

# 語の出現頻度と意味関係分析を用いた Webからのタスク検索

加藤 龍<sup>1,a)</sup> 大島 裕明<sup>1,b)</sup> 山本 岳洋<sup>1,c)</sup> 加藤 誠<sup>1,d)</sup> 田中 克己<sup>1,e)</sup>

**概要:**本研究では、クエリとしてある目的が与えられた際に、その目的を達成するために必要なタスク集合をウェブから発見するタスク検索を提案する。本稿では、タスク検索の第一歩として、入力クエリを効果的にクエリ拡張することで、元のクエリでは発見不可能なタスクを含んだウェブページを収集する手法について提案する。提案手法では、検索連動型広告に着目し、動詞の出現パターンを用いてタスクに関連した動詞を抽出することで、クエリ拡張を行う。

**キーワード:** タスク検索, 検索連動型広告, クエリ拡張

RYO KATO<sup>1,a)</sup> HIROAKI OHSHIMA<sup>1,b)</sup> TAKEHIRO YAMAMOTO<sup>1,c)</sup> MAKOTO P. KATO<sup>1,d)</sup>  
KATSUMI TANAKA<sup>1,e)</sup>

## 1. はじめに

ネット社会の進展に伴い、ウェブメディアには多様な情報が集積され続けている。このようなウェブ情報をどのように探索、発見し、取捨選択するかについて、さまざまな方向からのアプローチがなされている。現在、多くの検索エンジンが実用化されるとともに、ウェブ情報検索やウェブ情報マイニングについて広く研究が行われている。現行のウェブ検索エンジンは、キーワードクエリに基づく文書検索モデルに基づくものが多く、この意味では、Webから文書ではなく情報そのものを検索するという意味でのウェブ情報検索は、いまだ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ブから検索する、いわば、タスク検索も未解決の問題の一つである。

何かを成し遂げたいが、どうすれば達成できるかわからないとき、現行のウェブ検索エンジンを使って手段・方法を見つけることが日常的にも良く行われている。たとえば“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で、“立体マスクをつける”や“アレロク

ク錠を飲む”というタスクを発見することができるが、検索漏れも多く、この意味で、タスク検索の再現率向上は大きな課題である。多様なタスク(答え)が得られた段階で、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ができるようになる。本研究では、「目的となるタスクを達成するために、どのようなタスクがあるか」を高い再現率で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。たとえば“花粉症対策をする”というクエリをウェブ検索エンジンに入力すると、出力として

- 立体マスクをつける。
- アレロク錠を飲む。
- 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)を発見することができる。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)がそれぞれのみで解であり、各解は、元のタスク「花粉症対策をする」に対して、instance-ofという汎化関係が成立する。また、同じクエリ「花粉症対策をする」に対して、次のような出力が得られる可能性もある。

- スギ花粉対策をする。
- カモガヤ花粉対策をする。
- ヒノキ花粉対策をする。

<sup>1</sup> 京都大学大学院 情報学研究所  
Graduate School of Informatics, Kyoto University

a) r.kato@dl.kuis.kyoto-u.ac.jp

b) ohshima@dl.kuis.kyoto-u.ac.jp

c) tyamamot@kuis.kyoto-u.ac.jp

d) kato@dl.kuis.kyoto-u.ac.jp

e) tanaka@dl.kuis.kyoto-u.ac.jp

この場合は、“スギ花粉対策をする”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) という汎化関係が成立する。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べて含むページが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。
- その店舗で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つのページが1つの解に対応する場合や、複数のページ (に記載されたタスク) の集約が1つの解に対応する場合があります。現行のウェブ検索エンジンによる解ページのランキングでは不十分であることが予想される。花粉症対策の方法は非常に多様であり、高くランクづけされたページであっても花粉症対策の方法のごく一部を含んでいるにすぎない場合もある。また、高くランクづけされたページ同士は内容が重複していることが多く、高ランクのページを見て回っても、発見で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が実用的なレベルに達すれば、以下に示すような、多くの応用例が考えられる。

- 入力クエリに関連するタスクをクエリ推薦としてユーザーに提示する。
- “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ザーに推薦する。
- 1つのページに複数のタスクが網羅的に記載されているページの発見や、多様なタスクを含んだページ集合のランキングといった、タスクベースのページランキング。

このようなタスク検索を実現するための第一歩として、本研究では、入力となるクエリを達成するためのタスクを含んだウェブページを網羅的に集めるための手法を提案する。提案手法で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で、入力クエリ単体では得ることがで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である。まず、2章で関連研究を紹介する。3章では、本研究で扱うタスクとサブ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章で提案手法について説明し、5章で評価実験について述べた後、最後に本稿をまとめる。

## 2. 関連研究

### 2.1 タスク検索

1章で述べたような、ユーザーのタスクに基づく検索の支援を目的とした研究は、近年注目を集めはじめています。た

例えば、Yamamoto らは、*search goal* と *subgoal* という概念を提案し、検索連動型広告 (スポンサーサーチ) のログを利用し、入力クエリに関連した *subgoal* を関連クエリのクラスタリングにより発見する手法を提案している [1]。彼らの定義によれば、*search goal* とは“ユーザーが達成したいと考えている行動”であり、ある *search goal*  $x$  を達成することで別の *search goal*  $y$  の一部あるいは全てを達成するとき、 $x$  は  $y$  の *subgoal* である、と定義している。我々が提案するタスク検索は、タスク間の階層的な関係を考えている点で Yamamoto らの *search goal* と *subgoal* の関係に近い。しかし、本稿ではタスクを行動だけではなくユーザーの状態の遷移という点から捉え、より詳細にタスクをモデル化している (詳しくは3章で述べる)。

また、Hassan らも、ユーザーのタスクを達成するための支援を行うための手法を提案している [2]。彼らは、あるウェブページに関する Open Directory Project (ODP) から得られるカテゴリを1つのタスクとして捉え、検索ログから関連するタスク集合を抽出することで、あるクエリに関連したタスク集合を自動的に発見する手法を提案している。ほかにも、湯本らは手順情報に着目し、ハウツー的な文書から手順情報を抽出し構造化する手法を提案している [3]。こうした、ある目的を達成するために必要な手順情報を発見することも我々が扱うタスク検索の一種であると捉えることができる。

### 2.2 検索連動型広告を利用したクエリ変換

検索連動型広告を対象とした研究の多くは、広告の検索精度の向上 [4] やクリックスルーレートの予測 [5]、検索連動型広告におけるユーザーの検索行為の分析 [6][7] などが主である。

こうした研究に対し、本稿では検索連動型広告に着目し、タスクを含んだウェブページを得るためのクエリ修正を行う。検索連動型広告に着目した同様の研究として、田麥らの研究がある [8][9][10]。田麥らは、ギターの買い取りやホテルの予約といった、さまざまなサービスを提供しているウェブページを効率的に発見する手法について取り組んでいる。彼らは、“中古ギターを売りたい”という情報要求を持ったユーザーの場合、“中古ギター 売却”というクエリでウェブ検索を行うよりも、“中古ギター 買取”のように、中古ギターの買い取り業者というサービス提供者側の視点に立ったクエリに変換してウェブ検索する方が、目的のサービスに関するページを得やすいと指摘している。彼らは、“売却”に対する“買取”のようなクエリの変換ルールとして、以下の6パターンを提案している。

- (1) 語の逆意関係：“中古ギター 売却”に対する“中古ギター 買取”のような、動作の主体の違いにより動詞の意味が逆転する動詞。
- (2) 語の同義関係：“荷物 送る”に対する“荷物 配送”の

ような、同義語の関係にある単語。

- (3) 語の併立関係：“部屋 借りる”に対する“部屋 探す”のような、動作の併立 (entailment) 関係にある動詞。
- (4) 語の兄弟語関係：“ぬいぐるみ 処分”に対する“ぬいぐるみ 供養”のような、同一の結果をもたらすが異なる意味の動詞。
- (5) タスクの価値を表す語：“ブランド品 売却”に対する“ブランド品 売却 高価”のような、タスクの価値を宣伝するような語。
- (6) タスク提供者が提供する語：“ダンス 習う”に対する“ダンス 学校”のような、タスクの提供者側と関連した語。

彼らの研究では、入力クエリとそれに関連した検索連動型広告から、上記ルールに基づいて、変換すべきクエリを抽出し、それをクエリ拡張に用いることで、サービスを提供するウェブページに関する検索の適合率および再現率が向上できることを示している。

### 3. タスクとタスク検索

本章では、我々が取り組むタスク検索の検索モデルについて説明する。1章で述べたように、ある入力タスクと、それを達成することに関わるタスクには、instance-of 関係や subtype-of 関係といった関係が存在する。タスク検索を行うためには、まずこうした多様なタスク間の関係をモデル化する必要がある。

#### 3.1 行動と状態に基づくタスクのモデル化

本稿では、行動と状態という2つの概念を用いてタスクをモデル化する。図1は、“花粉症の対策をする”ということを行動と状態でモデル化した図であり、2つの状態と1つの行動で表現されている。図中の2つの状態のうち、左側の“花粉症の対策をしていない状態”は、タスクを実行する前の状態であり、右側の“花粉症の対策をした状態”は、タスク実行後の、目標とする状態を表している。それらの状態間を遷移させるものが行動であり、ここでは“花粉症の対策をする”という行動があることを表している。このように、タスクというものを考慮するにあたって、状態とその状態間を遷移する行動が重要な要因となると我々は考えている。

状態と行動という2つの概念に基づき、本研究ではタスクを、ある初期状態から目的とする状態への遷移を可能とする、一連の行動、と定義する。

ここで、行動と状態という2種類の概念を用いて、“花粉症の対策をした状態”という目標を達成するためのタスクについて考えてみる。その候補として、たとえば、“マスクをつける”、“アレロック錠を飲む”、“スギ花粉対策をする”といった3つのタスクが考えられる。これら3つのタスクは、全て与えられた目標を達成するものであると考

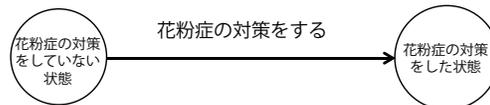


図1 “花粉症の対策”に関する状態と行動の例  
Fig. 1 States and its action for “treat pollen allergy.”

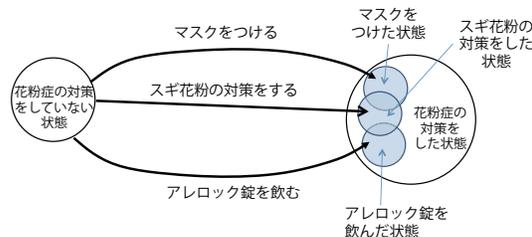


図2 “花粉症の対策をした状態”になるための複数の行動の例  
Fig. 2 Example actions for achieving “treat pollen allergy.”

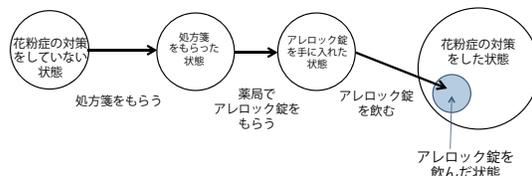


図3 複数の行動からなるタスクの例

Fig. 3 Example task achieved by multiple states and actions.

えられ、状態と行動のモデルに当てはめて考えると、図2のようなになる。ここでの考え方は、目標となる“花粉症の対策をした状態”は、“マスクをした状態”、“アレロック錠を飲んだ状態”、“スギ花粉対策をした状態”といった、花粉症対策に関連したさまざまな状態の集合であると捉えている。このように考えることで、“花粉”と“スギ花粉”のような subtype-of 関係の扱いと、“花粉”と“マスク”という概念構造的には階層関係が存在しない関係を、同様に扱うことが可能となる。

このモデルにおいて、行動は、一般に、複数の行動に分割することが可能である。行動と行動の間には、中間状態が存在することになる。たとえば、“アレロック錠を飲む”という花粉症の対策は、まず病院に行って処方箋をもらい、その後薬局でアレロック錠をもらい、そしてアレロック錠を飲むという、複数の行動のシーケンスととらえることができる。中間状態としては、“処方箋をもらった状態”や“アレロック錠を手に入れた状態”などが考えられる。これを、状態と行動のモデルで表現すると図3のようなになる。

#### 3.2 タスク検索

ここまでの議論を踏まえて、タスク検索について定義する。

まず、ユーザより入力として与えられるのは、Web 検索エンジンに入力されるようなクエリ文字列である。本稿におけるタスク検索とは、入力として与えられたクエリ文字列が明示的、または、非明示的に示す目標の状態への遷移

を実現する行動の集合を発見することと定義する。たとえば，“花粉症 対策”という入力クエリに対して，“マスクをつける”，“スギ花粉対策をする”，“アレロック錠を飲む”，“処方箋をもらう”といった行動が求める出力となる。

#### 4. 検索連動型広告を用いたクエリ拡張に基づくタスクを含むウェブページの発見

本稿では，3章で述べたタスク検索を実現するための第一歩として，タスクを含むウェブページを網羅的に発見するための手法を提案する。1章でも述べたとおり，通常のウェブ検索を行っただけでは，ある目的を達成するためのタスクを網羅的に収集することは難しい。そこで，本稿では，検索エンジンの検索連動型広告に着目し，そこからタスクに関連した動詞を用いて入力クエリを拡張することで，元の入力クエリから得られるウェブページには含まれていないようなタスクを含んだウェブページを発見する。

##### 4.1 検索連動型広告を用いたクエリ拡張

前章で説明したように，タスクとは，目的を達成するために必要な，ユーザが取るべき行動と考えることができる。Yamamotoらや田麥らは，そうしたタスクは検索エンジンの検索連動型広告と非常に関連が深いことを指摘している。現在，多くの商用検索エンジンでは，ユーザが入力したクエリに対して，広告主によって登録された広告のうち関連度の高いものが，検索結果ページ上に表示されるようになっている。この検索連動型広告とタスク検索の関係に関して，たとえば，Yamamotoらは，タスク検索は一般的に達成することが困難であるため，そのようなドメインに多くの広告が登録されており，1つ1つの広告は特定の行動やトランザクションに関連していると述べている [1]。また，田麥らは，あるサービスを検索する場合，その多くは検索連動型広告に出現しており，検索連動型広告を利用することでサービスに関するウェブを効率的に検索できることを示している [8]。

そこで，本稿ではクエリ拡張を行う際の拡張語を選択するための情報源として，検索連動型広告を利用する。検索連動型広告を利用することで，タスクに関連し，かつ通常のウェブ検索結果からでは得られないようなタスクに関連した語句がクエリ拡張語として得られることが期待される。

##### 4.2 手法の流れ

提案手法の概要を図4に示す。本手法は以下の流れでウェブページを取得する。

- (1) タスク（目的となる状態）を表したクエリ  $q$  を受け取る。
- (2)  $q_0$  でウェブ検索を行い， $n$  件の検索連動型広告  $\{a_1, \dots, a_n\}$  を取得する。
- (3)  $\{a_1 \dots a_n\}$  のタイトルおよびスニペットに対して形態

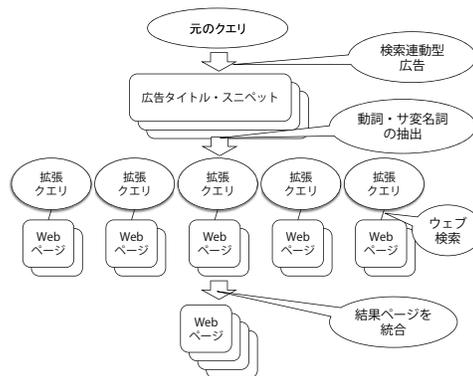


図4 提案手法の概要

Fig. 4 Overview of our method.

素解析を行い，出現する動詞およびサ変接続名詞を抽出し出現回数を計算する（得られた単語集合を  $W$ ，単語  $w \in W$  の出現回数を  $tf_w$  とする）。

- (4)  $tf_w$  の高い単語上位  $m$  件 ( $\{w'_1 \dots w'_m\}$  とする) を取得する。
- (5) ウェブページ集合を  $D = \phi$  とし，(4) で得られたそれぞれの単語  $w_i$  ( $1 \leq i \leq m$ ) について， $|D| = k$  となるまで以下のプロセスを繰り返す。
  - (a) 元のクエリに得られた単語を追加したクエリ  $q_i = q \wedge w_i$  を作成する。
  - (b) 得られたクエリ  $q_i$  でウェブ検索を行い，上位  $k'$  件のウェブページ  $\{d_1^{(i)} \dots d_{k'}^{(i)}\}$  を取得する。
  - (c) 得られたウェブページを  $d_1^{(i)}$  から順に  $D$  に追加 ( $D \leftarrow D \cup \{d_1^{(i)}\}$ ) する。このとき， $|D| = k$  となればステップ (6) へ移行する。
- (6) 得られたウェブページ集合  $D$  を返す。

上記の手法を適用することで，元のクエリ  $q$  で得られるウェブページ集合より多くのタスクを含んだ，つまり得られるタスクという観点で再現率の高いウェブページ集合を集めることができると考えられる。

#### 5. 評価実験

本章では，4章で提案したクエリ拡張手法が，元の入力クエリで得られた上位  $k$  件のウェブページから得られるタスクに対して，どの程度新しいタスクを含んだウェブページを発見できているのかを評価した。

##### 5.1 ベースライン手法と実験設定

提案手法の有効性を検証するため，検索連動型広告を用いず，通常のウェブ検索の検索結果を用いてクエリ拡張を行いウェブページを収集する手法をベースライン手法として用意した。具体的には，5.1節で示した手法のうち， $\{a_1, \dots, a_n\}$  を検索連動型広告ではなく，通常のウェブ検索結果上位  $n$  件とした手法である。

実験にあたりウェブ検索結果の取得には Google Custom

Search API<sup>\*1</sup>を用い、検索連動型広告の取得には、Yahoo! Japan が提供するスポンサードサーチ<sup>\*2</sup>を利用した。また、形態素解析には MeCab<sup>\*3</sup>を用いた。

節で述べた手法のパラメータについては、今回の実験では、 $n = 15$ ,  $k = 20$ ,  $k' = 5$  と設定した。すなわち、クエリ拡張を行い最終的に 20 件のウェブページを取得し、そのウェブページにどの程度、元のクエリで得られる 20 件のウェブページに含まれていないタスクが含まれているかを評価した。

### 5.1.1 実験に用いたクエリ

本稿では、“花粉症の対策をする”、“部屋を掃除をする”、“結婚する”、“コーヒーを淹れる”、“ダイエットをする” という 5 種類のクエリを用いて実験を行った。これらのクエリを選択した理由は、これらの目的を達成する手段が豊富に存在すると考えたためである。

## 5.2 評価尺度

提案手法の性能を評価するに当たって、以下の 2 種類の評価基準を設け、人手による検証を行った。

- **行動数**: クエリ拡張で得られた上位 20 件のウェブページ集合から得られた、目的の達成に必要な行動の数のうち、元のクエリで検索して得られた上位 20 件のウェブページには存在しなかった行動の数。
- **詳細度**: 新しく得られた行動のうち、その行動を達成するために必要な情報が得られたウェブページに書かれている場合はその行動に対して 2 点、行動が列挙されているだけでその行動に達成するために必要な情報が得られない場合は 1 点として上記の行動数を重みづけた尺度。

たとえば、クエリ拡張を行い新しく取得したウェブページの中から、“加湿器を使う”という行動が得られたとき、行動数を 1 として計算する。その際、もしウェブページに加湿器を購入するための情報や良い加湿器に関する情報などが含まれていたら詳細度を 2 点、そうでない場合は 1 点として計算する。

## 5.3 実験結果

得られた結果を表 1 に示す。まず、新しく得られた行動数に着目してみると、表 1 から分かるように、“部屋の掃除をする”、“結婚式する”、“コーヒーを淹れる”というクエリでは、目的を達成するために必要な行動として、提案手法によりクエリ拡張を行うことで新しいものがベースラインよりも多く取得できていることが分かる。また、詳細度について着目してみると、表 1 から分かるように、“ダイエットをする”以外の 4 つのクエリについて、提案手

表 1 クエリ拡張により新しく得られた行動の数と詳細度

Table 1 Number of obtained actions and their detailedness by query expansion technique.

		提案手法		ベースライン	
		行動数	詳細度	行動数	詳細度
入 カ ク エ リ	花粉症の対策をする	11	16	11	13
	部屋の掃除をする	16	32	7	7
	結婚する	10	20	3	6
	コーヒーを淹れる	5	9	4	4
	ダイエットする	5	9	5	10

法がベースライン手法を上回っていることが分かる。これは、検索連動型広告を用いてクエリ拡張を行うことで、よりタスクの達成に特化した、タスクの達成に必要な情報が記載されたウェブページが取得できたことを示していると考えられる。

クエリ拡張語としてどのようなものが得られていたのかを分析すると、“部屋の掃除をする”というクエリについては、提案手法では“クリーニング”、“配送”、“洗浄”といった単語が得られていた。一方ベースラインでは、“上がる”や“維持”といった単語が得られており、検索連動型広告を利用することで、タスクに関連した単語がクエリ拡張語として得られていることが分かる。また、“部屋の掃除をする”というクエリでは、提案手法は“掃除の代行業者に頼む”というタスクを含んだウェブページが得られており、検索連動型広告に出現するクリーニングや洗浄といった単語が、こうしたタスクに関連したウェブページの発見を可能にしたと考えられる。

一方で、“ダイエットする”というクエリについては、どちらの評価尺度ともベースラインと比較して、高い値は得られなかった。提案手法では、“ダイエットをする”に対して、“吸引”、“体験”といった単語が拡張語として得られており、実際に“吸引”という語を追加することで脂肪吸引に関するウェブページを発見することができていた。一方で、ベースライン手法では“食べる”という語がクエリ拡張語として得られており、元のクエリでは既にダイエットに有効な食事に関するウェブページが多く出現していたものの、“食べる”を元のクエリに追加することで、より多様な食事に関するダイエット方法を記載したページが得られ、結果として提案手法よりも得られる行動数が多くなっていた。

## 6. 考察

### 6.1 クエリ拡張語のランキング

今回提案手法で得られたクエリ拡張語の中には、節で示したように、タスクに関連した動詞が得られていることもあったが、タスクとは関連しない単語も多くクエリ拡張語として得られていた。今回は、単純に出現頻度だけを利用してクエリ拡張語を選択したが、将来的には 2.2 節で紹介

\*1 <https://developers.google.com/custom-search/?hl=ja>

\*2 <http://search.yahoo.co.jp/search/ss>

\*3 <https://code.google.com/p/mecab/>

表 2 クエリ拡張により新しく得られた行動の subtype-of 関係、instance-of 関係による分類結果

Table 2 Classification result of obtained actions in terms of subtype-of and instance-of relationships.

		提案手法		ベースライン	
		subtype-of	instance-of	subtype-of	instance-of
入力クエリ	花粉症対策をする	0	11	0	11
	部屋の掃除をする	1	15	0	7
	結婚する	2	8	2	2
	コーヒーを淹れる	0	5	2	2
	ダイエットする	1	4	0	5

したような、タスク提供者側と利用者間の単語の意味関係に踏み込んで、クエリ拡張に利用する語を抽出することが必要であると考えられる。

## 6.2 タスク間の関係に基づく結果の分析

3章では、目標となる状態を達成する際に、instance-of 関係や subtype-of 関係にある状態を達成するようなタスクがタスク検索の検索対象であることを述べた。そこで、クエリ拡張により新しく得られたタスクが、どの程度このような関係に関連しているのかを人手で分類した。このとき、あるタスクが元のクエリが示すタスクの subtype-of 関係であるとは、得られたタスクが示す名詞あるいは動詞が、元のクエリが示すタスクの名詞または動詞の subtype-of 関係となっていることとした。たとえば、“コーヒーを淹れる”という入力クエリに対して、“アイスコーヒーを淹れる”というタスクは、“アイスコーヒー”が“コーヒー”の subtype-of 関係であるため、“アイスコーヒーを淹れる”は“コーヒーを淹れる”の subtype-of 関係である。また、“花粉症の対策をする”に対して“マスクを着用する”のように、それ単体で入力クエリの解となっているようなタスクを instance-of 関係であると判定した。

表 2 は、得られたウェブページ中に記載されていたタスクを subtype-of 関係か instance-of 関係に基づいて分類した結果である。表から分かるとおり、今回得られたタスクの多くは instance-of 関係にないタスクが得られていることが分かる。これは、今回の手法は動詞に着目した手法であり、“コーヒー”に対する“アイスコーヒー”のような、subtype-of 関係にあるようなタスク間の関係を全く考慮していないためであると考えられる。このような、subtype-of 関係にあるようなタスクは、それ単体ではユーザの入力クエリとなる達成に役に立つ情報であるとはいえない。しかし、subtype-of 関係にあるようなタスクを発見することができれば、それを基にそのタスクの達成に必要なタスクを再帰的に発見していくことで、入力タスクにも適用可能なタスクが発見可能となると考えられる。さまざまなタスクを網羅的に取得するためには、動詞のみならず、名詞の階層の関係なども考慮しながらページを取得する必要があると考えられる。

## 7. まとめと今後の課題

本稿では、ある目的を達成するための方法を検索するというタスク検索を提案した。また、より多くのタスクを含んだウェブページを収集するために、検索連動型広告を利用したクエリ拡張手法を提案し、予備実験を行った。

今後タスク検索に関する研究を進めていくにあたり、さまざまな事柄について考えていく必要がある。今回提案した手法はウェブページを取得する手法であり、得られたウェブページがどのようなタスクを含んでいるかは人手で判断し評価を行った。今後は、単にウェブページを抽出するだけではなく、そこからどのようなタスクが存在するかを言語パターンなどを用いて抽出してることが必要になると考えられる。また、評価方法についても考案の余地がある必要がある。6章で述べたように、今回の評価尺度は順序関係を持ったタスクを正しく評価することができない。また、得られたタスク集合の多様性も評価において重要な観点であると考えられる。こうしたことを考慮した、タスク検索における統一的な評価尺度を考えることも今後の課題である。

**謝辞** 本研究の一部は、文部科学省科学研究費補助金(課題番号 24240013, 24680008)によるものです。ここに記して謝意を表します。

### 参考文献

- [1] Yamamoto, T., Sakai, T., Iwata, M., Yu, C., Wen, J.-R. and Tanaka, K.: The wisdom of advertisers: mining sub-goals via query clustering, *Proc. of CIKM*, pp. 505-514 (2012).
- [2] Hassan, A. and White, R. W.: Task tours: helping users tackle complex search tasks, *Proc. of CIKM*, pp. 1885-1889 (2012).
- [3] 湯本高行: 手順の対応関係に基づくハウツー情報の階層化, 第 5 回データ工学と情報マネジメントに関するフォーラム (2013).
- [4] Broder, A., Fontoura, M., Josifovski, V. and Riedel, L.: A semantic approach to contextual advertising, *Proc. of SIGIR*, pp. 559-566 (2007).
- [5] Richardson, M., Dominowska, E. and Ragno, R.: Predicting clicks: estimating the click-through rate for new ads, *Proc. of WWW*, pp. 521-530 (2007).
- [6] Jansen, B. J. and Resnick, M.: Examining Searcher Perceptions of and Interactions with Sponsored Results, *Workshop on Sponsored Search Auctions* (2005).
- [7] Danescu-Niculescu-Mizil, C., Broder, A. Z., Gabrilovich, E., Josifovski, V. and Pang, B.: Competing for users' attention: on the interplay between organic and sponsored search results, *Proc. of WWW*, pp. 291-300 (2010).
- [8] 田麥節治, 是津耕司, 木俣豊, 田中克己: サービス情報検索と適格性に基づくランキング, 第 156 回データベースシステム研究発表会 (2012).
- [9] 田麥節治, 赤星祐平, 是津耕司, 木俣豊, 田中克己: Web サービスを対象とした検索意図推定とその応用, 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012) (2012).
- [10] 田麥節治: サービス適格性と語彙分析に基づくサービス情報検索, 京都大学大学院情報学研究所 修士論文 (2013).