

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)} 田中 克己^{1,†1,d)} 加藤 誠^{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)} KATSUMI TANAKA^{1,†1,d)}
MAKOTO P. KATO^{1,†1,e)}

Abstract: This study is about subtask search. We propose a new method to find subtasks from the Internet.

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げら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「アレロック錠剤を飲む」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。

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

¹ 情報処理学会
IPSJ, 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)} ktanaka@i.kyoto-u.ac.jp

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



図 1 アプリケーションの例
Fig. 1 Application.

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。

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。

2. 関連研究

タスク検索に関する研究として、以下のものがある。

- ゴールにもとづいたクエリのクラスタリング
- タスクにもとづいた検索行動の推移
- タスクに適したウェブサイトを発見するサービス検索
山本らの研究 [1] において、スポンサードサーチ広告を用

いたクエリクラスタリングが試みられている。スポンサードサーチ広告は、できるだけ多くのユーザーの注目を集めることを目的としている。それも、広告の商品を使いたいだろうユーザーの注目を集めようとしている。こうした理由から、スポンサードサーチのスニペットやタイトルは、クエリに呼応した工夫がなされたものとなる。つまり、「クエリが意味するタスクを達成するために、広告の商品が必要になる」ことが、スニペットやタイトルを読むと理解できるような工夫がなされる。逆にいえば、タスクを表現するクエリを入力したときに出力されるスポンサードサーチ結果は、タスクを遂行するのに必要なものになる。本研究においてもこのアイデアを利用している。

ほかに、タスク遂行のための、複雑な検索の手順を研究したものとして Hassan らの [2] がある。タスクを遂行するのに、一度だけの検索では不十分なことが多い。タスクを達成するためには、通常、何度も検索を試行しなければならない。[2] は、こうした複数の検索におけるユーザーの行動を研究したものである。

田麦らによるタスク遂行のためショッピングサイトなどサービスを提供するページを発見する研究 [3] がある。「中古ギターを売る」といったタスクを遂行する際、「中古ギター 売却」というクエリで検索を行うよりも、「中古ギター 買い取り」というように、サービス提供者がわの視点での行動にクエリ変換したほうが、タスク遂行に適したページを発見しやすいことがある。[3] は、サービス提供者とサービス享受者の視点を軸にしたタスク検索研究といえる。

3. 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。

3.1 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。

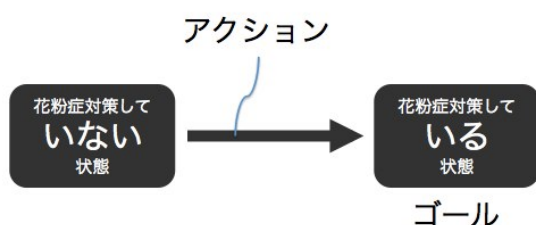


図 2 アクションとゴールの例
Fig. 2 Action.

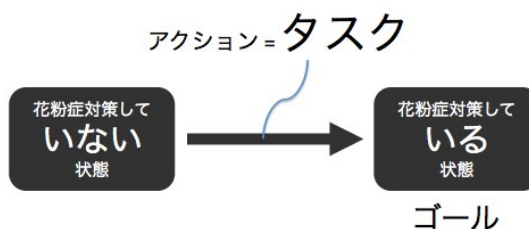


図 5 タスクの例
Fig. 5 Action Task.

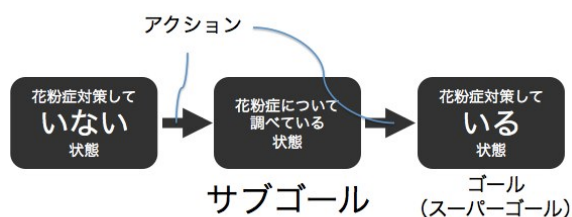


図 3 サブゴールの例
Fig. 3 Subgoal

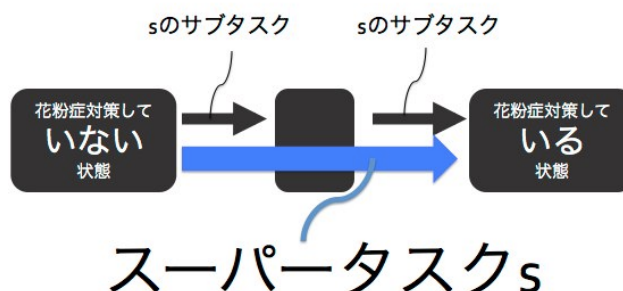


図 6 スーパータスクの例
Fig. 6 Supertask.

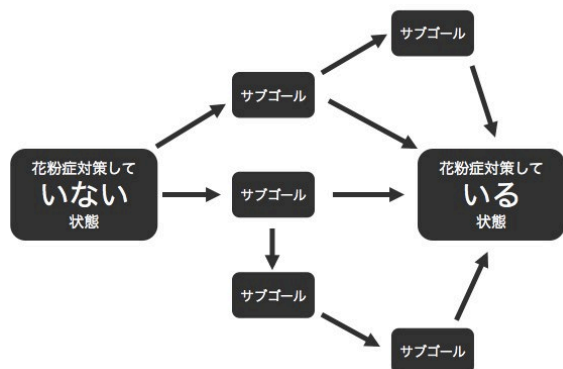


図 4 複数のサブゴール
Fig. 4 Many sub goals

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）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。

3.2 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、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。

3.3 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やるタ

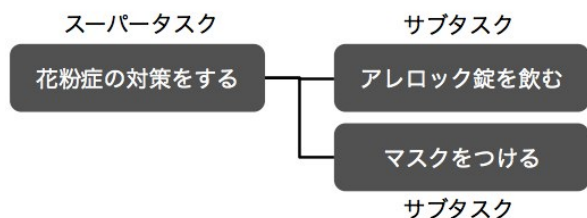


図 7 スーパータスクとサブタスクの関係

Fig. 7 Relationships of super task and sub tasks

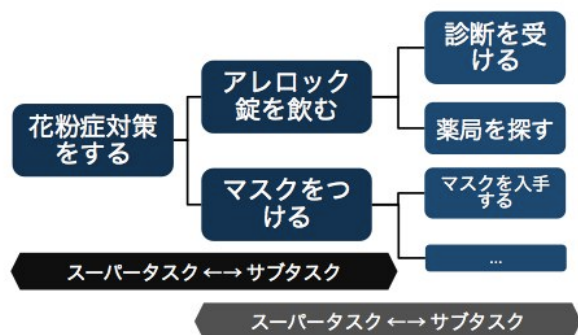


図 8 スーパータスクとサブタスクの多重関係

Fig. 8 Multiple relationships of super task and sub tasks

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「アレロク錠を飲む」、「マスクをつける」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「アレロク錠を飲む」はスーパータスクだが、「アレロク錠を飲む」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。

4. 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。

4.1 サブタスク検索における入力と出力

「(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」つまりゴールもタスク表現になりうる。図 9 に図示したように、初期状態を持ってないことを除けば、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

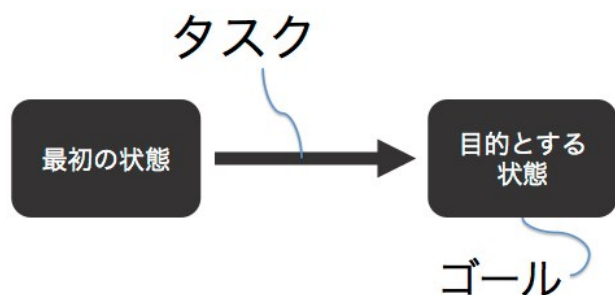


図 9 ゴールがタスク表現としても扱える

Fig. 9 Multiple relationships of super task and sub tasks

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。

4.2 提案手法

本稿ではスポンサード検索を用いたサブタスク検索を提案する。スポンサード検索とは、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検索を提供している。スポンサード検索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。

4.2.1 スポンサード検索を用いたサブタスク検索

スポンサード検索を利用して、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検索を行う
- (3) 得られたページ情報からサブタスクを抽出する

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。

5. 評価

本章では、提案手法によりサブタスク検索を行い、その性能を適合率と発見数により評価する。

5.1 実験の手法

提案手法では、Yahoo! Japan が提供するスポンサード検索^{*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 のスポンサード検索を行う
- (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) スニペットの文章に
- (6) 発見した文章に Mecab で形態素解析を行い、サブタスクとして適した表現に変換する

以上の手順でサブタスクを発見する。

またベースラインとして、単純な言語パターンを用いてサブタスク検索を行うシステムを使う。

ベースラインは上位にランクづけされたページをスクリーピングして、タスクに使われることの多い「使う」「おすすめ」といった単語を見つけるとその

- (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 100 件のページを取得する
- (5) ページ内から「使う」「おすすめ」といったタスクに使われる可能性の高い単語を含む文章を発見する
- (6) 発見した文章に係り受け解析を行い、サブタスクとして適した表現に変換する

以上の手順で、タスクと推測できる文を発見する。

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

以上のクエリから、サブタスクを探す。理想的なサブタ

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

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

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。ここで注意していただきたいことがある。「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」という連続的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する。

5.1.1 評価手法

適合率と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、発見できたサブタスク候補のうち、実際にサブタスクであったものの割合を適合率、実際にサブタスクであったものの個数を発見数とする。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

以上の出力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

Subtask Search

The image shows a web interface for 'Subtask Search'. It has three main sections: 1. A search bar with the text '花粉症対策をする' (Doing something about allergies) and a green button 'Search by search engine'. 2. Radio buttons for 'Google' (selected) and 'Bing'. 3. An 'Input URL' field with a green button 'Free scraping'. Below that is an 'Input your task' field with a green button 'Search by ads'.

図 10 ベースライン

Fig. 10 Multiple relationships of super task and sub tasks

表 1 花粉症対策の方法タスク検索

Table 1 An Example of Table.

	発見数	正解数	正解率
ベースライン	131	item 2,1	—
row4	item 1,4	item 2,4	item 3,4

- 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かわらず、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。

5.2 実験の結果

結果を図 7 に示す。ベースラインの結果ページは図 11 のようになった。

[花粉症対策を考えるよい食べ物、悪い食べ物って何？\(1/3ページ\)](#)

- を理解するために、もう少し詳しく免疫の仕組みを説明しましょう

[花粉症対策につながる4つの真実！-薬剤師ネット](#)

- たお薬を選ぶために、医療機関や薬剤師に相談してください
- で「眠くなりにくい内服薬」の特集をされていたので、気になる方はチェックしてみてください
- この5つのノウハウをメモしておいてください
- や薬剤師さんに相談してください
- 花粉症の症状かな・・・？と感じた方は、できるだけ早く医療機関を受診するようにしてください
- 合は、特に鼻の下はしっかり洗うようにしてください
- 記事のノウハウで乗り切ってください
- では防ぎきれない花粉が増えるそうで、そうなると効果が弱くなるとのことなので注意して下さい
- たお薬を選ぶ

[【花粉症対策】家に花粉を入れない方法・除去する方法-NAVERまとめ](#)

- 加湿器を使う

図 11 ベースラインの結果

Fig. 11 Multiple relationships of super task and sub tasks

6. 考察

図 11 のように、ベースラインの手法では、「医療機関や薬剤師に相談する」「加湿器を使う」「花粉症の市販薬を選ぶ」といったタスクを発見できる。しかし「むずむずするときに飲んでみてください」といったタスクの断片や、そもそもタスクではない「設定してください」といった文も発見してしまう。単純に言語パターンを用いるだけではタスクを表現する文章を発見しにくく、発見できた文章がタスクかどうかのフィルタリングも簡単ではないことが明らかとなった。

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。

7. 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。今後、サブタスク検索が役割を果たす。複数の応用例が考えられる

参考文献

- [1] The Widom of Advertisers: Mining Subgoals via Query Clustering 入手先 <http://research.microsoft.com/en-us/people/tesakai/cikm2012yamamoto.pdf> (2012.11.02).
- [2] Task Tours: Helping Users Tackle Complex Search Tasks: Ahmed Hassan, Ryan W. White 入手先 <http://research.microsoft.com/pubs/178868/HassanCIKM2012.pdf>
- [3] Tamugi search: 入手先 <http://somewhere.pdf>