

タスクの汎化と特化に着目した Web からのタスク検索

加藤 龍[†] 大島 裕明[†] 山本 岳洋[†] 加藤 誠[†] 田中 克己[†]

[†] 京都大学情報学研究科 〒 606-8501 京都府京都市左京区吉田本町

E-mail: [†]{r.kato,ohshima,tayamamoto,kato,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本稿では、クエリとしてある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タスク集合を Web から発見する「タスク検索」を提案する。提案手法では、実行することで入力タスクを達成できるタスクを Web 検索エンジンを用いて発見し、一連のタスクをまとめることでタスク集合を形成し、タスク集合の達成度（それを実行することで入力タスクを達成できる度合い）によってランキングを行う。タスクの汎化・特化関係を考慮することで、同種のタスクを同定しその頻度を利用することでタスクの達成度を推定する。

キーワード 情報検索, タスク検索

1. はじめに

検索エンジンが多くのユーザに利用されるようになるにつれ、その用途は情報獲得を超え多岐に渡るようになってきた。例えば、調査や学習、意思決定などのために行う検索は探索的検索として分類され近年注目を集めており、商用検索エンジンにおいても検索エンジンを「意思決定を支援するためのサービス」と位置づける動きがある。これらの例のように、情報獲得自体を主目的とするのではなく、現実世界においてあるタスクを達成することを主目的とした検索が徐々に注目され始めている。

我々は、タスク達成を主目的とした検索をより直接的に支援するため、「あるタスクを達成するために必要なタスクの検索」、すなわち、タスク検索を提案する。例えば、タスク「花粉症の対策をする」を入力としたとき、タスク検索では「マスクをつける」、「眼鏡をかける」、「抗アレルギー薬を飲む」、「耳鼻咽喉科に行く」など、入力されたタスクを達成するために実行すべきタスクを出力する。ユーザは従来の検索のように文書を逐次閲覧することなく、タスク遂行のために必要十分な情報を獲得することができるため、検索からタスク達成までの時間が短縮されることが期待できる。

本論文では、特にタスク検索において出力するタスクの構造、および、ランキングに焦点を当て、タスク検索を実現する方法について述べる。タスク検索では入力と出力は共にタスクであるため、以降、混乱を避けるため入力されるタスクを入力タスク、出力されるタスクを出力タスクと呼ぶ。

タスク間には様々な関係があるため、出力タスクの列を結果として出力することは適切ではない。単一のタスクを実行するだけでは入力タスクを達成できない場合があるためである。前述のタスク「花粉症の対策をする」を入力タスクとしたとき、「抗アレルギー薬を飲む」という出力タスクを遂行すれば、入力タスクを達成したことになる。しかし、「耳鼻咽喉科に行く」という出力タスクを遂行しても、それだけでは入力タスクを達成したことにはならない。「耳鼻咽喉科に行く」、「花粉症の治療を受ける」という2つの出力タスクを実行することで、初めて入力タスクを達成したことになる。前者のように、遂行される

ことにより入力タスクを完全を達成できる出力タスクは入力タスクと INSTANCE-OF 関係にあるとし、後者のように、遂行されることにより入力タスクの一部を達成できる出力タスクは入力タスクと PART-OF 関係にあるとしタスク検索において区別する。本論文で提案するタスク検索では、特に入力タスクと PART-OF 関係にあるような出力タスクを集合としてまとめて出力する問題として扱う。すなわち、タスク検索の出力は順序付きタスク集合族となる。

タスク検索を行うユーザにとって最も重要なのは「そのタスクに関する情報を見つける」ことではなく、「そのタスクが達成できる」かどうかである。花粉症対策の例を用いれば「タバコを控える」という出力タスクは花粉症対策への達成度は高くない、これも入力タスクを達成するための出力タスクとしてはあまり重要ではない。上記の例からわかるように、タスク検索において重要な尺度は、出力タスクの達成度（それを実行することで入力タスクを達成できる度合い）であり、従来の情報検索で考慮されてきたトピック適合性や多様性だけでは適切に出力タスクを評価することはできない。そこで、我々はタスク間の汎化・特化関係を考慮しつつ、タスクの言及頻度に着目することで達成度を考慮して出力するタスクをランキングすること提案する。我々は「あるタスクが出現する頻度が高いとき、そのタスクの達成度は高い」という仮定によって達成度を推定する。この仮定は必ずしも成り立つとは限らないが、多くの人に言及されるタスクはその有効性が広く認知されていると考えることができる。我々はタスクの汎化・特化関係を利用することでタスクの表記ゆれを考慮し、より精度良く達成度を推定する方法を提案する。

本論文の構成は以下の通りである。まず 2. 節において関連研究を紹介し、本研究の位置づけを述べる。次に 3. 節においてタスク間の関係と、タスク検索の定義、タスク集合を評価する尺度について説明する。4. 節にて提案手法について述べる。5. 節で評価実験に関する説明を行い、実験結果の考察を行う。最後に 6. 節で本論文の結論を述べる。

2. 関連研究

タスク、もしくはそれに類するゴールや手順、サービスなど、現実世界で何らかの行動を起こすことを前提とした検索がいくつか提案されている。たとえば Yamamoto [6] らは *search goal* と *subgoal* という概念を提案した。彼らは検索連動型広告を利用して、クエリログのクラスタリングを行うことで、ユーザーの入力したクエリに適した *subgoal* を発見する手法を提案している。*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 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。

また、Hassan らは、タスクを達成するための支援を行う、タスク検索と同じ目的の手法を提案している [2]。Hassan らの手法は、ある Web 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。このような手順により、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、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 [7]。こうした、「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と同様のものとみなすことができる。

ユーザーの目的に着目した同様の研究として、田麥らの研究がある。田麥らは、家具の売却やケーキの予約のような、サービスを提供する Web ページを検索することを目的としている [3] [4] [5]。彼らはクエリ変換を行うことで、より目的とするサービスを検索しやすくする手法を提案している。たとえば「自分が持っているギターを売りたい」というユーザーが中古ギターを売却するサービスを検索するとき、「中古ギター 売却」というクエリで検索するよりも「中古ギター 買取」というクエリで検索したほうが、より目的とする中古ギター買取サービスを発見しやすい。これは、サービスの受け手の視点に立ったクエリから、サービス提供者側の視点に立ったクエリへの変換である。田麥らは「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 種を提案している。

(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。

彼らの提案手法は検索連動型広告を利用したものである。提案手法は、入力クエリとそれに関連した検索連動型広告情報から、上述した 6 種のルールに基づいて、変換すべきクエリを抽出している。こうして得た語をクエリ拡張に用いて、サービスを提供する Web ページの検索を行う。実験の結果、Web 検索の適合率と再現率が向上できることを彼らの研究は示している。

3. タスク検索

本節では、タスク間の関係と、タスク検索の定義、タスク検索システムを評価する尺度について説明する。

3.1 タスク間の関係

タスク間には、INSTANCE-OF 関係と SUBTYPE-OF 関係、PART-OF 関係という 3 種の関係がある。以下ではそれぞれの関係について定義を与える。

3.1.1 INSTANCE-OF 関係

タスク t_i がタスク t_j の INSTANCE-OF 関係にあるということは、「 t_i を遂行することで、 t_i を完全に達成する」という関係を意味し、 $\text{instanceOf}(t_j, t_i)$ で表現する。たとえば「マスクをつける」という行動を遂行することで、「花粉症対策をする」というタスクを完全に達成したことになるため、「マスクをつける」というタスクは「花粉症対策をする」というタスクに対して INSTANCE-OF 関係にあるといえる。タスク t_j がタスク t_i と INSTANCE-OF 関係にあるとき、 t_j はタスク t_i を具体化したものであるともいえる。

3.1.2 SUBTYPE-OF 関係

タスク t_j がタスク t_i の SUBTYPE-OF 関係にあるということは、「 t_j が t_i のうち特定の種類のタスク」である関係を意味し、 $\text{subTypeOf}(t_j, t_i)$ で表現する。たとえば「スギ花粉の対策をする」というタスクは、「花粉症対策をする」というタスクの特殊な場合である。よって、「スギ花粉の対策をする」というタスクは「花粉症対策をする」というタスクの SUBTYPE-OF 関係にある。SUBTYPE-OF 関係も一種の IS-A 関係である。つまり、タスク t_j がタスク t_i の SUBTYPE-OF 関係にあるとき、 t_j は t_i が特化したものである。以降、簡単のため、入力タスクに対して SUBTYPE-OF 関係にあるタスクを SUBTYPE タスクと呼ぶ。

3.1.3 PART-OF 関係

タスク t_j がタスク t_i の PART-OF 関係にあるとは、「 t_j を遂行することで、 t_i の一部分を達成できる」という関係を意味し、 $\text{partOf}(t_j, t_i)$ で表現する。たとえば「耳鼻咽喉科に行く」というタスクは「花粉症対策をする」というタスクに対して PART-OF 関係にある。「耳鼻咽喉科に行く」といった行動を遂行しただけでは花粉症対策にはならないが、「花粉症の治療を受

ける」という行為のためには必要なタスクである．このように，PART-OF 関係は複数の連続的なタスクにおいて発生する．また，タスク t_j がタスク t_i の PART-OF 関係にあるとき， t_j は t_i が具体化したものであるといえる．

3.2 タスク検索

タスク検索とは，入力タスクを t_q としたとき，その要素をすべて遂行することで t_q を達成することができるタスク集合の集合を出力する問題である．ここで， t_i を達成できるようなタスク集合 T とは，以下のいずれかである：

(1) タスク t_i と INSTANCE-OF 関係にあるタスクを唯一の要素とするタスク集合 $T_i = \{t_j\}$ ($\text{instanceOf}(t_j, t_i)$)，および，

(2) 集合中の要素すべてを実行することで t_i を達成できる， t_i と PART-OF 関係にあるタスクの集合 $T_p = \{t_j | \text{partOf}(t_j, t_i)\}$ ．

タスク検索の出力 $\mathcal{T}$ は上記のいずれかのタスク集合を要素とするタスク集合族となる．達成例えば，入力タスクとして「花粉症対策をする」が与えられたとき，出力 $\mathcal{T}$ は INSTANCE-OF 関係にあるタスクを唯一の要素とするタスク集合 { 抗アレルギー薬を飲む } や PART-OF 関係にあるタスクの集合 { 耳鼻咽喉科に行く，花粉症の治療を受ける } などを含む．また，出力 $\mathcal{T}$ には順序が存在し，達成度（もしそのタスクを実行したとき，どの程度入力タスクを達成するかを示す尺度）が高い順に順序つけられる．

3.3 タスク検索の評価尺度

前述の通り，タスク集合の出力は順序が定義された集合族であり，各タスク集合はその要素すべてを実行することで入力タスクが達成できることが要求される．しかし，この条件を満たすタスク集合を形成するためには正確なタスク間関係の推定が必要となるため，現実的には以下で述べるようなタスク検索の評価尺度によってタスク検索の出力を評価することになる．タスク検索の出力は達成度（Completeness），純度（Purity），多様度（Diversity）という3つの尺度で行われる．

3.3.1 達成度

達成度は，もしそのタスクを実行したとき，どの程度タスクを達成できるかを示す尺度である．例えば，「花粉症対策をする」という入力タスクに対し「病院に行く」というタスクが出力されたとき，その達成度は「水を飲む」や「説明書を読む」よりも高い．タスク検索の出力は順序づけされたタスク集合であるため，集合内のすべてのタスクを実行した場合に入力タスクを達成できる度合いで達成度の評価を行う．

3.3.2 純度

タスク検索の出力の定義においては要求されていないが，あるタスク集合の中に異なるタスクを遂行するためのタスクが含まれることは出力として望ましくない．例えば，「コーヒーを淹れる」という入力タスクが与えられたとき，それと SUBTYPE-OF，PART-OF 関係にあるようなタスクを図1に示す．「コーヒーを淹れる」という入力タスクに対して { 氷を作る，カップを用意する，湯を沸かす，豆を選ぶ，湯を注ぐ } というタスク集合が出力された場合，「湯を沸かす」や「湯を注ぐ」といったタス

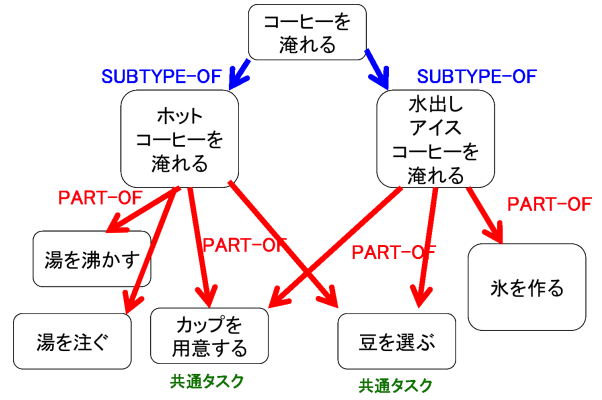


図1 複数の SUBTYPE-OF と PART-OF の関係

クは入力タスクの SUBTYPE タスクである「ホットコーヒーを淹れる」の達成にのみ寄与するが，「氷を作る」というタスクは「水出しアイスコーヒーを淹れる」という別の SUBTYPE タスクの達成にしか寄与しない．これらのタスクを同一のタスク集合に含めるのはタスク検索の利便性を損ねるため，出力されるタスク集合の一貫性を示す尺度である「純度」を用いて評価する．

ある入力タスク t_q に対して，その SUBTYPE タスクの集合を $S = \{t | \text{subtypeOf}(t, t_q)\}$ ，ある出力タスク集合を T とする．このとき，タスク集合 T の純度は，ある SUBTYPE タスクと PART-OF 関係にあるタスク集合のうち，大きさが最大のものがタスク集合 T において占める割合，と定義され，以下の式で求められる：

$$\text{Purity}(T) = \frac{\max_{s \in S} |\{t | \text{partOf}(t, s) \wedge t \in T\}|}{|T|}. \quad (1)$$

たとえば，「コーヒーを淹れる」という入力タスクに対して { 氷を作る，カップを用意する，湯を沸かす，豆を選ぶ，湯を注ぐ } というタスク集合 T が出力されたとする．「コーヒーを淹れる」という入力タスクには「水出しコーヒーを淹れる」と「ホットコーヒーを淹れる」という SUBTYPE タスクがあり，タスク集合 T の部分集合 { カップを用意する，湯を沸かす，豆を選ぶ，湯を注ぐ } が SUBTYPE タスク「ホットコーヒーを淹れる」，{ 氷を作る，カップを用意する，豆を選ぶ } が SUBTYPE タスク「水出しコーヒーを淹れる」と PART-OF 関係あるため，タスク集合 T の純度は SUBTYPE タスク「ホットコーヒーを淹れる」と PART-OF 関係にあるタスクの割合，4/5 となる．

3.3.3 多様度

多様度もタスク検索の定義では要件とされていなかったが，実際にタスク検索を実現する上では重要な尺度である．類似したタスク集合が検索結果として提示される場合，各々のタスク集合の達成度が高かったとしても，得られる情報量は多くないと考えられるためである．多様度は検索結果中の上位10件における SUBTYPE タスクの数で評価を行った．

4. 提案手法

本節ではタスク検索を実現する方法について述べる．ただし，ここで説明する手法は3節で定義したタスク検索の出力に必

ずしも沿うものではない．これは，タスク間の関係や達成度を正確に推定することが困難であるためである．本論文の手法では，タスク集合のうち達成度が高いと推定されたものを出力することによって，これらの集合の要素を実行することで入力タスクが達成されることを期待する．また，達成度の順にタスク集合が出力されているかを情報検索における評価指標を利用することで評価する．

我々はまず本手法におけるタスクの抽出方法について述べ，その後，汎化・特化関係に基づいた達成度の推定方法について説明する．また，一貫性のあるタスク集合を形成する手法，および，タスク集合族の多様性を高める方法についても提案する．

4.1 タスクの抽出

まず本提案手法にて利用されるタスクの表現方法をここで定義する．日本語によるタスクの表現は多彩だが，多くのタスクは「水を撒く」「証明書に記名する」のように，目的語と述語だけで言い換えることが可能である．「結婚する」のように目的語のないタスクも存在するが，今回の手法では主に目的語と述語で表現されるようなタスクにのみ着目する．また，多くの場合，目的語に続く「を」や「に」のような格助詞は異なっていたとしても，意味が同じ場合が多いため，今回は機能語は考慮せずタスクを目的語と述語の組で表現する．より正確には，あるタスク t は名詞 n と動詞 v の組によって表現され下記のように表す：

$$t = \langle n, v \rangle \quad (2)$$

例えば，入力タスク〈花粉症, 対策する〉に対して，〈マスク, つける〉や〈耳鼻咽喉科, 行く〉などが出力タスクとして得られることになる．

次に，入力タスクを達成するために実行する必要があるタスクの候補を，Web 検索エンジンを用いて発見する方法について述べる．入力タスクをそのまま検索クエリとして Web 検索エンジンで検索を行った場合，入力タスクを達成するために必要なタスクが検索結果中に現れることは多い．しかし，検索結果中のすべての名詞と動詞の組がタスクを表すわけではないため，まず，タスクを含む可能性が高いセンテンスを発見し，そこからのみ名詞と動詞の組を抽出して出力タスクの候補とした．

本提案手法では，タスクが含まれているセンテンスは命令文の形式を取っていることが多いことに着目し，命令文特有の表現を含むようなセンテンスを検索結果から取得する．その後，得られたセンテンスに対して形態素解析を行い，機能語を除いた後で連続する名詞と動詞の組をタスクとして抽出する．以下にこのタスク抽出の方法についてまとめる：

- (1) 入力タスク t_q を検索クエリとして Web 検索を行い， n 件の Web ページを取得する．
- (2) 各 Web ページから，命令文に頻出するパターン集合 P のうちいずれかを含むセンテンスのみを抽出する．
- (3) センテンスそれぞれに対して形態素解析を行い，機能語を除いた後で連続する名詞 n と動詞 v の組をタスク $t = \langle n, v \rangle$ とする．

命令文に頻出するパターン集合 P の詳細については 5. 節にて

詳しく説明する．

4.2 達成度の推定

Web 検索によって出力タスクの候補を抽出した後，個々のタスクの達成度を求め，達成度が低いタスクをノイズとして除外する．我々は「あるタスクの Web 上での言及頻度が高いときそのタスクの達成度は高い」という仮説を立て達成度の推定を行った．ある程度以上の達成度を持つタスクは Web ページに記述される頻度が高くなる傾向があるといえる．なぜなら，高い達成度のタスクであれば，その有効さが口コミや体験談で広まり，より多くの Web ページにおいて紹介されると考えられるためである．インターネットの利用が広く一般化した現在において，Web 上にタスクを遂行した実体験を投稿する行為は多く見られ，達成度の定義にある「あるタスクを実行することで別のタスクを実際に達成できた」ことを記述する例は多い．こうしたことから，言及頻度の高いタスクは達成度が高い傾向にあると仮定している．ただし，言及頻度の高いタスクが，必ずしも達成度の高いものであるとは限らない．Web 上に頻繁に出現するが，実際には達成度の低いタスクは存在する．たとえば，低品質の商品を販売する企業が，販売業績を伸ばすために，達成度が低いにもかかわらず多くのページで宣伝する事態が考えられる．

タスクの言及頻度を計測するにあたり，タスク表現が多様であるという問題がある．同じタスクであっても，Web 上では異なる表現で記述されている場合が多々ある．「耳鼻咽喉科に行く」と「診療科で受診する」がその一例である．「花粉症対策」で Web 検索を行った結果，これらの文が検索結果ページから発見されたとき，前者の「耳鼻咽喉科」が後者の「診療科」とおなじものを指していることを推測できる．また前者の「行く」が「受診する」を意味すること，Web ページの文脈から理解できる．しかしながら，単純な単語の一致によるタスク同定では上記の推測を行うことはできない．

この問題に対応するため，我々はタスクの汎化に着目した．タスクの汎化により，ふたつのタスクが同じ行動を指しているかどうかを推定することができる．本研究では，2 つの語の組み合わせという表現に統一したタスクを汎化するにあたり，名詞の上位下位関係と，動詞における含意関係に着目した．上位関係と含意関係は，どちらも IS-A 関係であるとみなせ，名詞と動詞をそれぞれを上位語，含意する動詞で汎化することができる．あるタスクが $t = \langle n, v \rangle$ で表現されているとし， $\text{hyp}(n)$ を n の上位語集合， $\text{ent}(v)$ が v が含意する動詞集合であるとき，名詞 $n' \in \text{hyp}(n)$ と動詞 $v' \in \text{ent}(v)$ を含む以下のタスクは元のタスク t を汎化したタスクであるという： $\langle n', v \rangle$ ， $\langle n, v' \rangle$ ， $\langle n', v' \rangle$ ．タスク t' が上記のいずれかであるとき，タスク t' はタスク t を汎化したタスクであるといい， $t \rightsquigarrow t'$ と表記する．たとえば「耳鼻咽喉科」の上位語には「診療科」や「診療科」があり，「受診する」が含意する動詞には「行く」，「相談する」，「探す」，「訪れる」といったものがある．よって，もとのタスクが〈耳鼻咽喉科, 受診する〉であるとき，名詞と動詞の汎化した語の組み合わせにより以下のように汎化されたタスクが得られる．

- $\langle n', v \rangle$ 例: $\langle \text{診療科, 受診する} \rangle, \langle \text{診療科, 受診する} \rangle,$
- $\langle n, v' \rangle$ 例: $\langle \text{耳鼻咽喉科, 行く} \rangle, \langle \text{耳鼻咽喉科, 相談する} \rangle,$
 $\langle \text{耳鼻咽喉科, 探す} \rangle, \langle \text{耳鼻咽喉科, 訪れる} \rangle,$
- $\langle n', v' \rangle$ 例: $\langle \text{診療科, 行く} \rangle, \langle \text{診療科, 相談する} \rangle, \langle \text{診療科, 探す} \rangle, \langle \text{診療科, 訪れる} \rangle.$

タスクの汎化を考えることにより, 同種のタスクを同定し, その言及頻度をより正確に推定できる. あるタスク t の Web 上での言及頻度を $f(t)$ としたとき, タスク t と同種のタスク集合 $S = \{s | \exists t'(t \rightsquigarrow t' \wedge s \rightsquigarrow t')\}$ を用いて, 汎化を考慮した言及頻度 $f'(t)$ は以下のように求められる:

$$f'(t) = \sum_{s \in S} f(s). \quad (3)$$

すなわち, 同種のタスクの言及頻度の和をそのタスクの言及頻度であるとする. 達成度は汎化を考慮した言及頻度 f' に等しいとし, ある閾値以下の達成度をもつタスクは除外される. また, 次小節で説明される手法によりタスク集合を形成した後, タスク集合のランキングに達成度が利用される.

4.3 タスク集合の形成

我々はあるページ内に出現するタスク集合 T がすべて同一のタスクを達成するためのタスクであると仮定する. また, このページが入力タスク t_q を含むとき, タスク集合 T はすべてを実行することで入力タスク t_q を達成できるようなタスクであるとも仮定する. これはそのページを執筆したユーザが, 入力タスク t_q を達成するために必要なタスクをそのページに記述し, また, それらがすべて同じ t_q の SUBTYPE タスクを達成するためのタスクであることを仮定することと同じである. ただし, これらの仮定は必ずしも正しいとは限らない. 単一のページに複数種類の SUBTYPE タスクの達成に必要なタスクが記述されている場合もあり, あるページのタスク集合 T が入力タスク t_q を達成するために十分である保証はない. しかし, これらの例外が比較的起こりにくいことから, 本小節の冒頭で述べた仮定を用いることでタスク集合を形成する. すなわち, あるページ内に出現するタスク集合 T をタスク検索により出力するタスク集合族の要素として用いる.

4.4 多様性を考慮したタスク集合のランキング

Web ページからタスクを抽出しタスク集合を形成した後, 得られたタスク集合の集合は含まれるタスクの達成度の和によってランキングされる. しかし, 似たタスク集合は同じ達成度になりやすいため, 上位にランキングされたタスク集合が互いに似てしまう場合がある. そこで, 出力するタスク集合族の多様性を高めるため, 我々はタスク集合の達成度とタスク集合間の非類似度を同時に考慮することによって, タスク集合のランキングを行う. この目的のため, 我々は Maximal Marginal Relevance [1] に基づいたランキングを行う. 入力タスク t_q に対し, まずタスク検索の出力であるタスク集合族を空集合で初期化しこれを $\mathcal{T}$ とする. その後, 逐次, 下記で定義される関数 $rank$ の値が最も高いタスク集合 T を $\mathcal{T}$ に追加していく:

$$rank(t_q, T) = \lambda \sum_T f'(t) - (1 - \lambda) \max_{T' \in \mathcal{T}} \text{Sim}(T, T'). \quad (4)$$

ここで, タスク集合間の類似度は $\text{Sim}(T, T') = |T \cap T'| / |T \cup T'|$ で定義される. 最終的にタスク集合族 $\mathcal{T}$ を上記の手順において追加した順でランキングを行う.

5. 実験と考察

本節では, 提案手法の有効性を確かめるため, 我々が行った実験を説明する.

5.1 設定

達成度

Web 検索には Bing Search API^(注1) を用いた. 得られた Web ページ中の文章を「」「!」「?」のいずれかで区切り, 各ページにひもづいたセンテンスを得た. 得られたセンテンスから命令文に頻出するパターン集合 P のいずれかを含むようなセンテンスだけから, 名詞と動詞の組で構成されるタスクを抽出した. 本実験では, パターン集合 P として下記のパターンを利用した.

- ください・下さい
- なさい
- べき
- ましょう
- しょう
- といい・とよい・と良い
- 必要が
- ては

フィルタリングしたセンテンスを, MeCab^(注2) で形態素解析し「目的語, 格助詞, 述語」の形式にのっとったタスク候補を多量に集める. そして格助詞の部分は無視し「目的語と述語」のペアを作る.

上位語と下位語を求める際は ALAGIN フォーラム^(注3) の上位下位関係抽出ツール^(注4) で作ったデータベースを利用した. 含意関係には同じく ALAGIN フォーラムの動詞含意関係データベース^(注5) を使った.

5.2 結果

我々は提案手法とベースライン手法, 2つの手法で実験を行った. 提案手法は 4. 節で説明した手法を用いた. ベースライン手法は提案手法とほぼ同じであるが, 汎化と特化による言及頻度計算を行わず, 単純な出現頻度でランキングしている. つまり, 4. 節において達成度をそのタスク自体の言及頻度 f によって求めている. また, MMR アルゴリズムによるランキングの入れ替えを行っていない. 表 1 に, “部屋を掃除する” という入力タスクでタスク検索した結果の上位 3 件を例として示す. 提案手法は評価手法に比べ, 一貫性に優れている, すなわち, 純度が高いことがわかる.

表 1 提案手法とベースライン手法の結果比較

順位	提案手法	ベースライン
1	新聞 使う	きれい する
	全体 拭く	額 差し引く
	拭き する	状態チェック 済ませる
	不動産業者 確認する	明細 貰う
	契約 確認する	新聞 使う
	説明書通り 利用する	説明書通り 利用する
	換気 充分する	確認 する
	きれい する	管理者 連絡する
	状態チェック 済ませる	契約 確認する
	額 差し引く	様 する
	管理者 連絡する	拭き する
	確認 する	賃貸借契約書 チェックする
	賃貸借契約書 チェックする	掃除 する
	換気 する	不動産業者 確認する
		換気 充分する
		全体 拭く
		換気 する
2	一緒 やる	運氣 乱す
	掃除 する	きれい する
	引越し業者 決める	リビング ある
	ハタキ 利用する	状態 整理する
	汚れ ためる	世話 する
	きれい する	植物 育てる
	洗剤 かける	優先 片付ける
	フキン 拭く	そこ 片付ける
	マスク 活用する	リビング 置く
	雑巾がけ する	掃除 する
	商品内容 確認する	分別 行う
	物 置く	数 減らす
	洗剤 使う	
3	掃除 する	空間 散らかる
	世話 する	公園デビュー する
	運氣 乱す	工作 する
	状態 整理する	空間 想像する
	リビング ある	モテ護身術 マスターする
	優先 片付ける	身 持つ
	きれい する	クリスマスイヴ 掃除する
	数 減らす	ゴミ 出す
	分別 行う	入浴剤 解消する
	リビング 置く	これから 書く
	そこ 片付ける	掃除 する
		鬼 退治する
		Arufa さん 見習う
		アイツ 感謝する
		変質者 撃退する
		写真 見る
		これから 頑張る

5.3 評価

表 2 に提案手法でタスク検索を行った結果を、表 3 に提案手法での結果を示す。両者とも 8 つのクエリで検索を行った。検索結果のうち、上位 10 件で、達成度・純度・多様性で評価した。

5.3.1 達成度の評価

4 人の評価者が出力された各タスク集合を評価をした。各タスク集合とそのクエリに関して、「そのタスク集合をすべて実行

表 2 提案手法の評価

クエリ	達成度	純度	多様度
部屋を掃除する	17.5	0.35	3
ピリヤードで優勝する	8.25	0.52	3
家庭菜園を始める	14.5	0.40	5
保育園に入園させる	14.25	0.59	3
花粉症対策をする	16.75	0.41	2
犬を育てる	6.75	0.25	3
胃もたれを防止する	13.75	0.43	3
クレー射撃を体験する	4.75	0.14	2
平均	12.06	0.39	3.00

表 3 ベースライン手法の評価

クエリ	達成度	純度	多様度
部屋を掃除する	15.0	0.28	3
ピリヤードで優勝する	5.75	0.29	3
家庭菜園を始める	15.75	0.37	5
保育園に入園させる	11.75	0.34	2
花粉症対策をする	17.75	0.41	2
犬を育てる	10.5	0.22	2
胃もたれを防止する	14.75	0.48	3
クレー射撃を体験する	6.5	0.14	1
平均	12.22	0.32	2.63

することで、入力タスクを達成できる度合い」を達成度として評価した。達成度は 3 段階で評価されその基準は以下の通りである：

- 0 点: タスクをすべて遂行しても目的を全く達成しない。
- 1 点: タスク集合を遂行することで、目的をわずかでも達成する。
- 2 点: タスク集合を遂行することで、目的を 50%以上達成する。
- 3 点: タスク集合を遂行することで、目的を 80%以上達成する。

表 2 と表 3 における「達成度」は、出力された上位 10 件のタスク集合の達成度の累計値である。達成度を比較したとき、両者の値にはほぼ差はないが、提案手法の方がやや低い。「部屋 掃除する」や「ピリヤード 優勝する」では提案手法の達成度が優れているが、「保育園 入園させる」や「胃もたれ 防止する」ではベースライン手法が優れている。

5.3.2 純度の評価

純度及び多様度の評価のため、各入力タスクと SUBTYPE-OF 関係にあるタスクを列挙した。しかし、すべての SUBTYPE タスクを発見するのは困難であるため、実験により出力されたタスク集合をもとに SUBTYPE タスクを列挙した。たとえば、「部屋 掃除する」を入力タスクとしたときの SUBTYPE タスクは以下ようになった。

- 風呂を掃除する
- 引越の際に掃除する

(注 1): <http://datamarket.azure.com/dataset/bing/search>

(注 2): <http://mecab.googlecode.com/>

(注 3): <https://alaginrc.nict.go.jp/index.php>

(注 4): <http://alaginrc.nict.go.jp/hyponymy/>

(注 5): <https://alaginrc.nict.go.jp/resources/nictmastar/li-resource-info/li-resource-outline.html>

- 部屋を掃除する

純度の評価にあたり、我々はタスク集合に SUBTYPE のタスクでラベリングを行った。表 2 と表 3 における純度の点数は、各クエリの検索結果の上位 10 件の純度の平均値である。

2 つの手法の純度を比較すると、多くのクエリにおいて提案手法はベースライン手法を上回っている。特に「ピリヤード優勝する」が最も対照的であり、提案手法での平均純度は 0.52 で、ベースライン手法の平均純度は 0.29 である。この結果は、タスク集合から不要と思われるタスクを排除する手順で、タスク集合の純度を高めたことが原因と考えられる。「ピリヤード優勝する」ではノイズであるタスクが多かったことから、特にこの手順が大きな成果を上げたと思われる。

5.3.3 多様度の評価

多様度もまた純度同様に、SUBTYPE-PF の関係に着目して評価される。多様度の評価に用いた SUBTYPE タスクは純度の評価に使ったものと同じである。表 2 と表 3 に、その結果を示す。提案手法とベースライン手法をくらべると、多様度においてはほぼ変わりがない。これは、Web ページから発見できたタスクがそもそも同一であるがゆえに、頻繁に出現するとされるタスクの種類も同様のものになることが原因と考えられる。

実験結果をまとめると、提案手法はベースライン手法とくらべ、達成度と多様度はほぼ同じであるが、純度において優れている。

6. ま と め

本稿では、与えられた入力タスクに対して、実行することにより入力タスクを達成できるタスク集合を出力する検索であるタスク検索を提案した。検索するために、語の汎化と特化に着目してタスクの達成度を推定する手法とタスク集合を形成する方法、入力タスクに対し出力されるタスク集合族を多様化する方法を提案した。また、出力された結果を達成度・純度・多様度という評価尺度で評価し、提案する手法の有効性を検証した。

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。

文 献

- [1] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 335–336, 1998.
- [2] Ahmed Hassan and Ryan W White. Task tours: helping users tackle complex search tasks. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1885–1889. ACM, 2012.
- [3] Setsuji Tamugi. Service information search based on service eligibility and vocabulary. In *Master Thesis, Department of Social Informatics Graduate School of Informatics Kyoto University*, 2013.
- [4] Setsuji Tamugi, Yuhei Akahoshi, Koji Zettsu, Yutaka Kidawara, and Katsumi Tanaka. Prediction of search intent for web services and application (in Japanese). In *DEIM*

Forum 156, 2012.

- [5] Setsuji Tamugi, Koji Zettsu, Yutaka Kidawara, and Katsumi Tanaka. Service information search and ranking by eligibility (in Japanese). In *IPSJ SIG DBS*, 2012.
- [6] Takehiro Yamamoto, Tetsuya Sakai, Mayu Iwata, Chen Yu, Ji-Rong Wen, and Katsumi Tanaka. The wisdom of advertisers: mining subgoals via query clustering. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 505–514. ACM, 2012.
- [7] Takayuki Yumoto. Hierarchizing how-to information based on correspondence of procedure (in Japanese). In *DEIM2013*, 2013.