1. はじめに

ネット社会の進展に伴い,ウェブメディアには多様な情 報が集積され続けている. このようなウェブ情報をどのよ うに探索,発見し,取捨選択するかについて,さまざまな 方向からのアプローチがなされている. 現在,多くの検索 エンジンが実用化されるとともに,ウェブ情報検索やウェ ブ情報マイニングについて広く研究が行われている.現行 のウェブ検索エンジンは,キーワードクエリに基づく文書 検索モデルに基づくものが多く,この意味では,Web から 文書ではなく情報そのものを検索するという意味でのウェ ブ情報検索は,いまだ多くの問題を抱えている. ある目的 を達成する手段や方法をウェブから検索する,いわば,タ スク検索も未解決の問題の一つである.

Recently, more and more information is stored in web media as the Internet become bigger. Researchers are trying to build a new method to search, find and collect information from diverse sources. At this time, they are developing many kinds of search engines and studying about information retrieval and web mining. Most of web search engines are based on document search model, which is based on keyword queries. Therefore, information retrieval systems have many faults when they try to find pure information not documents. Task search is also a unsolved problem. Task search is a search way that looks for a method to accomplish the purpose from web.

何かを成し遂げたいが,どうすれば達成できるかがわか らないとき,現行のウェブ検索エンジンを使って手段・方 法を見つけることが日常的にも良く行われている. たとえ ば “花粉症対策をする” をそのままクエリとして指定し,検 索を実行することで,“立体マスクをつける” や “アレロック錠を飲む” というタスクを発見することができるが,検 索漏れも多く,この意味で,タスク検索の再現率向上は大 きな課題である.多様なタスク(答え)が得られた段階で, 初めて,安心して各タスク(方法)を比較したり,自分に とって最適な方法を考えることができるようになる. 本研 究では,「目的となるタスクを達成するために,どのような タスクがあるか」を高い再現率で発見するタスク検索手法 を提案する.

タスク検索の例を説明する. たとえば “花粉症対策をす る” というクエリをウェブ検索エンジンに入力すると,出 力として

• 立体マスクをつける. • アレロック錠を飲む. • 医師の診断を受ける.

といったように,複数の異なったタスク(方法)を発見 することができる. この例は,クエリのタスクに対して, 検索された各々のタスク(方法)がそれ自身のみで解であ り,各解は,元のタスク「花粉症対策をする」に対して, instance-of という汎化関係が成立する.また,同じクエリ「花粉症対策をする」に対して,次のような出力が得られる 可能性もある.

• スギ花粉対策をする. • カモガヤ花粉対策をする. • ヒノキ花粉対策をする.

この場合は,“スギ花粉対策をする” や “ヒノキ花粉対策を する” という解は,クエリ “花粉症対策をする” というタス クに対して,a-kind-of(もしくは subtype-of)という汎化 関係が成立する.さらに,“花粉症対策をする” というク エリに対する出力として,以下のような項目をすべて含む ページが検索されることも考えられる.

* 花粉症専用のマスクを販売している店を見つける.
* その店舗で花粉症専用のマスクを購入する.
* 隙間の無いように花粉症専用マスクを装着する.

与えられたタスク(クエリ)に対する検索解は,1 つのペー ジが 1 つの解に対応する場合や,複数のページ(に記載さ れたタスク)の集約が 1 つの解に対応する場合があり,現 行のウェブ検索エンジンによる解ページのランキングでは 不十分であることが予想される. 花粉症対策の方法は非常 に多様であり,高くランクづけされたページであっても花 粉症対策の方法のごく一部を含んでいるにすぎない場合も ある. また,高くランクづけされたページ同士は内容が重 複していることが多く,高ランクのページを見て回っても, 発見できる花粉症対策の方法の数は増えないことも考えら れる.   このようなタスク検索が実用的なレベルに達すれば、以下に示すような,多くの応用例が考えられる.

* 入力クエリに関連するタスクをクエリ推薦としてユー ザに提示する.
* “マスクをつける” という入力クエリに対して,“アレ ロック錠を飲む” という,ある同一の目的を達成する 代替タスクをユーザに推薦する.
* 1 つのページに複数のタスクが網羅的に記載されてい るページの発見や,多様なタスクを含んだページ集合 のランキングといった,タスクベースのページランキ ング.

 このようなタスク検索を実現するための第一歩として, 本研究では,入力となるクエリを達成するためのタスクを 含んだウェブページを網羅的に集めるための手法を提案す る.提案手法では検索連動型広告に着目し,入力クエリに 関連した動詞を自動的に抽出し,入力クエリを拡張するこ とで,入力クエリ単体では得ることができないタスクをよ り多く発見する手法を提案し,その有効性を検証する.  本稿の構成は以下のとおりである.まず,2 章で関連研 究を紹介する.3 章では,本研究で扱うタスクとサブタス クの概念を定義するとともに,その関係性を整理する.4 章で提案手法について説明し,5 章で評価実験について述 べた後,最後に本稿をまとめる.