

語の出現パターンと意味関係分析を用いた Webからのタスク検索

加藤 龍^{a)} 大島 裕明^{1,b)} 山本 岳洋^{1,†1,c)} 加藤 誠^{1,†1,d)} 田中 克己^{1,†1,e)}

概要: 本研究では、クエリとしてタスクが与えられた際に、そのタスクを達成するために必要なサブタスク群を Web から発見する手法を提案する。提案手法では、動詞の含意関係や逆意関係にもとづいたルールによりクエリの変換を行う。変換したクエリで Web 検索を行う。得られたページから、タスク特有の言語パターンを用いて、サブタスク候補となるフレーズを発見する。

キーワード: タスク検索

Subtask search

RYO KATO^{a)} HIROAKI OHSHIMA^{1,b)} TAKEHIRO YAMAMOTO^{1,†1,c)} MAKOTO P. KATO^{1,†1,d)}
KATSUMI TANAKA^{1,†1,e)}

Abstract: This study is about subtask search. We propose a new method to find subtasks from the Internet. When a query is given, our method find subtasks which are needed to accomplish the task represented by the query. In our method, the system convert the query according to relationships of words such as the relationship of includingness and contradiction. The system find phrases from the pages which are listed up by our method. After that, we use language patterns which are specified in tasks to find phrases which are possible to be correct subtask.

Keywords: Task search

1. 動機

誕生以来、Web には多様な資料が集積され続けている。情報をどう走査、発見し、取捨選択するか、様々な方向からのアプローチがなされている。現在、いくつかの検索エンジンが実用化され、Web を舞台に情報発見や集約を行っている。だが、Web における情報検索は、いまだ多くの問題を抱えている。そのうちの一つとして、目的を達成する

手段を発見したいとき、できるだけ多くの方法を探そうとしても意外と見つけれないことがあげられる。

なにかを成し遂げたいが、どうすれば成功するかわからないとき、Web 検索を使って実現方法を考える行為がよく行われている。たとえば「花粉症対策をする方法」をクエリに検索することで、「立体マスクをつける」や「アレロク錠剤を飲む」を発見することができる。だが、そうして発見できた「花粉対策をする方法」を採用すべきなのか、簡単にはわからない。まだ発見できていない「花粉対策をする方法」のほうがよいかもしれないからだ。

こうした状況では、多くのユーザーは「タスクを遂行するためにどんな方法があるか」をできるだけ多く発見するサブタスク検索を求める。多様な答えが得られた段階で、初めて、安心して各方法を比較したり、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。

¹ 情報処理学会
IPSI, Chiyoda, Tokyo 101-0062, Japan

^{†1} 現在、京都大学
Presently with Kyoto University

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

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

^{c)} yamamoto@dl.kuis.kyoto-u.ac.jp

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

^{e)} ktanaka@i.kyoto-u.ac.jp



図 1 アプリケーションの例
Fig. 1 Application.

サブ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
- アレロック錠剤を飲む
- 医師の診断を受ける
- 植物に近寄らない

といったように、複数の異なった選択肢を複数発見する。これがサブタスク検索の例である。このように、できるだけ多くの手法を探そうとする検索は現在の一般的な検索エンジンでは困難である。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。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。

このようなサブタスク検索が実用的なレベルに達すれば、図 1 のように、現在の検索エンジンが実装しているサジェスト機能にサブタスク検索を組み込むといった応用が考えられる。

サブタスク検索は、タスクを遂行する方法が複数あり、どれがベストかわからない状態で役割を果たす。図 1 は、まさしくその一例である。クエリに「iPhone ゲーム 開発」と入力した結果、通常の Web 検索であっても iPhone ゲームの作り方についての Web ページを非常に多量に発見できる。しかし、通常の検索のみでは人気の高いページ

だけが上位にランクづけされてしまう。そのため、iPhone のゲームを開発するためにやるべき行動をすべて網羅できない。このとき、「App store の審査を受ける」といった、やらねばならないことだが、上位ページではあまり言及されないタスクに気づかない、といった問題が発生する。こうした状況でサブタスク検索を行うことで、iPhone のゲームを開発するために必要なタスクがすべて明らかになり、タスク遂行のためのロードマップを描きやすくなる。

逆にサブタスク検索が得意なこととして、遂行する方法がひとつのタスクがある。たとえばすでに手元に本があるとき「手元の本を読む」というタスクを遂行する方法はひとつである。

サブタスク検索が有効に働くのは、多くの、時間のかかる作業をせねばならないときである。

タスク検索は、Web 検索のなかでもかなりの割合を占めている。サブタスク検索を実現することは、Web での検索体験を向上させる点で非常に有意義である。

以下、本稿では可能な限り多様な方法を Web から発見する手法を提案し、評価する。

1.1 タスク検索

1 章で述べたような、ユーザ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る。たとえば、Yamamoto らは、*search goal* と *subgoal* という概念を提案し、検索連動型広告（スポンサードサーチ）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“ユーザが達成したいと考えている行動”であり、ある *search goal* x が別の *search goal* y の *subgoal* であるとは、“ y を達成することが x の一部あるいは全てを達成するとき、 y は x の *subgoal*”である、と定義している。我々が提案するタスクとサブタスクの関係は、ユーザの行動を基に階層的な構造を考えている点で 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の状態の遷移という点から捉え、さらにタスクの並列性と直列性といった関係を提案している。また、彼らの研究は関連クエリをサブタスクという観点でクラスタリングしていたのに対して、我々の手法はサブタスクを含むウェブページを発見しているという点で、彼らの研究と異なる。

また、Hassan らも、ユーザ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]。

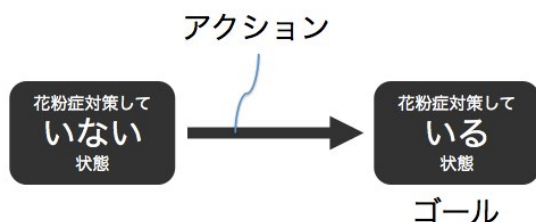


図 2 アクションとゴールの例
Fig. 2 Action.

こうした、ある目的を達成するために必要な手順情報も、その目的のサブタスクの 1 つであると捉えることができる。

1.2 検索連動型広告

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、サブタスクを得るためのクエリ修正を行うが、検索連動型広告をそうした目的で利用した研究は少ない [1]。

2. タスク関係の概要

本章ではタスクとゴールの概念を述べ、タスクを遂行する行動がどのようにモデル化されるかを説明する。

2.1 ゴールとアクション

ある目的を達成するためにどのような順番でどのような行動を取るかを考察する際、タスク、ゴール、アクションといった用語の定義が必要となる。本稿ではゴールとアクションを以下のように定義する。

- ゴールとは、目的の一部または全部を達成した状態である
- アクションとは、ある状態から別の状態に移行する行為である

ゴールとアクションを図示すると図 2 のようになる。「花粉症対策している状態」がゴールであり、「花粉症対策をしていない状態」から「花粉症対策している状態」に移行する行為がアクションである。図 2 では、ひとつのアクションでゴールに移行しているが、実際はひとつのアクションだけでゴールに移行できるとは限らない。多段階でのゴール移行を表現すると、図 3 のようになる。

図 3 はサブゴールという概念を表現している。「花粉症対策をしていない状態」から「花粉症対策している状態」に移行する間に、別の「花粉症について調べている状態」というものを経由している。

図 4 に図示したように、ひとつのゴール（スーパーゴール）

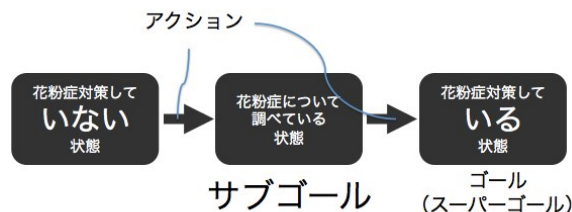


図 3 サブゴールの例
Fig. 3 Subgoal

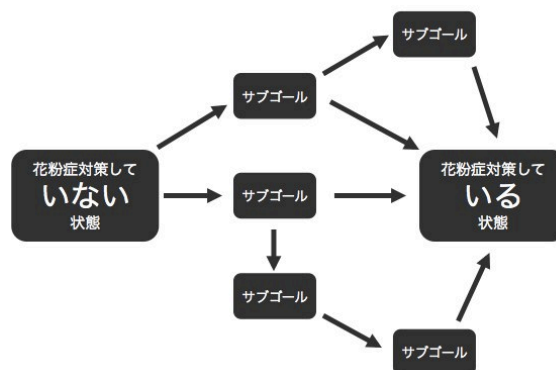


図 4 複数のサブゴール
Fig. 4 Many sub goals

ル) にサブゴールは複数ありえる。スーパーゴールとサブゴールの関係は、後述のスーパータスクとサブタスクの関係と同様に、相対的なものである。サブゴールからサブゴールへの移行もアクションのひとつである。

2.2 タスクとサブタスク、スーパータスク

次に、タスクの定義を説明する。

- タスクとは、遂行すると、目的の一部あるいは全部を達成したことになるアクションである
- サブタスクとは、タスク同士の関係に着目したとき、「遂行するともう一方のタスクの一部を遂行したことになる行動」である。このとき「もう一方のタスク」はスーパータスクである

つまり、なんらかのゴールに至るアクションがタスクといえる。ゴールとタスクの関係を図 5 に図示する。状態を変化させる行為すべてがアクションであるが、アクションのうち、目的の一部あるいは全部を達成したものだけがタスクである。

タスクもゴールと同様、複数の階層に分割できる。図 6 にスーパータスクとサブタスクの関係を図示する。

このタスクを分割すると、まず「花粉症対策をしていない状態」から「花粉症対策している状態」に移行するタスクをスーパータスクとみなし、「花粉症対策をしていない状態」から別の何らかの状態への移行がひとつめのサブタスク、

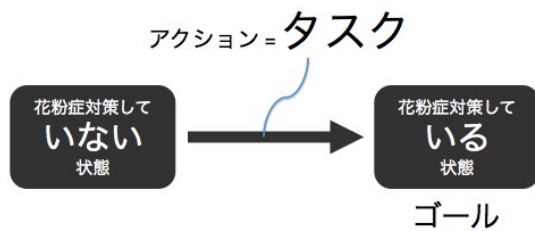


図 5 タスクの例
Fig. 5 Action Task.

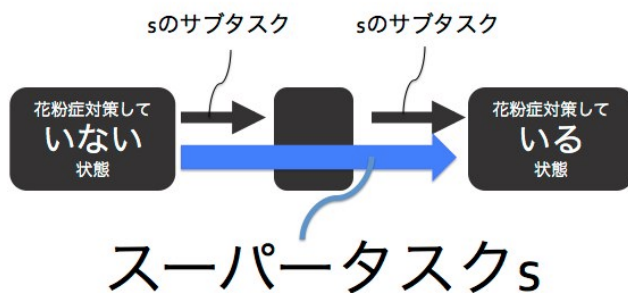


図 6 スーパータスクの例
Fig. 6 Supertask.

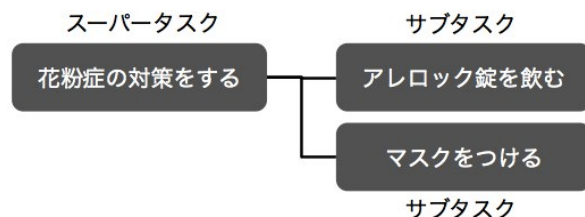


図 7 スーパータスクとサブタスクの関係
Fig. 7 Relationships of super task and sub tasks

何らかの状態から「花粉症対策している状態」への移行がふたつめのサブタスクだとみなせる。

2.3 木構造によるタスク分割

上記のタスクの説明は状態遷移図によるタスク構造のモデル化である。表現を変え、タスクを and-or 木構造と見なすこともできる。実際はタスクを遂行する順序やあるタスクを遂行できて初めて遂行可能になるタスクなども存在するため、and-or 木構造よりももっと複雑な構造となる。

図 7 に示すように、サブタスクとスーパータスクは相対的な関係である。あるサブタスクは、別のタスクをサブタスクだとして見ればスーパータスクにあたることもある。

例えば「花粉症対策をする」がスーパータスクだと仮定すると、サブタスクとして「アレロック錠を飲む」、「マ

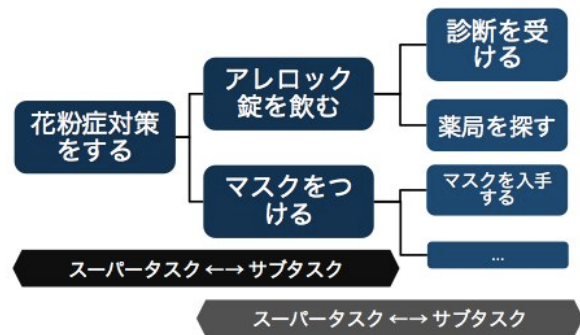


図 8 スーパータスクとサブタスクの多重関係
Fig. 8 Multiple relationships of super task and sub tasks

スクをつける」といったタスクがあげられる。ここで一段階視点を下げて「アレロック錠を飲む」をスーパータスクと捉えた場合、「診断を受ける」、「薬局を探す」といった行動がサブタスクとなる。逆の順番でいえば、「薬局に行く」というサブタスクから見たとき「アレロック錠を飲む」はスーパータスクだが、「アレロック錠を飲む」をサブタスクとした場合スーパータスクは「花粉症対策をする」となる。

ここで注意してほしい点がある。「花粉症対策をする」のサブタスクとして「薬局に行く」もありうる。二段階、三段階下のタスクであってもそれはサブタスクなのである。逆に、「薬局に行く」のスーパータスクとして「花粉症対策をする」がある。つまり、複数段階の連続的な関係と、サブ・スーパーの一段階の関係は両立できる。多重のスーパータスクーサブタスク関係を図 8 に図示する。サブタスクのサブタスクのサブタスク、といったように、タスク分割は複数段階行えることが多い。

ただし、それ以上分割できないタスクが存在し、これをアトミックタスクと呼ぶ。アトミックタスクは空の状態から何らかの状態に移行するアクションである。

ある目的を達成するために行動を起こす際、どのような手順で状態が変異するのか整理をする。まず最終的な目的として「花粉対策をする」というタスクがあるとすると、これは「花粉症対策をした」というゴールと等価である。

ユーザーは「花粉症対策をしていない」という状態から「花粉症対策をした」という状態に遷移するために必要なアクションを発見しようとする。

「花粉症対策をする」というタスクを遂行する方法はいくつも存在する。「アレロック錠を飲む」、「マスクをつける」、「医師の診断を受ける」、「薬局に行く」これらすべてが「花粉対策をする」のサブタスクである。

2.4 サブタスク、スーパータスクの問題

2.3 のようにスーパータスクをサブタスクに分割する際、どの程度の粒度でタスクを分割すべきか、という問題があ

る。スーパータスクをサブタスクに分割するとき、サブタスクとスーパータスクは is-a 関係にある。極論すればタスク分割は何度でも行えるが、分割されすぎたタスクは望ましいタスク検索の結果とはならない。

3. サブタスク検索の手法

本章では、スポンサードサーチを利用したサブタスク検索を提案する。

3.1 サブタスク検索における入力と出力

本稿で扱うサブタスク検索の入力と出力を述べる。ユーザーは「(動詞)」または「(名詞)(助詞)(動詞)」というクエリを入力する。一例として「花粉症の対策をしたい」という意図でタスク検索を行う場合、「花粉症対策をする」といったクエリが考えられる。入力の例として、以下のものがあげられる。

- 花粉症対策をする
- 部屋を掃除する
- 二日酔いを覚ます
- ダイエットする
- 快眠する
- おいしいコーヒーを淹れる

以上の例はタスクを動詞として表現したものであるが、実際にはタスクの表現としては動詞以外のものも考えられる。「花粉症対策」「掃除」「快眠」といったものがその例である。

さらに、タスクの表現として「花粉症対策をした状態」つまりゴールをそのまま使うこともできる。ゴールとタスクは別のものであるが、クエリとして用いる場合、どちらも同様のものと見なせる。ゴールとタスクの違いは前述の通りであるが、サブタスク検索においては、違いを無視して利用できるのである。

図 9 に図示したように、初期状態を持っていないことを除けば、ゴールもタスクとして扱える。本来のタスクとは、状態が目的の状態に変化するアクションであるので、ゴールも初期状態を考えなければ「状態に変化する」というアクションと同様に、クエリに使える。整理すると、まずタスクは以下のようなものである。

- タスクは初期状態とゴールが与えられれば成立する
- タスクはアクションである
- ゴールは以下のようなものである。
- ゴールには初期状態が与えられない。不要である。
- ゴールは状態である

以上の特徴から、タスクとゴールは明らかに違うものであるが、ゴールを「初期状態のないタスク」として扱うことができる。

しかし本稿ではゴールによるタスク表現は扱わない。あくまで「(動詞)」または「(名詞)(助詞)(動詞)」のみを

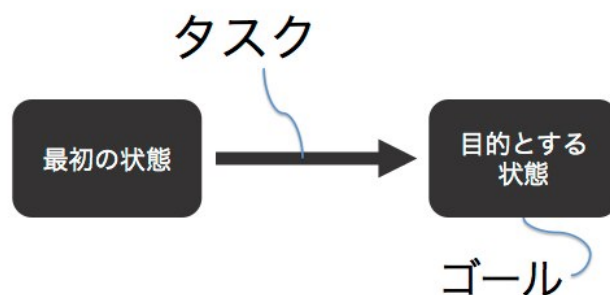


図 9 ゴールがタスク表現としても扱える

Fig. 9 Multiple relationships of super task and sub tasks

タスクとして扱う。

「花粉症対策をする」をクエリとした場合の、サブタスク検索の出力を例示する。

- 立体マスクをつける
- アレロック錠剤を飲む
- 医師の診断を受ける
- 植物に近寄らない

以上のように、クエリをスーパータスクとした場合のサブタスクである「(名詞)(助詞)(動詞)」の文章が出力される。

3.2 提案手法

本稿ではスポンサードサーチを用いたサブタスク検索を提案する。スポンサードサーチとは、ユーザーの入力した検索クエリに対応した広告を表示するシステムである。Google や Bing, Yahoo! など、多くの検索エンジンがスポンサードサーチを提供している。スポンサードサーチによりユーザーは自分の入力したクエリに適した広告ページを見ることができ、広告主は広告の商品に適したユーザーにアプローチすることができる。

3.2.1 スポンサードサーチを用いたサブタスク検索

スポンサードサーチを利用して、以下のような手法でサブタスク検索を実現できる。

- (1) 言語パターンに基づいてクエリを変換する
- (2) 変換したクエリをもとにスポンサードサーチを行う
- (3) 得られた広告のタイトルとスニペットから動詞・サ変接続名詞を抽出する
- (4) 抽出した単語でクエリを拡張する
- (5) 拡張したクエリで検索を行う
- (6) 検索結果ページよりサブタスクを発見する

本稿は、このような手法によるサブタスク検索を提案する。最初の手順である言語パターンによるクエリ変換は、単語間の上位語、下位語などのエンティリメント関係に着目して行う。

スポンサードサーチに着目した理由を述べる。スポンサードサーチには、商品を購入しそうなユーザーの目に留まるように多くの工夫がなされている。具体的には、商品に対応したクエリを入力したときに広告が表示されるように設定されており、スニペットの文章は商品のニーズを端的に表現している。これは?が指摘した通りである。このような特徴を利用すれば、サブタスク検索に使えると我々は考えている。

ユーザーが目的とするゴールに到達するために、スポンサードサーチに広告を出している企業の商品を使うことは有効な場合が多い。本稿の4で用いた「花粉症対策をする」といったクエリであれば、「立体マスクをつける」「花粉症の薬を飲む」といったサブタスクはどちらも商品を使用することで達成できる。

また実際に商品そのものを使わなくとも、何らかの手段で商品の効能が発揮された状態になれば、目的が達成されることも多い。「花粉をシャットダウンする」という広告スニペットから「花粉をシャットダウンすることが花粉症対策のサブタスクである」と推論することができる。

4. 評価

本章では、提案手法によりサブタスク検索を行い、その性能を相対的再現率により評価する。

4.1 実験の手法

提案手法とベースライン手法を述べる。

4.1.1 スポンサードサーチを利用したサブタスク検索

提案手法では、Yahoo! Japan が提供するスポンサードサーチ^{*1}を利用してサブタスク発見を試みる。言語パターンよりサブタスク表現を探す際、形態素解析に MeCab^{*2}を用いる。提案手法の処理を順番にすると以下ようになる。

- (1) クエリを形態素解析して名詞と動詞を抽出する
- (2) 名詞と動詞をキーワードにした新たなクエリを作成する
- (3) 新たなクエリで Yahoo! Japan のスポンサードサーチを行う
- (4) 最大 15 件のページとタイトル、スニペットを取得する
- (5) 発見したスニペットに Mecab で形態素解析を行い、動詞とサ変接続名詞を取得する
- (6) 取得した単語のうち、出現頻度の高いもののみをピックアップする
- (7) 頻出単語でクエリを拡張し、Web 検索を行う
- (8) 検索結果ページトップ 20 件内で発見できたサブタスクを数える

以上の手順で発見できたサブタスクを数える。クエリ拡張を行わずに Web 検索を行い、上位 20 件中から取得でき

るサブタスクと比較し、クエリ拡張を行ったことにより発見できたサブタスクを数え、評価する。数え方、評価の手法は 4.1.4 で述べる。

4.1.2 ベースライン手法

またベースラインとして、単純な言語パターンを用いてクエリ拡張を行い、サブタスク検索を行うシステムを使う。提案手法と異なる点はクエリ拡張を行う際の手順のみであり、拡張したクエリで検索を行うこと、また、それ以降は提案手法と同様の手順で実験を行う。

- (1) クエリを形態素解析して名詞と動詞を抽出する
- (2) 名詞と動詞をキーワードにした新たなクエリを作成する
- (3) Google Custom Search API を使い、新たなクエリで Web 検索する
- (4) 上位 20 件のページを取得する
- (5) 取得したページのタイトル、スニペット内から品種つする動詞とサ変接続名詞を取得する
- (6) 頻出単語でクエリを拡張し、Web 検索を行う
- (7) 検索結果ページトップ 20 件内で発見できたサブタスクを数える

提案手法と同様に、ベースラインもクエリ拡張を行わずに Web 検索した結果と比較し、クエリ拡張を行ったことにより発見できたサブタスクを数え、評価する。

4.1.3 実験に用いるクエリ

実験に用いるクエリとして、サブタスク検索に使われるだろう以下の一般的なクエリを使用する。

- 花粉症対策をする
- 部屋を掃除する
- 二日酔いを治す
- ダイエットする
- 結婚する
- コーヒーを淹れる

以上のクエリから、サブタスクを探す。理想的なサブタスクとしては、たとえば「部屋を掃除する」をクエリとした場合、

- 掃除機をかける
- パナソニックの掃除機を買う
- 家事代行に依頼する
- 友達に掃除を頼む
- ルンバを使う
- お茶の出がらしをまく

といったサブタスクが出力されれば理想的である。「掃除機をかける」と「パナソニックの掃除機を買う」は並列したタスクではなく、「パナソニックの掃除機を買って、それから掃除機をかける」という連続の関係にあるタスクかもしれない。しかし本研究では連続したタスクであっても、別個のタスクとして扱う。よって、「掃除機をかける」と「パナソニックの掃除機を買う」は別のタスクだと判断

^{*1} <http://search.yahoo.co.jp/search/ss>

^{*2} <https://code.google.com/p/mecab/>

する。

ただし、この点は検討の余地がある。異なる製品であっても同一の結果を導くならば、同一タスクと見なすべきかもしれない。本稿では異なる製品であれば別タスクとしているが、

4.1.4 タスクの基準

発見数で評価を行う。スーパータスクをクエリとして検索を行い、発見できたページ内で発見できたサブタスクの個数を発見数とする。提案手法とベースライン、両者において発見できたサブタスクすべてを正解セットとし、正解セット中どれだけ発見できたかを評価、比較する。サブタスク候補がサブタスクとして正解であったかどうかは、人の手で判断する。判断基準を説明する。例えば、「花粉症対策をする」というクエリを入力したとして、

- 乾燥機やエアコンで代用する
- 加湿器を使う
- 冷えを解消する
- ほどほどにする
- 花粉と触れる部位を減らす
- 寄生虫を飼う
- 自分の体調に合えば使う
- 汁を摂って綿棒で鼻内の粘膜に塗る
- 薬剤師に相談する
- 鼻の下はしっかり洗う
- ノウハウのメモをとる
- ラジオ体操をする

以上の文を含むページが返ってきたとする。このうち、実際のサブタスクだと判断できるものは以下の通りである。

- 加湿器を使う
- 花粉と触れる部位を減らす
- 寄生虫を飼う
- 薬剤師に相談する
- 鼻の下はしっかり洗う
- 冷えを解消する
- ラジオ体操をする

サブタスクではないと判断できるものは以下の通りである。

- 乾燥機やエアコンで代用する
- ほどほどにする
- 自分の体調に合えば使う
- 汁を摂って綿棒で鼻内の粘膜に塗る
- ノウハウのメモをとる

判断の基準として「それ単体で、遂行すべきサブタスクが何なのか明確であること」があげられる。「乾燥機やエアコンで代用する」では何の代用をするのかわからず、明確ではない。だが「ラジオ体操をする」は、判断が難しい。「ラジオ体操をする」ことが花粉症対策として有効であるかどうか、簡単には結論を下せない。世界のどこかに「花

表 1 「部屋の掃除をする」でのサブタスク検索の結果

Table 1 Result 1

	スコア	サブタスク数
ベースライン	7	2
提案手法	30	13

表 2 「コーヒーを淹れる」でのサブタスク検索の結果

Table 2 Result 2

	スコア	サブタスク数
ベースライン	4	4
提案手法	9	5

粉症対策にはラジオ体操がよい」と考える人々があり、彼らの間ではラジオ体操は花粉症対策として一般的かもしれないからだ。

こういった、判断しにくいサブタスク候補に対しては、「実際にサブタスクとして遂行している人が存在しているか」が基準となる。サブタスク候補を発見した Web ページを訪問し、サブタスク候補の文章が「スーパータスク遂行にはサブタスク候補の遂行が必要」と主張しているかで判断する。明らかに、サブタスクだと意図していない文脈であった場合、サブタスクではないと判断できる。

4.1.5 タスクの粒度

どの程度の粒度まで分割されたタスクを 1 タスクとして評価すべきか、議論すべきところである。本稿においては、「ぬれ雑巾で掃除をする」といった、具体的な行動の記述があれば 1 タスクと数える。ある程度具体的なスーパータスクとサブタスク関係にある 2 つのタスク、たとえば「マスクをつける」と「二重マスクをつける」を発見した場合、2 ではなく 1 タスクと数える。「免疫機能を制御する」といった抽象的なサブタスクは 1 タスクとは数えない。「免疫機能を制御するために薬を飲む」という記述があれば、十分具体的だと判断し、1 タスクと数える。

4.1.6 タスクの得点

サブタスクの記述の詳細度をもとに、サブタスクに 1 点あるいは 2 点を与える。具体的には、まとめサイトに頻出する、「クエン酸を使う」「クイックルを使う」といったサブタスクの列挙例は 1 点。Amazon の商品ページや、掃除代行料金見積もり計算の手順が詳しく載ったページであれば、記述されたサブタスクに 2 点を与える。

4.2 実験の結果

結果を図 7 に示す。クエリ拡張をした結果新たに発見できたサブタスクを、提案手法とベースラインで比較する。なお、スコアとサブタスク数はクエリ拡張をしないで検索した結果、発見できたサブタスクを除外していることに留意してほしい。つまり、クエリ拡張をした結果新たに発見できたサブタスクだけをカウントしている。

表 3 「ダイエットする」でのサブタスク検索の結果

Table 3 Result 3

	スコア	サブタスク数
ベースライン	10	5
提案手法	7	4

表 4 「結婚する」でのサブタスク検索の結果

Table 4 Result 4

	スコア	サブタスク数
ベースライン	2	1
提案手法	16	8

5. 考察

ベースラインの結果は、クエリ拡張を行わなかった場合の結果とさほど変わらない。それに対し、提案手法では、ベースラインが発見できなかった「掃除代行業者に見積もりを頼む」のようなサブタスクを発見できた。この違いの理由として、「掃除代行業者に依頼する」といった方法は一般的に知られていないことが挙げられる。

ベースラインで多様なサブタスクを発見できなかった理由として、Web ページのタイトルやスニペットはサブタスクやサブタスクの目的を記述していないことがある。

提案手法は提案は 15 件までしかスポンサードサーチ結果を取得できず、十分な量のスニペットを得ることができなかった。もっと多くのスポンサードサーチ結果を取得できれば、より性能の高いサブタスク検索を実現できると期待できる。

またスニペットは商品のニーズを端的に表したものであるが、必ずしもサブタスクをそのまま表現しているわけではない。これはスポンサードサーチのスニペットが広告文章であることが原因である。たとえば「花粉症対策をする」というタスクに対し、スポンサードサーチを用いたサブタスク検索では「注目」といったノイズを出力する例があった。「注目」という言葉は広告でユーザーの注意を引くためによく用いられる単語である。商品のアピールとしては有効であるが、サブタスクではない。

6. 結論

本稿では、スポンサードサーチに着目し、サブタスク検索を行った。スポンサードサーチの広告意図を利用したサブタスク検索を提案し、評価した。本研究の結果を整理すると以下ようになる。

- ゴール、タスクを定義した。
- スポンサードサーチを利用したサブタスク検索を提案した。
- 提案手法は、サブタスク発見のために有用であることを示した。

サブタスク検索を必要とするユーザーは多く、活用の状況がいくつか考えられる。将来の研究課題として、サブタスク候補からのノイズ除去、より広範なサブタスク発見があげられる。本稿ではスポンサードサーチとショッピングサイトを利用したサブタスク発見を行った。これは、広告やショッピングサイトにおいて m ユーザーのニーズに適した商品、商品に適した文章を用いて商品をアピールする工夫がなされていることを利用したものである。しかし、サブタスクを発見できる Web ページはこれらショッピングサイト、スポンサードサーチに限ったものではない。より多くのサブタスクを発見するために、別の点に着目した検索が必要になるだろう。

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, *Proceedings of the 21st ACM international conference on Information and knowledge management*, ACM, pp. 505–514 (2012).
- [2] Hassan, A. and White, R. W.: Task tours: helping users tackle complex search tasks, *Proceedings of the 21st ACM international conference on Information and knowledge management*, ACM, 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).