う3つのキーワードを検索式(検索クエリー)としてキーワードベースの検索システムに入力した場合、ユーザの意図が「塩素ガスによる劣化を低減させる」ことなのか、「塩素ガスによって劣化が低減する」現象なのか、あるいは「塩素ガスの低減によって劣化する」ことなのか判別することができませんが、自然文で入力できればユーザの意図はより明確になります。そこで、特許検索においてはいわゆる概念検索と呼ばれる検索方式も導入されていて、ユーザは単語ではなくて自然文(文章)を検索クエリーとして入力できるようになっています。しかしながら概念検索システムもその多くが、システム内部では文章から単語や複合語を切り出して、検索対象文書内でのそれらの出現頻度をもとに検索を行っています。その意味で多くの概念検索システムもキーワードベースの検索であると言えます。

富士通研究所では、長年の機械翻訳研究で培った高精度の構文解析および意味解析技術を活用して、クエリー中の語と語の間の意味的繋がりを検索の類似度評価に取り入れることにより、従来のキーワードベースの検索システムに比べ、よりユーザの意図に沿った表現を多く含む特許を検索する技術を研究開発しています。今回の展示ではこの特許意味検索システムのプロトタイプをご紹介します。

特許意味検索システムの概要

特許意味検索システムでは文書が表す意味を、文書中に含まれる文の意味構造の総和と定義しています。本研究における意味構造とは格文法に基づいたグラフ構造により表現されます。すなわち、意味構造は、単語の概念を表すノードおよびノード間の関係を表すアークからなる有向グラフにより表されます。図1は、「太郎は花子に本をあげた。」の意味構造を有向グラフにより表した例です。図1において、○で囲まれたものがノードを表し、□で囲まれたものがアークを表します。この例ではノードは、「GIVE」、「HANAKO」、「TARO」、「BOOK」の4つです。アークは「中心」、「目的」、「動作主」、「対象」、「述語」、「過去」の6つです。アークはノード間の関係を表します。1ノードにしか繋がらないアークはそのノードの属性を表します。

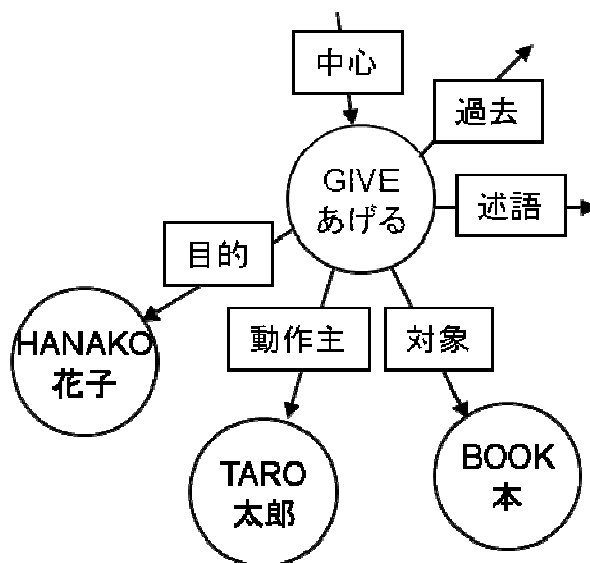


図1. 意味構造

意味を有向グラフとして表現したとき、意味検索はグラフマッチングと考えることができますが、単純なグラフマッチングでは、グラフの複雑性や表現の多様性、また意味解析処理の精度の問題などによりマッチ率に限界があります。そこで、本システムでは、意味構造をアークとそのアークにつながるノード(1つか2つ)の部分グラフに分解し、この部分グラフをベースとする検索アルゴリズムを考案し実装しました[1]。意味構造の抽出には、日英翻訳エンジン ATLAS[2]の翻訳過程で生成される意味解析結果を用いています。ATLAS は中間言語方式の翻訳方式を採用しており、原文を辞書と文法規則に基づき解析して意味構造を計算しています。

[1]大倉清司,潮田明(2012) 意味検索のプロトタイプシステムの構築. 言語処理学会第18回
年次大会予稿集. 2012.

[2] 富士通.英日・日英翻訳ソフト ATLAS.
<http://software.fujitsu.com/jp/atlas/>