Master Thesis

Web Task Search Based on Task Generalization and Specialization

Supervisor Professor Katsumi Tanaka

Department of Social Informatics
Graduate School of Informatics
Kyoto University

Ryo KATO

Feburary 7, 2014

Web Task Search Based on Task Generalization and Specialization

Ryo KATO

Abstract

There are many methods that retrieve information from the web. Approach of web search varies from method to method, however, the most popularized search engines are based on document search model indexed with keywords. Therefore, web search as retrieving pure information not webpages are imperfect. Task search, pursuing the way to achieve the goal, is a great problem that should be solved.

People often use traditional search engines to know what to do something in order to accomplish a goal; For example, search with the query "treat hay fever" when you want to know how to treat hay fever. If you perform web search with "treat hay fever", you will discover webpages such as medicine makers' official sites, how-to sites and QA sites describing ways how to treat hay fever.

Although, these search results are not perfect solutions. Webpages are scored by the system based on hyperlink relations and terms' frequency, meanwhile, this ranking system ignores whether the user can accomplish the goal by the contents. Therefore, even if webpages are scored highly, contents in the webpages might be useless to achieve the goal. Moreover, they sometimes do not contain ways how to accomplish the goal. Since ordinary search engines are not based on tasks, users must change queries creatively and read many result pages.

Task search aims for solve this problem; If it is realized, its social value will be high. For example, you can evaluate a web search engine based on tasks. In this system, when a user wants to know the way to achieve the goal, this evaluation system add scores to webpage based on a special criteria. This system uses a rule that add a high score to pages when the webpage contains tasks that if user perform them then the goal will be accomplished. The system add rank to webpages based on task so that it make users possible to discover more useful webpages suitable for achieving the goal.

I propose the task search that retrieves tasks directly in this paper. I define

a task as “ An action that will achieve the whole goal or a part of the goal ”. To take an example, a query “Treat hay fever” has a goal “Become a state that has been treated hay fever”. If you want to become a state that has been treated hay fever, “Put on a mask”, “Put on glasses”, “Take a medicine of Allelock”, “Go to an nose and throat clinic” and “Keep pollen out of your home” are the tasks to be done. However, these tasks have relations of INSTANCE-OF, SUBTYPE-OF and PART-OF so that if you simply perform one task of them, you will not acomplish the goal perfectly. I focused attention on transitions of state that occurs when people perform tasks. Thereby, I classified relationships among tasks into three types.

Task search, which I propose in this paper, reads a task as input and outputs a list of ranked task sets. Inputs in the task search are tuples of a noun and a verb that represents the user's goal. Outputted tasks must meet the requirement that when a user perform the task then the whole goal or a part of the goal will accomplished. I propose three evaluation measures that checks how the outputted tasks are desirable: Completeness, purity and diversity.

I propose a query expansion method based on sponsored search. The purpose of this method is finding words suitable for task search. I made an experiment for the method and demonstrated that it is superior in terms of discovering tasks rather than conventional search. The results of the experiments indicate that the query expansion method is excellent in the variety of tasks and degree of detail.

Furthermore, I propose a novel method that searches tasks utilizing entailment relations and hyponym relations. I conducted an experiment that demonstrates the usefulness of these relations in the situation of comparing tasks. I implemented a task search system to make the experiment. Finally, I demonstrated that the task search method has an advantage for purity comparing with the baseline method.

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

内容梗概

Web上の情報を探索する手法は現在まで数多く発案され、そのアプローチも様々である。現在広く普及している検索エンジンはキーワードに基づいた文書検索モデルを出発点としている。そのため、Web上の文書ではなくWeb上の情報そのものを検索するという意味での情報検索はいまだに不完全である。「目的を達成するために何をすればいいか」を求める検索、つまりタスク検索も、解決すべき大きな問題である。

「目的を達成するために、何をすればいいか」を求めるために、現行の検索エンジンを使うことは一般的によく行われる行為である。たとえば「花粉症対策をする」ためには何をすればいいかを知るために「花粉症対策」で検索することがそれである。実際に「花粉症対策」で検索すると、医薬品メーカーの公式情報サイトやHow-toサイトの花粉症対策特集ページ、QAサイト、厚生労働省のQA集などが見つかる。

しかしそれらの検索結果は完全な答えではない。検索エンジンが提供するページは主にハイパーリンク関係や語の頻度を基準として計算されたスコアによりランクづけされているが、このランキングは、「ページ内に目的を達成する行動が書かれているか」という視点に立ってはいない。そのため、検索エンジンに高くランクづけされたページであっても、目的を達成するために有効とは限らない。また、目的を達成するためにどんな行動をとればよいか、Web検索の結果から探そうとしても、見つからないことがある。検索結果ページに目的を達成する方法が書いてあるとは限らないためである。タスクに基づいた検索ではないため、ユーザーはクエリを工夫したり、複数の検索結果ページを読み込んで思考することを迫られる。

タスク検索は、このような問題を解決するものであり、もし実現すればその社会的意義は大きなものとなる。たとえば、Webページのランキングに、タスクの観点に基づいた尺度を使うことが考えられる。ユーザーが「目的を達成するために、何をすればいいか」を求める際に、現行のWeb検索エンジンが用いる指標ではなく「Webページ内に書かれた行動を遂行することがどのような結果を生み、どのように目的を達成するか」といった指標でスコアをつけ、ラン

クづけをすることで、より目的達成に役立つ Web 検索が可能となる。

本研究において私は直接タスクを検索するタスク検索を提案する。私はタスクを「遂行することで、目的の一部または全部を達成する行動」と定義した。「花粉症対策をする」を例とすると、目的は「花粉症対策をした状態になること」である。花粉症対策をした状態になるには、「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」「自宅の花粉を取り除く」といったタスクを遂行すればよいが、これらのタスクには INSTANCE-OF や SUBTYPE-OF、PART-OF といった関係があり、単純に、どれかひとつを遂行するだけで目的が完全に達成できるわけではない。タスク検索を実行するにあたり、望ましいタスク集合を探すために、私はタスクの遂行による状態変化に着目し、タスク同士の関係を分類した。

私が提案するタスク検索は、タスクを入力すると、ランクづけされた、タスク集合のリストを出力するものである。タスク検索における入力は、目的を表現した主語と目的語の組である。出力されるタスクは、入力される目的の状態に対して、「タスクを遂行することで、目的の一部または全部を達成する」という要件を満たす必要がある。このタスクが望ましい答えかどうか、評価するため、Completeness, purity, diversity という尺度を導入する。

タスク検索に適したクエリを自動的に発見する手法として、私は広告連動型検索にもとづいたクエリ拡張を提案する。この手法に関して評価実験を行い、従来の検索クエリによりタスクを検索する場合よりも、発見できるタスクの種類と詳細さにおいて優れていることを示した。

また、語のエンテイルメント関係と上位下位関係を利用してタスクを直接する検索する手法を提案する。これらの関係がタスク同士を比較する際に有用であることを示し、Web 上よりタスクを検索するシステムを実装し、実験を行った。その結果、このタスク検索手法が purity において比較手法よりも優れていることを示した。

Web Task Search Based on Task Generalization and Specialization

Contents

Chapter 1 Introduction

1.1 Background

Researchers and developers have created a lot of web search methods in these decades[9]. Their approach varies from method to method, however, the most popular search engine depends on the model of document search. This model is based on the frequency of keywords appeared in the documents. Therefor, it is not perfect when you want to search information itself, not documents on the web. Task search, which looks for the way to achieve the goal, is an important problem that should be solved.

People frequently use an ordinary search engine to find ways to accomplish the goal. For example, when people want to know ways to treat hay fever, they sometimes search with a query such as "hay fever treatment". If you actually search with "hay fever treatment", you will discover websites of medical product maker, how-to sites, QA sites and Ministry of Health, Labor and Welfare's one.

However, These results are not an ideal solution. In these systems, webpages are ranked with the score based on their hyperlink relationships and term frequency. Search engines do not pay attention to the contents themselves, so they are careless about the usefulness of the information in the result pages. When it comes to task search, good result pages should contain information that helps users to achieve the goal. Consequently, if you search with ordinary search engine and obtain high-ranked webpages, they might contain useful information to achieve your goal. Furthermore, when you try to find the ways from the result webpages to achieve the goal, you might not find them out. This is because sometimes result pages might not contain the means to achieve the goal. Ordinary search engines do not search based on task, so when you want to find a way to achieve your goal, you must change the query effectively and read the result pages deeply.

Task search, which I am going to propose, will solve this problem. If this concept is realized, it will greatly contribute to information society. For instance, search engines can adopt a scoring system based on the viewpoint of tasks. This scoring system adds scores to webpages based on the measure that

evaluates the usefulness of the contents in them. It evaluates the degree of achievement that "If a user actually does the action written in the webpage, what result will occurs and how it works?" Such systems can rank webpages in a better way to find means to achieve the user's goal.

In this research, I suggest task search, which is finding tasks directly. I define a task as "An action that will achieve the whole goal or a part of the goal". For example, when you regard "treat for hay fever" as a task, its goal is "to be a state that the treatment for hay fever is done". If you want to be the state that the treatment for hay fever is done, you would "put on a mask", "put on glasses", "take an anti-allergy medicine", "go to a clinic" and "remove pollen in your house".

Task search, which I am going to suggest, is a web search that accepts a task as user's input and outputs task set. I am going to describe the concrete expressions of tasks later. Output tasks must fulfill the requirements that "When a person perform the task, a part of his/her goal or the whole goal must be achieved". I adopted the measures of usefulness, concreteness, and coverage in order to evaluate the tasks.

1.2 Difficult point

Tasks have complicated relational structure. There are several relationships such as sequence and entailment among tasks. Therefore, It is difficult to rank tasks in a straightforward method. The most difficult problem is how to add ranks to tasks based on the relationships among them.

Here I explain the reason why ranking based on the relational structure of tasks. For instance, the task "treat hay fever" has a relational structure. If you want to treat fay fever, because hay fever is an allergic reaction, you can "take an anti-allergic medicine" and achieve the goal of "treat hay fever." However, if you just do the action of "go to an nose and throat clinic", you will not achieve the whole goal. If you want to achieve the whole goal, you must go to a nose and throat clinic, obtain a prescription of anti-allergy medicine, buy the medicine in a pharmacy and take it with water. On the other hand, if you have achieved the goal of "treat cedar fever", it does not mean that you achieved the goal of "treat

hay fever". This is because other several plants, such as ragweed and hinoki cypress, also produce pollen and cause hay fever. In addition to that, you can achieve the whole goal by performing multiple tasks although they each would achieve only a limited part of the goal. For example, when you perform the task "treat cedar fever", you achieved only a small part of the goal, however, if you also performed "treat hinoki cypress fever", "treat ragweed fever" and all other tasks to treat plants' fever, it means that you achieved the whole goal of "treat hay fever".

As I explained above, since tasks have complicated relational structure, it is difficult to evaluate tasks in the viewpoint of task-relationships. Therefore, I had to adopt some tricks in order to construct the model of tasks. I will describe them in Chapter 3.

The rest of the paper is organized as follows. Chapter 2 presents related works. I am going to describe modeling of tasks in Chapter 3. The method using sponsored search is presented in Chapter 4. After this, I will explain the method that finds tasks based on generalization and specialization in Chapter 5.Moreover, I am going to describe the experiment that demonstrates the advantages of the task search method in Chapter 6. Finally, I conclude the paper in Chapter 7.

Chapter 2 Related Work

There are some researches related to task search published recently. For example, Yamamoto et al [19] suggested the concept of search goal and subgoal. They proposed a method to finds subgoals fit for the query by using sponsored search advertisements. According to their paper, the definitions of search goal and subgoal are "A search goal is an action that the searcher wants to achieve, often represented by a verb plus possibly a noun phrase" and "A search goal x Is a subgoal of another search goal y if achieving x helps the searcher to also achieve y either wholly or partially. Task search, which I am going to suggest, is close to Yamamoto's search goal and subgoal in a sense that tackle the hierarchical relationships among tasks. However, my approach has different viewpoint that suppose tasks as not only actions but also transitions of user's state. I am going to describe my task model in Chapter 3.

Many researchers pay attention to sponsored search, however, most of them focus on other topics such as improvements of search precision [1], predicting click through rate [11] and analysis on user's behavior around sponsored search [7, 3].

Hassan et al [6] proposed a method, which helps users to achieve tasks, aiming for the same purpose to my task search. Their method obtains the category of the webpage from Open Directory Project (OPD) and regards it as a task. It extracts tasks which are related each other from the search log. They find task sets related to the query automatically. In addition to that, Yumoto et al [20] paid attention to procedure information. They supposed a method to extract procedure information from how-to documents and construct it into a hierarchical model.

Widyantoro and Yen [17] proposed a search engine for semantic web search based on ontology. Zhang et al [21] tackled service search. Finin et al [4] also researched on semantic web and proposed a method that extracts metadata and rank webpages with ontology-based scores. Jones and Klinkner proposed concepts of *search mission* and *search goal* [8]. They try to classify search session by a method based on search mission.

There is another related research based on user's purpose. Tamugi et al [16] [15] [14] suggested service search aims for find webpages that offer a service such as furniture shopping and reservation for cakes. Their method converts queries and make the search better to find services, which the user wants to use in order to achieve the purpose. For example, when a user, who wants to sell the guitar, searches a service that purchases used guitars, the query "used guitar purchase" is better than "used guitar sell". This is a query conversion from a query that is based on the viewpoint of service receiver to a query that is based on the viewpoint of service provider. Tamugi et al proposed six rules described before used for query conversion such as from "sell" to "purchase".

1. **Convertion to provider-consumer word:** A pattern to convert the verb in the search query words that the user express an intention to implement the service into the words from the viewpoint of the providers. This is one type of the antonym. (e.g. from "sell guitar" to "purchase guitar")
2. **Conversion to synonym:** A pattern to convert the verb in the search query words that the user expresses an intention to implement the service into the synonym to express an intention more exactly on service vocabulary. (e.g. from "send package" to "deliver package")
3. **Conversion to entailment word:** A pattern to convert the verb in the search query words that the user expresses an intention to implement the service into the other verb entailed by the original verb. (e.g. from "rent an apartment" to "find an apartment")
4. **Conversion to sibling word:** A pattern to convert the verb in the search query words that the user expresses an intention to implement the service into the other verb that entails the original verb too. (e.g. from "throw away furniture" to "purchase furniture")
5. **Addition to service value word:** A pattern to add the word describing the characteristic of service to the search query that the user inputs. (e.g. from "sell brand items" to "sell famous brand items")
6. **Conversion to service provider word:** A pattern to convert verb in the search query words that the user expresses an intention to implement the service into the noun that express something the service provider offers.

(e.g. from “leave a cat” to “cat hotel”)

Their method utilizes sponsored search. It extracts words from the input query and sponsored search information, which is related to the query following the six rules above. After that, it uses the words to query expansion and search webpages offer services. According to their paper, the experiments done by them show the improvement of recall and precision.

The most similar approach to my task search is Tamugi et al’s one. Both my method and their method use sponsored search, therefore, their insights helped me a lot. However, the task search, which I am going to describe, is a little different from their purpose. Therefore, even though method’s procedure is similar, task search must be analyzed from the different point.

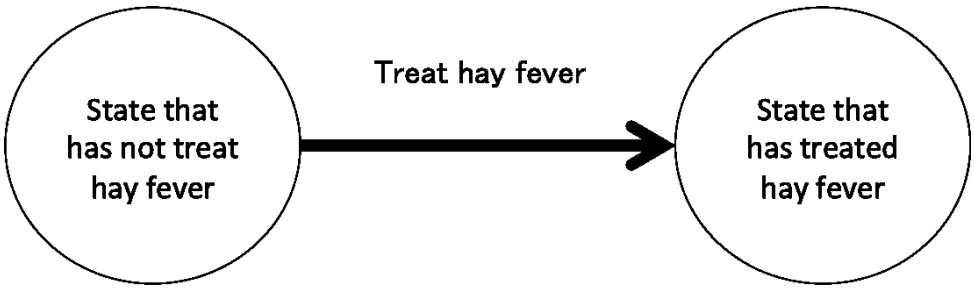

Figure 1: States and its action for "treat pollen allergy."

Chapter 3 Task Model

In this chapter I am going to describe the concepts of terms appear in task search and their model.

3.1 Modeling based on state

In this research I construct the model of tasks by using the concepts of action and state. Figure 1 shows the model of the task that "treat hay fever". Before you have performed the task, you are in "the state before you have treated hay fever", which is in the left circle. If you have performed the task, you are in the right circle that is "the state already have treated hay fever". This figure suggests that the action "treat hay fever" makes the transition. To sum up, a task is an action that causes a transition of state.

As has been noted in Section 1.1, I defined a task as "An action that will achieve the whole goal or a part of the goal". In the definition, the goal means a state, therefor, task is an action that make a transition from a state that have not achieved the goal to a state that already achieved the goal. As I described above, tasks can be dealt with states and actions that make transitions.

Figure 2 shows three tasks that related to the task "treat hay fever". I am going to describe relationships among tasks. There are three relationships among tasks: **INSTANCE-OF relationship**, **SUBTYPE-OF relationship**, **PART-OF relationship**.

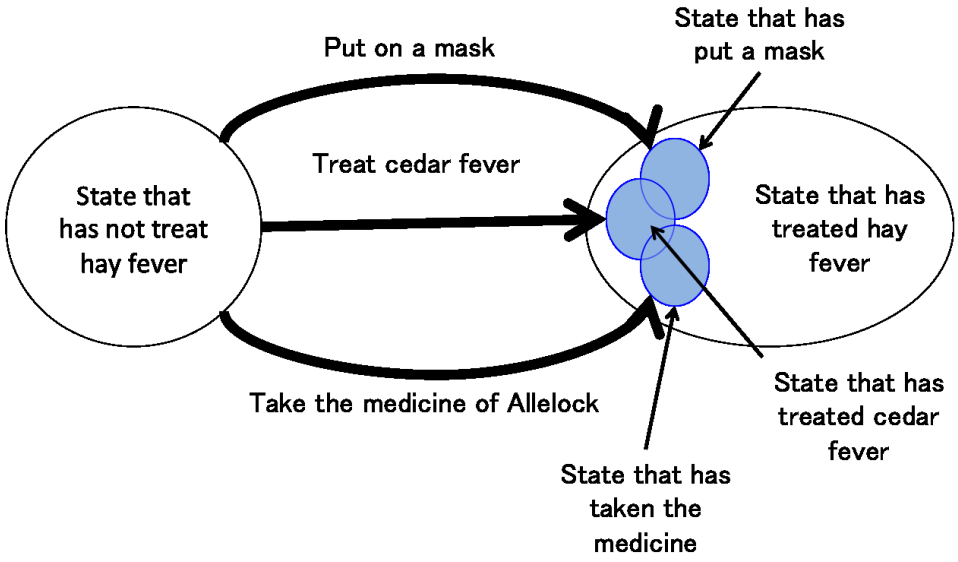

Figure 2: Example actions for achieving "treat pollen allergy."

3.2 Relationships among Tasks

3.2.1 INSTANCE-OF relationship

When the task t_1 is in the INSTANCE-OF relationship to the task t_0, it means that "If you perform t_1, the goal, which would be achieved when t_0 is performed, will be wholly achieved"; for example, when you have performed the action "put on a mask" shown in Figure 2, your state would be "the state that is already treated hay fever". Therefor, the task "put on a mask" in a INSTANCE-OF relationship to the task "Treat hay fever". In addition, in the real world, it might not true that if you put on a mask then you have treated hay fever. However, in this paper, I regard the task "Put on a mask" would be a perfect treatment for hay fever in order to simplify the problem.

INSTANCE-OF relationship is a kind of IS-A relationship; Namely, when the task t_1 is in the relationship of the task t_0, t_1 is a task that is specialized from t_0. I consider a specialization in INSTANCE-OF relationship to be an instantiation.

3.2.2 SUBTYPE-OF relationship

When the task t_1 is in the SUBTYPE-OF relationship to the task t_0, it means that "If you perform t_1, the goal, which would be achieved when t_0 is performed, will be partly achieved"; for example, the task shown in Figure 2 achieves only

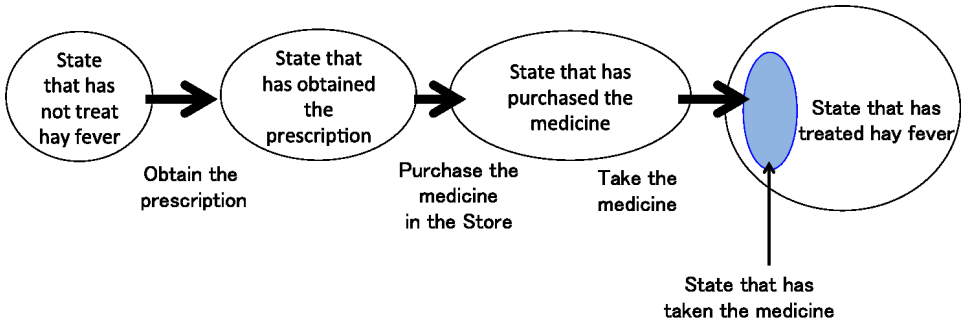

Figure 3: Example actions for achieving "treat pollen allergy."

a part of the goal to have treat hay fever. As I mentioned in Section 1.2, there are other plants cause hay fever. Therefor, the task "treat cedar fever" is in a SUBTYPE-OF relationship to the task "treat hay fever".

SUBTYPE-OF relationship is also a kind of IS-A relationship; in other words, when the task t_1 is in SUBTYPE-OF relationship to the task t_0, t_1 is a task that is specialized from t_0.

3.2.3 PART-OF relationship

The task "Take the medicine of Allolock" shown in Figure 2 achieves the whole goal alone. Therefor, this task is in INSTANCE-OF relationship to the task "treat hay fever". However, there is another aspect in it if you consider other tasks before you take the medicine.

When the task t_1 is in the PART-OF relationship to the task t_0, it means that "The t_1 is consisting of the task t_0, which is an sequential task"; for example, the task "obtain a prescription of an anti-allergy medicine" shown in Figure 2 is in PART-OF relation to the task "Treat hay fever". Figure 3 describes a procedure that a person takes the medicine of Allolock and eventually achieves the whole goal that treats hay fever. If you want to take the medicine of Allelock, you must obtain a prescription of the anti-allergy medicine, hand it to a pharmacy, purchase the medicine of Allelock and finally take it with water. The action that "Obtain the prescription" and "Purchase the medicine of Allelock" can be regarded as a part of the sequential procedure. If you perform the single action, which "Obtain the prescription", it does not means that you have treated hay fever, nevertheless, this action is necessary to perform the sequential procedure. PART-OF relationships appear in sequential multiple tasks.

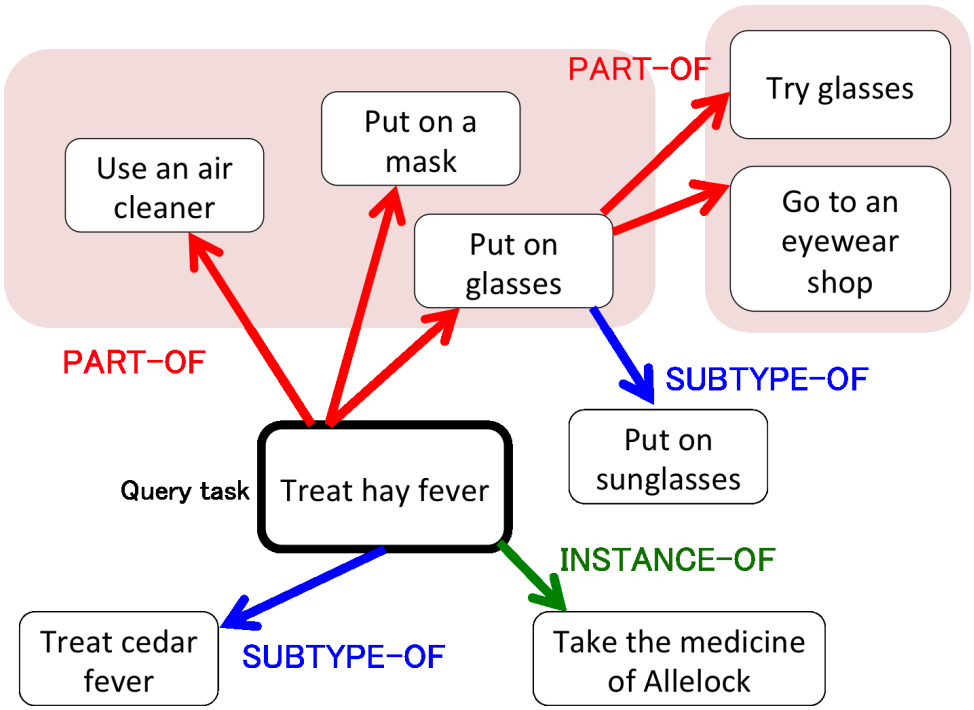

Figure 4: Task graph represents the relations among the query "treat hay fever".

In the example above, when you perform the action "Obtain the prescription" your state has been changed so that you have already obtained the prescription. The state, which you have obtained the prescription, is a mid-state between the state of the first phase and the final phase that you have treated hay fever.

PART-OF relationship is also a kind of IS-A relationship; namely, when the task t_1 is in PART-OF relationship to the task t_0, t_1 is a task that is specialized from t_0.

3.2.4 Task Graph

I made a model of Tasks as a directed acyclic graph, which have a single task node that means the query task. Figure 4 represents a part of structure of a task search. Each nodes stands for tasks and every edge is a relationship between tasks. The central node is a task that is given as the query. When you want to treat hay fever, you can choose any way to do that, however, every measure you can select is related to the query task.

A task graph is composed of different types of edges. Five edges connect the query task "treat hay fever" to task nodes "Treat cedar fever", "Take the

medicine of Allelock", "Use an air cleaner", "Put on a mask" and "Put on glasses". These task nodes can be classified into three types: SUBTYPE-OF, INSTANCE-OF and PART-OF. For example, the edge links from the query node to "treat cedar fever" is a SUBTYPE-OF edge, therefore it means that "treat cedar fever" is a subtype task of "treat hay fever". The edge connects "treat hay fever" and "Take the medicine of Allelock" is a INSTANCE-OF edge, so when you perform the task "Take the medicine of Allelock", it make you achieve the whole goal. PART-OF edges have a special feature like no others; they make groups in the hierarchy. For example, the three tasks "Use an air cleaner", "Put on a mask" and "Put on glasses" are in the same group.

SUBTYPE-OF, INSTANCE-OF and PART-OF are the all relations possible when you choose a way, so what the task search system should do is find tasks by their relationships and evaluate them in certain evaluation measures.

3.3 Evaluation Measures

In order to evaluate task search, I adopted three evaluation measures: Completeness, purity and diversity.

3.3.1 Completeness

Completeness indicates the degree of achievement for the goal when the task is actually done; for instance, when a user input the task "Treat hay fever" and obtains the task "Attend a hospital", its usefulness is higher than the usefulness of the task "Finish a marathon". You can adopt completeness to a single task, a set of tasks and a list of sets of tasks. The example I described above is a completeness about a single task. When you evaluate a set of tasks by completeness, you must think about a procedure of actions when you achieve the goal.

Completeness strongly depends on the tasks that have a relationship of SUBTYOE-OF; For example, there are completely different ways to achieve goals such as "Brew iced coffee" an "Brew hot coffee" in a case when the query "Brew a cup of coffee". Figure 5 represents this example. "Brew iced coffee" an "Brew hot coffee" are connected to "Brew a cup of coffee" in SUBTYPE-OF. When you try to brew iced coffee, you must chill the coffee by some means. Off

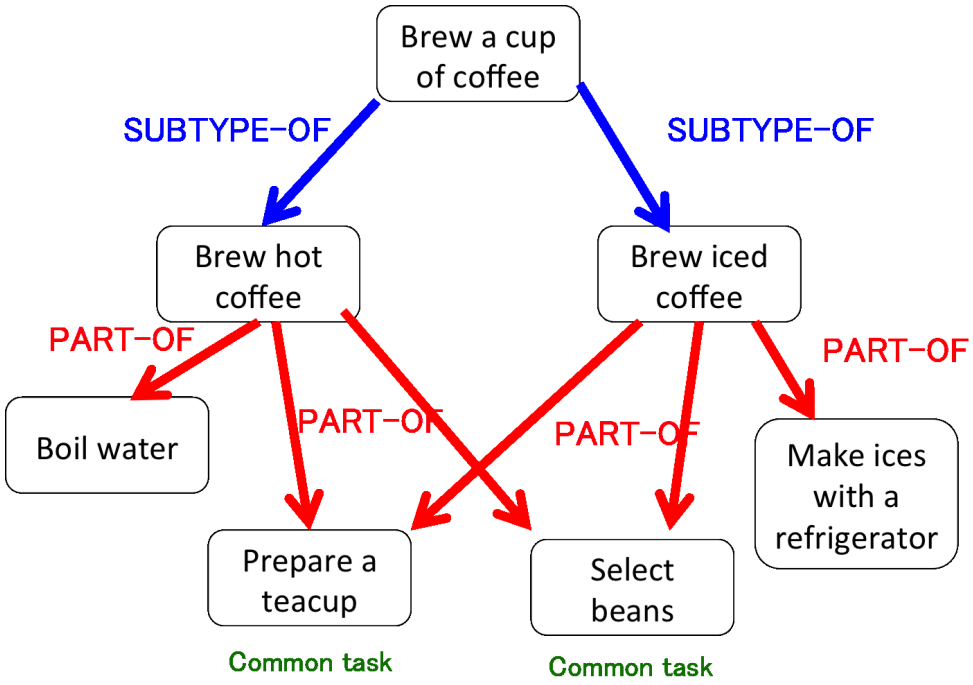

Figure 5: Evaluation with SUBTYPE-OF and PART-OF Relations

course, when you brew a cup of hot coffee, you need not chill it with ices. When the query "Brew a cup of coffee" is given, this query does not choose one of them, therefore, you can select one of them freely. The task search's result may be {"Make ices with a refrigerator", "Buy instant coffee", "Prepare a teacup"} or {"Boil water", "Select beans", "Pour hot water"}, however, they must not appear in the same time. When the task "Brew a cup of coffee" is given, usually it does not mean "Brew a cup of hot coffee and a cup of iced coffee", therefore, task search result must choose one of them.

When a set of tasks in results contains both tasks related to different SUBTYPE-OF tasks, I evaluate completeness by tasks that is classified in the only one tasks. For example, when a query "Brew a cup of coffee" is given and task search result is {"Make ices with a regrigerator", "Prepare a teacup", "Boil water", "Select beans", "Pour hot water"}, its completeness is the same to the completeness of {"Prepare a teacup", "Boil water", "Select beans", "Pour hot water"}. This is because tasks peculiar to "Brew iced coffee" are {"Make ices with a regrigerator"} only and tasks peculiar to "Brew hot coffee" are {"Boil water", "Select beans", "Pour hot water"}, therefore the latter SUBTYPE-OF

task contains more tasks in the results.

3.3.2 Purity

Purity indicates the consistence of a set of tasks. Good task search results should be consistent; therefore I suggest the evaluation measure purity is suitable for task search. Purity is frequently used when you evaluate a set of documents. It is calculated by the number of contains and the number of consistent documents. In this research, I adopt purity to tasks; hence I regard tasks as documents. For instance, when you input a query "Brew a cup of coffee" and obtain a result with {"Make ices with a refrigerator", "Prepare a teacup", "Boil water", "Select beans", "Pour hot water"}, its purity is lower than {"Prepare a teacup", "Boil water", "Select beans", "Pour hot water"}. The only difference between them is the existence of "Make ices with a refrigerator". This task is classified as a PART-OF task of a SUBTYPE-OF task "Brew iced coffee" and the other tasks are classified as a PART-OF task of a SUBTYPE-OF task "Brew hot coffee" or common tasks of "Brew iced coffee" and "Brew hot coffee". When you evaluate task sets by purity, you must select tasks by its belonging of SUBTYPE-OF task.

I describe how to evaluate a set of tasks by purity. Consider that you obtain a set of tasks $T = \{t_{i1}, t_{i2}, \dots, t_{h1}, t_{h2}, \dots\}$. It contains two types of tasks, which classified as PART-OF a SUBTYPE-OF task. One type tasks $T_i = \{t_{i1}, t_{i2}, \dots\}$ are considered to be PART-OF a SUBTYPE-OF task s_i. The other type tasks $T_h = \{t_{h1}, t_{h2}, \dots\}$ are considered to be PART-OF a SUBTYPE-OF task h_i. To sum up, task set T composed of these two task sets. Purity of a task set s $Purity\,(T)$ can be evaluated by this equation below.

$$Purity\,(T) = \frac{\max\{|T_j|, j = i, h\}}{|T|}.$$

When you evaluate purity, you must select the most dominant SUBTYPE-OF tasks in it; in the case above, {"Prepare a teacup", "Boil water", "Select beans", "Pour hot water"} are selected. "Prepare a teacup" and "Select beans" are common tasks shared by "Brew iced coffee" and "Brew hot coffee", hence you can select them anyway.

3.3.3 Diversity

Diversity indicates how a list of sets of tasks contains various actions. Diversity is also based on SUBTYPE-OF relations.

I will describe the procedure how to evaluate Completeness, Purity and Diversity actually in Chapter 6.

Chapter 4 Query expansion using sponsored search

In this chapter, I propose a new method to find webpages that include many tasks. This is the first step to realize the task search I mentioned in Section 1.1 and the method is used in the task search in Chapter 6.

The novel method uses sponsored search to expand the query based on the verbs related to the tasks. This method can find web pages, which are including tasks that are impossible to find from the web search by original inputted query.

I am going to describe the idea of the method in Section 4.1 and then explain detailed procedure in Section 4.2. After that, I will describe the experiment performed to demonstrate the advantages of my method in Section 4.4. Finally, I will discuss the character of method in Section 4.5.

4.1 Idea of the Method

Finding tasks is a difficult problem for conventional web search engine users as Hassan et al described [6]. Search engines, such as Google and Bing, mainly are based on indexes with terms. Thereby, they are useful when you perform informational search or navigational search, however, not useful enough when you do task search as I described in Chapter 1.

In contrast with conventional web search, sponsored search has a special advantage. It provides commercial information connected with user's query. Advertising clients try to buy terms related to their products. Therefore, texts on advertisements are strongly related to the query and tasks. Sponsored search is useful to guess user's intent, hence some researchers focused on this advantage and tried to classify information with sponsored search [19].

The main idea of my method focuses on terms frequently appearing in advertisements; For example, when you search with "部屋の掃除をする (Clean a room)" in a sponsored search, you will obtain several advertisements such as "お部屋の清掃はIRヘルパーズ (Ask IR Helpers to clean your room)" and "シェルの炭化水素系洗浄剤 (Shell's hydrocarbon detergent)". Figure 6 represents the sponsored search results. These advertisements inform us of products related to the query "部屋の掃除をする (Clean a room)". Frequently appearing

Figure 6: Sponsored Search by Yahoo Japan

terms such as "洗剤 (detergent)" and "クリーニング (Ask professionals to clean a room)" are good words for query expanding.

Advertising clients aim for selling products so that they try to make user pay attention to the advertisements. Therefore, texts appearing in advertisements contain useful words that user wants to find. When you do task search, such words are often good words for query expanding.

4.2 Procedure of the method

My query expanding method can be divided into four steps. First, do sponsored search with q given query. After that, expand query with the texts in the advertisements. Then, do web search with the expanded queries. Finally, combine the results pages.

Figure 7 represents the summery of the novel method. This method fetches web pages by the following detailed procedure.

1. Obtain the query q which represents the task that means the state the user want to achieve.
2. Execute sponsored search by q and obtain n advertisements $\{a_1 \dots a_n\}$
3. Do text segmentation to the titles and snippets of $\{a_1 \dots a_n\}$ and extract verbs and sahen nouns. After that, calculate the number of occurrences. Define W as the set of given words. Define tf_w as the number of occur of

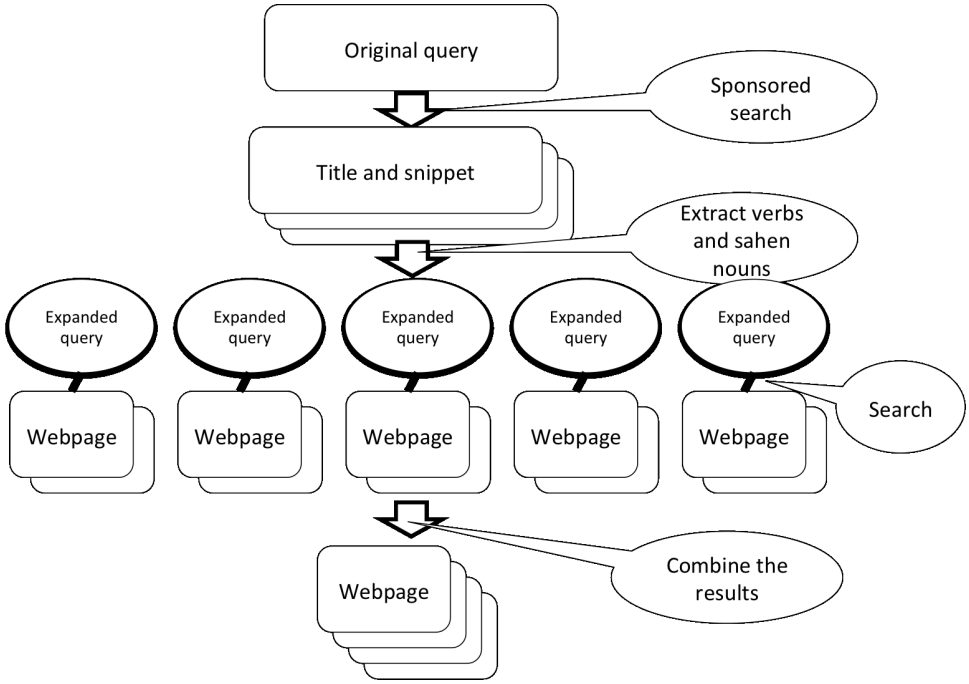

Figure 7: Example actions for achieving "treat pollen allergy."

the word $w \in W$.

4. Gain m words $\{w'_1 \dots w'_m\}$ which have higher in tf_w.
5. Let the webpages set $D = \Phi$. Repeat the following steps below until $|D| = k$ to each words w_i $(1 \leq i \leq m)$ given in step 4.
   (a) Add the gained word to the original query so that make query $q_i = q \wedge w_i$.
   (b) Execute web search with the query q_i and obtain k' webpages $\{d^i_1 \dots d^i_{k'}\}$.
   (c) Add gained webpages to D sequentially from d^i_1 $\left(D \leftarrow D \cup \{d^{k'}_1\}\right)$. When $|D| = k$ then go to step 6.
6. Return the webpage set D.

4.3 Experiment

In this section, I am going to describe how I evaluated the results of the experiment I described in Section 4.2. Since the number of tasks and tasks' degree of detail are important, I propose two barometers to evaluate the results. I compared the top k web pages from the results of the query expansion method and the baseline method.

4.3.1 Baseline procedure and experiment setting

I used another method as a baseline in order to verify the utility of the query expansion method. The baseline method does not use sponsored search. It collects web pages using the results of a conventional web search to expand queries.

This baseline method uses n pages from normal web search results. $\{a_1 \dots a_n\}$ are given from normal search, not from sponsored search. I used Google Custom Search API to obtain the results of web search. I also used MeCab [1)] to morphological analysis. I set the parameters of the method that $n = 15$, $k = 20$ and $k' = 5$. It means that the system expands the query and fetches 20 webpages. I evaluated the number of the tasks included in the 20 web pages.

4.3.2 Evaluation

I manually evaluated the novel method by following two evaluation measures.

- Action count
- Degree of detail

Action count represents the number of the actions that can be found in the web pages from the top 20 ranking. These actions must not be included in the web pages in the results of the original query.

Degree of detail indicates the number represents how detail the description of the action is. When a newly found action, which is searched by the query expansion, is detailed enough then I added 2 points. Otherwise, I added 1 point. For example, when the query is "花粉症対策をする (Treat hay fever)" and the result webpage includes the text "Use a humidifier" only, then I add 1 point to the webpage. If the webpage includes information about the way to purchase a humidifier or how to recognize good humidifiers from bad ones, 2 points will be added.

I evaluated the webpages by the two measures because the number of tasks and their degree of details are important to accomplish goals. When people perform a task after reading documents about the task, how they achieve the goal depends on the degree of detail. Moreover, if they did not find tasks on

[1)] http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html

Table 1: Converted queries

		Proposed method	Baseline
Queries	部屋の掃除をする (Clean a room)	部屋 掃除 する クリーニング 部屋 掃除 する 配送 部屋 掃除 する 清掃 部屋 掃除 する 洗浄 部屋 掃除 する 炭化	部屋 掃除 する 維持 部屋 掃除 する しよ 部屋 掃除 する 思う 部屋 掃除 する やっ 部屋 掃除 する 上がる
	ダイエットする (Go on a diet)	ダイエット する 痩せ ダイエット する 吸引 ダイエット する 通信 ダイエット する 保証 ダイエット する 体験	ダイエット する やめ ダイエット する 〜 ダイエット する 食べる ダイエット する やせる ダイエット する 測る
	結婚する (Get married)	結婚 する 紹介 結婚 する 離婚 結婚 する し 結婚 する 提携 結婚 する 安心	結婚 する し 結婚 する ? 結婚 する 祈 持つ 結婚 する 開業
	コーヒーを淹れる (Brew coffee)	コーヒー 淹 れる 通信 コーヒー 淹 れる 取れる コーヒー 淹 れる 学べ コーヒー 淹 れる 提携 コーヒー 淹 れる 紹介	コーヒー 淹 れる いれる コーヒー 淹 れる れ コーヒー 淹 れる 抽出 コーヒー 淹 れる し コーヒー 淹 れる ドリップ
	花粉症対策をする (Treat hay fever)	花粉 症 対策 する 改善 花粉 症 対策 する 治療 花粉 症 対策 する 清浄 花粉 症 対策 する 配送 花粉 症 対策 する 飲む	花粉 症 対策 する し 花粉 症 対策 する い 花粉 症 対策 する 抑制 花粉 症 対策 する 紹介 花粉 症 対策 する 予防

the web, they might not try to accomplish the goal. Thus, Action count and degree of details are suitable evaluation measures for the query expansion for task search.

4.4 Result

Figure 8 and Table 2 represents the result. First, my method's score is better than the base line for the queries "部屋を掃除する (Clean up the room)", "結婚する (Get married)" and "コーヒーを淹れる (Brew coffee)". Moreover, our novel method has higher score in the degree of detail for 4 queries; "花粉症対策をする (Treat hay fever)", "部屋を掃除する (Clean a room)", "結婚する (Get married)" and "コーヒーを淹れる (Brew coffee)". These results implies

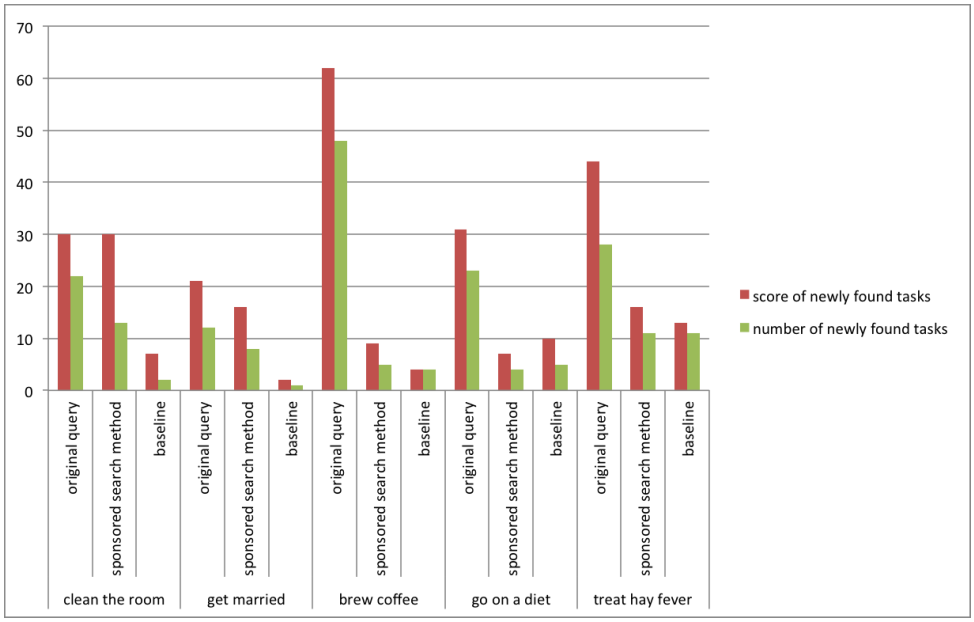

Figure 8: Example actions for achieving "treat pollen allergy."

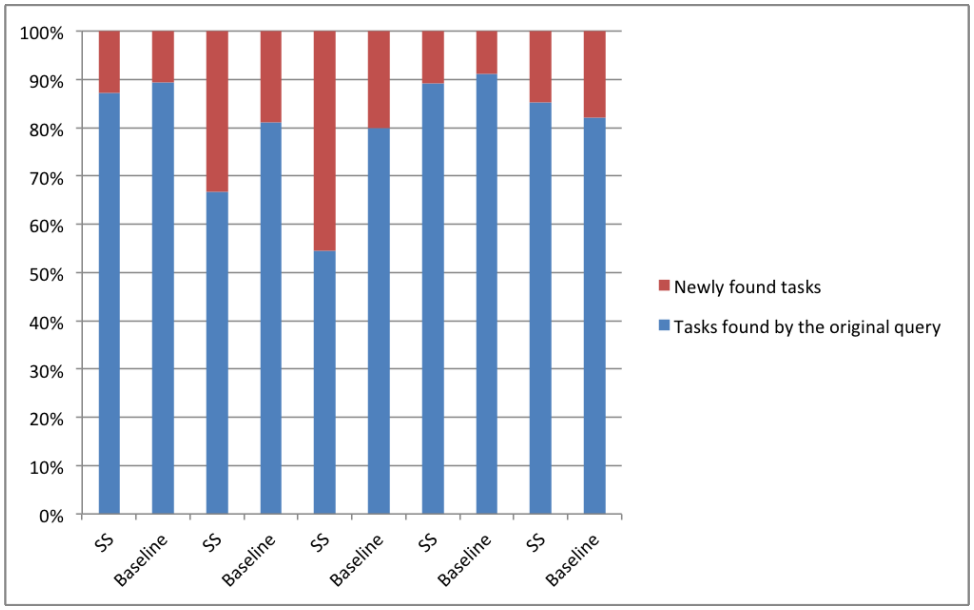

Figure 9: Example actions for achieving "treat pollen allergy."

that the web search based on sponsored search is specialized for task search, therefore it can find more web pages including information needed to achieve the task.

I analyzed the expanded queries. When we input "部屋を掃除する (Clean a room)", our method found words such as "クリーニング (Cleaning)", "配送

Table 2: Number of obtained actions and their degree of detail by query expansion technique.

		Proposed method		Baseline	
		Action count	Degree of detail	Action count	Degree of detail
Query	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

(Delivery)" and "洗浄 (Wash)". Meanwhile, the base line found words such as "上がる (Up)" and "維持 (Maintain)". These results imply that the novel method successfully gained the words related to the tasks as query expansion words using sponsored search. Furthermore, when we input the query "部屋の掃除をする (Clean a room)", our novel method found the web pages including the task "Ask a professional cleaner company to clean the room". The words gained by sponsored search such as "クリーニング (Cleaning)" and "洗浄 (Wash)" make it possible to discover the web pages related to the task.

On the other hand, when the query "ダイエットする (Go on a diet)" was given, our method results in lower score than the baseline in action count and degree of detail both. Proposed method gained words from sponsored search such as "吸引 (liposuction)" and "体験 (experience)". When we search with the expanded query by these words, we found web pages related to liposuction. However, the baseline gained the word "食べる (eat)" as expanded query. The baseline searched with the expanded query with "eat", therefore it discovered many web pages that describe how cook a food good for people in diet. Thus, the baseline found variety of web pages related to "Go on a diet" and its score was higher than the proposed method.

Table 3 represents the classification of the search results based on SUBTYPE tasks. In the search results, there were only a few SUBTYPE-OF tasks comparing with other tasks. This is because SUBTYPE-OF tasks are few in these queries.

Table 3: Classification result of obtained actions in terms of subtype-of and instance-of relationships.

		Proposed method		Baseline	
		SUBTYPE-OF	not SUBTYPE-OF	SUBTYPE-OF	not SUBTYPE-OF
Query	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

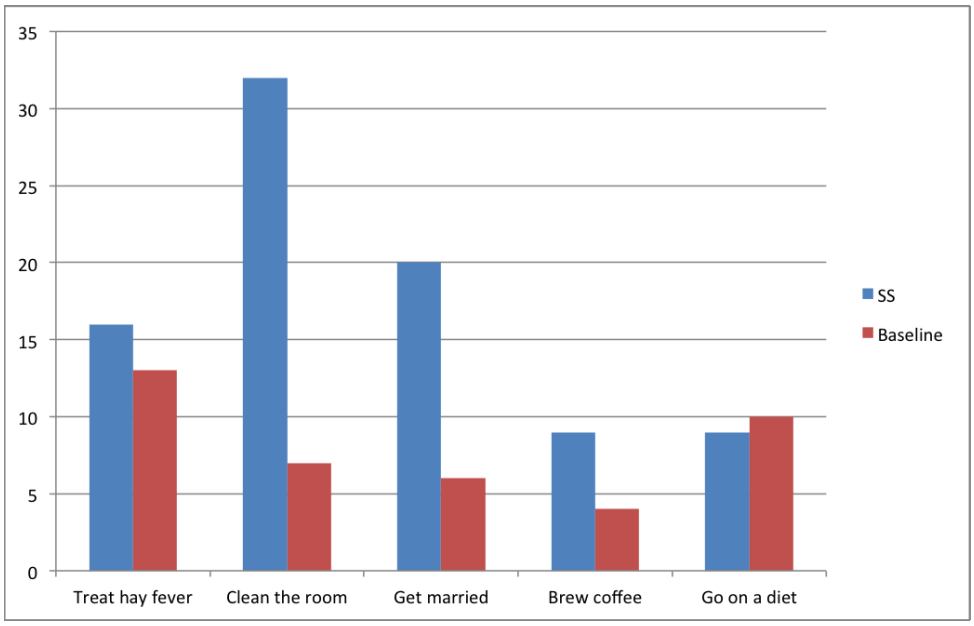

Figure 10: Example actions for achieving "Treat hay fever"

4.5 Discussion

I define task search in this paper that "finding a collection of actions that would achieve the goals composed of several actions that make state transition". When you input a string query, it explicitly of implicitly implies the goal statement. For example, when the input query "花粉症 対策する (Treat hay fever)" is given, the output would be the actions such as "Put on a mask", "Take a medicine" and "Get the medical prescription". The experiment indicates that the query expansion is useful to task search.

The experiment I described this chapter demonstrates that the method

based on sponsored search is excellent in terms of finding tasks. I am going to describe the whole task search method including the query expansion method.

Chapter 5 Generalization and Specialization

I propose a new method to find tasks directly, not webpages but tasks themselves.

I already proposed a new method to use sponsored search in Chapter 4 , but it does not search tasks directly but searches webpages, which contain more queries. Task search finds tasks directly so that it must detect tasks themselves. Therefor, it needs a novel method to discover not webpages but task sentences.

I am going to describe the novel method, which contains several tricks to find task sets with high completeness, purity and diversity. First, I will describe the way to frequently appearing tasks by generalization and specialization. Then, I will explain how to find relationships among tasks based on entailment relationship and hyponym relationship. Finally, I will explain the detailed procedure of the method.

5.1 Finding Frequently Appearing Tasks by Generalization and Specialization

In this section, I am going to describe how my novel method finds frequently appearing tasks.

5.1.1 Finding Similar Tasks

My method pays attention to frequency of tasks in webpages. I formed a hypothesis that when a task appears frequently in webpages, its usefulness is higher.

It is true that frequently appearing tasks are not always useful. There are some tasks appearing in webpages frequently although are useless. For example, it is possible that a sales company write articles on the web that the products of the company are good to perform the task. Even if the products are in bad quality, they can spread good rumors on the web.

Admittedly there are some unfair articles in the Internet; nonetheless, I would suggest that tasks with high usefulness tend to be written on the webpages frequently. This is because highly useful tasks spread mouth-to-mouth and via other media so that they will be shown in many webpages. Moreover,

most people use the Internet in their daily life recently, so they casually post and share their experience of performing tasks on the web. Therefore, Internet users often write facts that achieved the task actually, which are written in the definition of the usefulness. This is the reason why I suppose that frequently appearing tasks tend to be useful.

When I tried to evaluate the frequency of tasks, I tackled the problem of expression of tasks; Even if they are the same tasks, they sometimes described by different expressions on the web. For example, when you search with a query "treat hay fever" and find tasks "go to an ear, nose, and throat clinic" and "consult a doctor" at the same time, you can guess that "ear, nose, and throat clinic" and "doctor" indicates the same object. Furthermore, you can understand that "go to" and "consult" mean the same action from the context of the webpages. This reasoning is what ordinary human users do commonly; nevertheless, the search system must do automatically in my task search method.

I paid attention to generalization of tasks in order to solve this problem. I suppose that the system can guess the sameness of tasks by generalization of tasks.

I started this method from the unification of expressions using the template of a noun and a verb.

Japanese task expressions can take many kinds of forms, however, most tasks can change into the template of a object and a predicate such as "水を撒く (mizu wo maku, add water)" and "証明書に記入する (sho-meisho ni kinyu-suru, fill out the certificate)". Whereas I recognize that some tasks such, as "結婚する (kekkon suru, get married)", are not composed of object and predicate, they are comparatively rare. Since most tasks can be expressed with a object and a predicate, I concluded that the system should ignore tasks that are impossible to express with the template. Moreover, for most tasks, case particles such as "を (wo)" and "に (ni)" in predicates are not needed to understand the meaning of the task. Therefore, I decided to remove them from the predicate so that the templates composed of a noun and a verb without case particles such as $\langle$ 花粉症, 対策する $(treat, hayfever)\rangle$ express tasks in my task search system. Accordingly, I refer to a template as $\langle n, v\rangle$ hereafter, with n and v referring to

a noun and a verb.

Thanks to the simplification with the template of a noun and a verb, namely $\langle n, v \rangle$, the system can accept tasks as a combination of two words. This step makes the generalization of tasks easier so that it can compare tasks. I generalized tasks by generalizing nouns and verbs that compose of tasks separately, thereby I compared generalized tasks and check what the original tasks mean. I will describe how I generalized nouns and verbs in Section 5.1.2.

5.1.2 Generalization of tasks

In this research, I focused on hyponymy relationships among nouns and entailment relationships among verbs to generalize tasks, which are standardized into a pair of two words [13]. Both of the relationships are related to generalization and specialization of words, thereby, they make it possible to convert a task into generalize one. This section describes generalization of tasks based on hyponymy relationships among nouns and entailment relationships among verbs.

In this paper, I define "A noun n_1 is a hypernym to noun n_0" as "a noun n_1 is a kind or instance of n_0"; For example, Allelock is a kind of medicines, therefore medicine is a hypernym to Allelock. In addition, I do not distinguish concept-instance relation (such as medicine-Allelock) from hyponymy relation between concepts (such as medical items and medicine) so that I regard them as the same relationship. This definition is based on the definition of Sumida et al [12].

My definition of entailment is the same as that in Hashimoto et al [5] and WordNet; a verb v_1 entails v_2 if v_1 cannot be done unless v_2 is, or has been, done. For instance, when "divorce" has been done, "get married" had been done certainly, therefor; "divorce" entails "get married".

Hyponymy relations and entailment relations can be regarded as a kind of IS-A relation; Therefor, you can generalize nouns and verbs separately for each. I generalized nouns and verbs in tasks by converting nouns into their hypernyms and changing into their entailing verbs.

When the task given as a query is $\langle n_0, v_0 \rangle$, n_1 is a generalized word of n_0 and v_1 is a generalized word or v_0, I suppose that the three tasks $\langle n_1, v0 \rangle$,

Table 4: Generalized tasks

Generalization types	Tasks
	⟨ 診療科, 受診する ⟩
$\langle n_1, v_0 \rangle$	⟨ 診察科, 受診する ⟩
	⟨ 耳鼻咽喉科, 行く ⟩
	⟨ 耳鼻咽喉科, 相談する ⟩
	⟨ 耳鼻咽喉科, 探す ⟩
$\langle n_0, v_1 \rangle$	⟨ 耳鼻咽喉科, 訪れる ⟩
	⟨ 診療科, 行く ⟩
	⟨ 診療科, 相談する ⟩
	⟨ 診療科, 探す ⟩
	⟨ 診療科, 訪れる ⟩
	⟨ 診察科, 行く ⟩
	⟨ 診察科, 相談する ⟩
	⟨ 診察科, 探す ⟩
$\langle n_1, v_1 \rangle$	⟨ 診察科, 訪れる ⟩

$\langle n_0, v1 \rangle$ and $\langle n_1, v1 \rangle$ are the generalized tasks of the original task $\langle n_0, v_0 \rangle$; For example, "耳鼻咽喉科 (ear, nose, and throat clinic)" has hypernyms such as " 診療科 (clinic)" and "診察科 (clinic)", and also "受診する (go to the Clinic)" entails "行く (go)", "相談する (consult)", "探す (find)" and "訪れる (visit)". Therefore, when the given task is ⟨ 耳鼻咽喉科 $(ear, nose, and throat clinic)$, 受診する $(go to the clinic)$⟩, you can obtain following generalized tasks by combine generalized nouns and generalized verbs.

I generalized tasks by generalizing nouns and verbs for each in tasks. I consider that the system can check sameness of tasks by comparing generalized tasks so that it will evaluate the appearance frequency precisely.

5.1.3 Connecting Task Nodes by Sameness

I define an original task as a task that is obtained from the web directly; Namely, when you obtain a task from generalization, it is not a original task.

First of all, in my method, the system adds each tasks obtained from the web. At this time, all nodes in the task graph are not connected each other. Then, generalize them by hypernym and entailment. Generalized tasks are connected to the original task of them. At this time, some original nodes in the task graph are connected indirectly. Figure 11 represents the generalized tasks

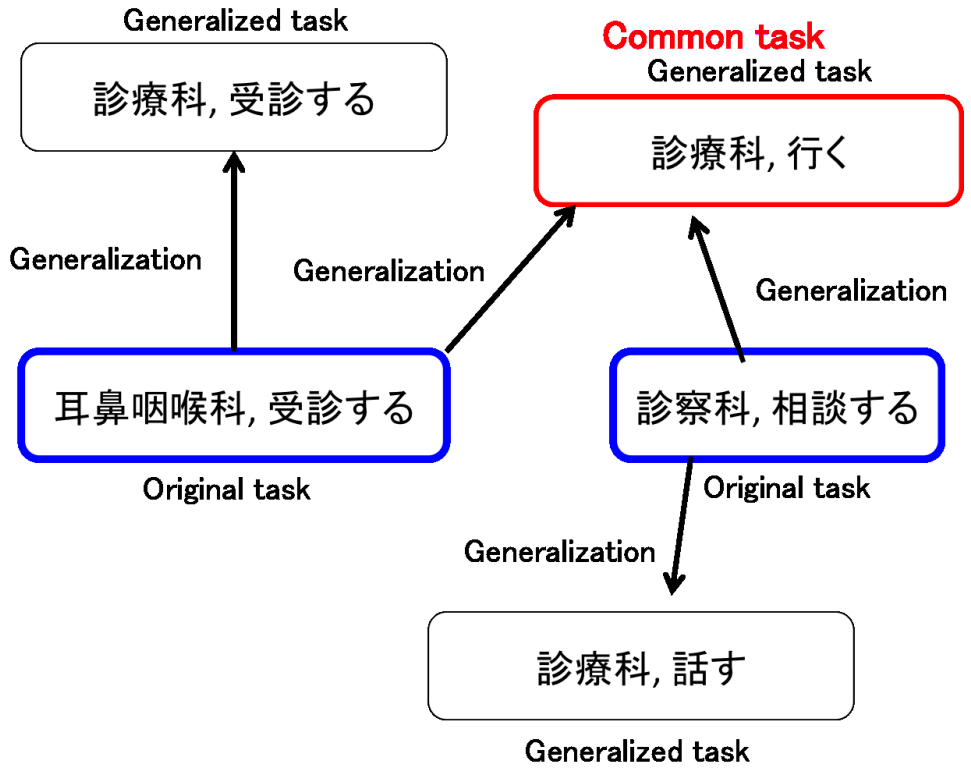

Figure 11: Task Generalization

in the task graph shown above. ⟨ 耳鼻咽喉科, 受診する ⟩ leads edges to ⟨ 診療科, 受診する ⟩, ⟨ 診療科, 行く ⟩ and ⟨ 診療科, 相談する ⟩. ⟨ 診療科, 相談する ⟩ also leads edge to ⟨ 診療科, 行く ⟩, so from this graph you can consider that ⟨ 耳鼻咽喉科, 受診する ⟩ and ⟨ 診療科, 相談する ⟩ are similar tasks.

In my method, I suppose that the similar tasks are actually the same tasks, so I count frequency of tasks by this mean. For example, if the graph shown in Figure 11 was given, my method count frequency of original task as below.

$$Freq\left(\langle \text{耳鼻咽喉科科}, \text{受診する} \rangle\right) = 2$$
$$Freq\left(\langle \text{診察科}, \text{相談する} \rangle\right) = 2$$

5.2 Finding Relations between Tasks

In this section I am going to describe how to find relationships among tasks by several means. As I wrote above, there are three kinds of relationships among tasks. Those relationships connect tasks in a complicated way, so I

adopt multiple methods to find relationships. First, I will describe methods to find SUBTYPE-OF relations. After that, I am going to explain how to discover PART-OF relationships. In addition, there are tricks to remove noises in PART-OF tasks. Finally, I will describe how to detect INSTANCE-OF relations from the task graph nodes.

5.2.1 Finding SUBTYPE-OF relations

I use hyponymy to detect SUBTYPE-OF relations. SUBTYPE-OF relations can be found when the tasks' objects are in hyponymy relationship. For instance, $\langle treat, ragweed - fever \rangle$ is SUBTYPE-OF $\langle treat, hay - fever \rangle$. You can suppose that when the given task's noun is a hyponym of the other task and the given task's verb equals to the others then they are in SUBTYPE-OF relationship.

When the system finds another pair of SUBTYPE-OF tasks that share the same task node, the system connects them; By this mean, the task search system build chained SUBTYPE-OF nodes.

5.2.2 Finding PART-OF relations

My method focuses on the sameness of webpage that the tasks are written in. When tasks are written in the same page, they tend to be in the PART-OF relationship to the same task. For example, most how-to sites are composed of multiple paragraphs that contain several tasks. Such tasks are usually not a single action but a sequence of multiple actions. Therefore, when you find multiple tasks in the same webpage, they may be parts of another task.

When the task search system finds multiple tasks in a single webpage, it supposes that they are in the same hierarchy from the aspect of PART-OF relationship. The parent node of them may be the inputted query task usually.

PART-OF tasks in task nodes frequently share the same tasks and sometimes they should be in the same hierarchy, therefore the novel method unites them based on their similarity and frequency of task elements in them. This system compares PART-OF tasks each by each. When a PART-OF task set T_1 and T_2 is given and $Similarity\,(T_1, T_2)$ is higher than the threshold th_1, then T_1 and T_2 is united. Iterate this procedure till all PART-OF tasks are tried. When I calculate similarity of sets of tasks, I adopted Jaccard similarity.

5.2.3 Removing Noises from PART-OF Tasks

When the system unites task sets, it calculates scores of tasks in them. If a certain task's score is lower than the threshold th_2, the task is removed from the task set. th_2 is calculated by the score of the tasks in task set. Scores' mean or average is used.

5.2.4 Finding INSTANCE-OF relations

Some tasks are not regarded as SUBTYPE-OF task and PART-OF task. My system regards them all as INSTANCE-OF tasks.

Scores rank all tasks and task sets. The task search method calculates score by task's frequency and webpages' attributes that the tasks were found.

5.3 Scoring and Ranking

In this section, I describe how to score tasks and task sets. After that, I will explain the method to rank task sets.

5.3.1 Calculating Score

In the task search method, tasks and task sets are mainly scored by their frequency as I wrote in 5.1. Original tasks are scored by their frequency and their hypernym's hyponym's frequency.

Moreover, I focused on webpage's features that indicates tasks' characters; For example, government official websites, such as Japan's Ministry of Health, Labour and Welfare's web site [1)] are more reliable than a ordinary person's blog. I suppose that tasks written in government websites or companies' websites tend to have more completeness, so the task search method use the feature that indicates the website's owner.

In addition, there are much false rumor in the Internet, therefore I need to remove doubtful tasks from the task search results. I realized that websites that sell something tend to contain unreliable information. In the task search method, I score tasks found in a commercial site lower than ones found in non-commercial sites.

To sum up, the task search method adds scores by these steps below.

[1)] http://www.mhlw.go.jp/

1. Calculate frequency of a task t by generalization and specialization.
2. Multiply f: the coefficient for official webpages.
3. Multiply c: the coefficient for commercial webpages.

5.3.2 Improving Diversity

An ideal task search result should contain diverse items in the list; Therefore, I evaluated the task search result in the aspect of diversity. Diversity indicates how many SUBTYPE-OF tasks' part-of tasks in the results, in addition to that, variety of task sets should appear in the highly ranked items. In accordance with this evaluation, the task search method adopts a trick to improve diversity in the result list.

My task search method uses MMR Algorithm[2] to improve diversity of the results. MMR Algorithm sorts documents by the equation below.

$$MMR = \arg\max_{d_i \in D \setminus S} \left[\lambda \left(Score\left(d_i\right) \right) - \left(1 - \lambda\right) \max_{d_j \in S} Sim\left(d_i, d_j\right) \right] \tag{1}$$

In this paper, D stands for sets of tasks and $Score\left(d_i\right)$ means a score for the task set d_i scored by frequency and webpages that the tasks belong to. Moreover, S stands for a list of task set selected as the output. $Sim\left(d_i, d_j\right)$ is Jaccard similarity of d_i and d_j. $\lambda \in [0, 1]$ is a parameter represents how much the system focuses on the importance of the task set and similarity of it and already chosen task sets. I set $\lambda = 0.5$ in the experiment.

5.4 Detailed Procedure

I describe the whole procedure of my task search method in this section. First, a user input a query that means a task. Then, the system converts the query into new queries by the method described in Chapter 4. After that, it executes web search and obtain webpages. The system collects texts that are likely to contain task words by the special patterns. It converts them by the template described in Section 5.1.1 and obtain tasks that are expressed by $\langle n, v \rangle$. Finally, add scores to them based on their frequency, commerciality and the degree of official and sort them by MMR algorithm. At this time, the system check the sameness of tasks not by simple coincidence but by generalization explained in

Section Chapter 5.

Here I describe the sequence of the method.

1. Obtain a query q that indicates a task.
2. Convert q into query set Q by the method explained in Chapter 4.
3. Perform web search by $q_{1,\ldots} \in Q$ and obtain n webpages $\{p_1, \ldots, p_n\}$.
4. Split texts in webpages by marks m stand for the end of a sentence. After that, obtain all sentences from each webpages $\{p_1 \ldots p_n\}$.
5. Obtain sentanses $\{s_1, \ldots, s_m\}$, which contains a pattern that is frequently appering in imperative sentences by morphological analysis.
6. Convert sentences into pairs of words by the template $\langle n \in N, v \in V \rangle$ and generate tasks $t \in T$. N is a set of nouns and V is a set of verbs. Pairs of words stand for tasks in this paper.
7. Do these actions below to original tasks $t \in T$.
   (a) Obtain hypernyms $\{h_{t1}, \ldots, h_{tk}\}$ from the task t.
   (b) Obtain entailing terms $\{e_{t1}, \ldots, e_{tj}\}$ from the task t.
   (c) Join $\{h_{t1}, \ldots, h_{tk}\}$ and $\{e_{t1}, \ldots, e_{tj}\}$ so that obtain $\{\langle h_{t1}, e_{t1}\rangle, \ldots, \langle h_{tk}, e_{tj}\rangle\}$.
8. Do these actions below to original tasks $t \in T$.
   (a) Do these actions below to original tasks $t_k \in T$.
      i. Compare t with t_k. When a generalized task of t is $t_g \in \{\langle h_{t1}, e_{t1}\rangle, \ldots, \langle h_{tk}, e_{tj}\rangle\}$ equals to a generalized task of $t_k, t_k \in \{\langle h_{t1}, e_{t1}\rangle, \ldots, \langle h_{tk}, e_{tj}\rangle\}$, add one to the score $score_t$ that indicates frequency of t at this time.
9. Do these actions below to original tasks $t \in T$.
   (a) Do these actions below to original tasks $t_k \in T$.
      i. If t and t_k were found on the same webpage, generate a new task set T_p that contains t and t_k. Then add T_p to the list of set tasks L.
10. Do these actions below to all the generated task sets $T_p \in L$.
    (a) Compare other generated task sets and if they share the same webpage, unite them.
11. Do these actions below to original tasks $t \in T_p, T_p \in L$.
    (a) When the webpage that t was found contains characteristic words for

official webpage of governments or companies, multiply $score_t$ by f; f is the coefficient for official webpages as I described in Section 5.3.1.

(b) When the webpage that t was found contains characteristic words for commercial webpages that sell something, multiply $score_t$ by c; c is the coefficient for commercial webpages as I described in Section 5.3.1.

12. Sort the sets of tasks $T_p \in L$ by MMR with the total score of tasks in them.

The task search system converted queries by the method I described in Chapter 4, however, its procedure is not completely the same. I selected only sahen-nouns when the system collected frequently appearing terms from sponsored search advertisements; hence, the system did not select verbs. This is because verbs in a query sometimes cause useless search. Furthermore, in this experiment the expanded queries include the original query.

I set $n = 1000$ so that the task search system retrieved one thousand webpages per query; Namely, when the system generated four expanded queries, it searches 200 webpages per expanded query including the original query.

I used Bing Search API [1] to perform web search. I also used ALAGIN forum's entailment relation database [2] [18] to generalize tasks. This entailment dictionary includes eight types of dictionaries. I adopted proper entailment relation dictionary, premise relation dictionary and non-entailment prediction dictionary. Moreover, I utilized ALAGIN forum's hyponymy extraction tool [3]; It contains about 6,000,000 pairs of hyponymy relations extracted from Wikipedia's Japanese site [4].

I used MeCab [10] to parse Japanese texts, furthermore, I also putted the word dictionary generated from Wikipedia and Hatena Keyword [5] to parsing. I used NetworkX [6] to generate graphs.

I used these patterns following in Step 5.

- ～ください・～下さい

[1] http://datamarket.azure.com/dataset/bing/search
[2] https://alaginrc.nict.go.jp/
[3] http://alaginrc.nict.go.jp/hyponymy/
[4] http://ja.wikipedia.org/
[5] http://d.hatena.ne.jp/keyword/
[6] http://networkx.github.io/

- ～ましょう
- ～なさい
- ～しよう
- ～べき
- ～といい・～とよい・～と良い
- ～必要が
- ～ては

In Step 5, I removed some tasks including stop words. Here I show some stop words below.

- お気に入り
- コメント
- メールアドレス
- JavaScript 対応ブラウザ

These words appear very frequently in the Internet, however, most tasks include them are useless; Therefore, I removed them from the task search results as stop words.

Chapter 6　Experiment

I performed the experiment to evaluate the task search method. I am going to show the results and discuss them.

6.1　Task Search Results

I performed task search by two methods. The first one is which I described in Chapter 5 and the other one is a baseline. I am going to explain the baseline method and show differences between them and finally discuss advantages and disadvantages of the method.

I built the baseline method with almost the same way to my task search method except for scoring and organizing. The baseline method does not score tasks by their frequency by generalization, but only by simply appearing. In addition, it does not care if the tasks belong to a commercial site or a governmental site. Moreover, it does not organize tasks by their entailment relations and hyponym relations. Besides, it does not rerank the result by MMR Algorithm so that it has not a trick to improve diversity.

Table 5 represents the highest three rankings of the task search with a query "部屋　掃除する (clean the room)". Scores on the table were calculated by tasks' frequency of appearing, websites the tasks belong to.

The results of the task search method are comparatively consistent rather than the baseline method; For instance, the baseline method's tasks in the third ranking seem to be strange from the aspect of task search. They are from a personal blog article and the baseline method does not regulate them. Moreover, my task search method improved purity of the first ranked task set. The first ranked task set of each methods are from the same webpage, therefore comparing them make the difference clear. The proposed method removed the task "掃除 する (Clean it)" from the task set. This procedure improves purity, but maintains diversity and completeness because they are not an important task to clean the room; The task "掃除 する (Clean it)" is not meaningful information because it is too natural and less in entropy. The proposed method also removed the task "明細 貰う (take the detailed statement) " from the

task set. This procedure cause decline of purity and completeness since it is a meaningful task to accomplish the query task “部屋 掃除する (clean the room)”.

6.2 Evaluate Task Search Results

Table 6 represents the results of the task search performed by the proposed method. Table 7 represents the results of the baseline. Each method performed task search by common eight queries. I evaluated them by ten ranked task sets from the top.

6.2.1 Evaluating Completeness

Four people evaluated completeness of the task sets in the two methods. When I explained how to evaluate completeness in the experiment, I did not tell them which is the proposed method. I let the evaluators read the tasks in the task sets, after that, let them add completeness scores from zero point to three points by the criterion below.

- Score 0: Even if you perform the tasks in the task set, you cannot accomplish the goal represented by the query.
- Score 1: If you perform the tasks in the task set, you achieve at least ten percent of all goals that are represented by the query.
- Score 2: If you perform the tasks in the task set, you achieve at least fifty percent of all goals that are represented by the query.
- Score 3: If you perform the tasks in the task set, you achieve at least eighty percent of all goals that are represented by the query.

Moreover, the evaluators add completeness scores as the whole ten task sets for each method. When they evaluate completeness of multiple sets of tasks, they regard a list of task sets as a greater task set includes the all tasks. “Completeness” in Table 6 are scores stands for the whole ten task sets. “Sum of completeness” indicates the sum of each completeness scores in a method.

Generally, each completeness of the task search method is the same or higher than the baseline method. The scores of sum of completeness indicates the same tendency. The task search method is, at least, not obviously worth than the baseline. In some queries such as “部屋 掃除する (Clean a room)” and “ビリヤード 優勝する (Win on cue sports)” my task search method is better

than the baseline, although, in some queries, namely “保育園 入園させる (Join a kindergarten)” and “胃もたれ 防止する (Prevent stomach problems)”, the baseline method overtopped the task search method. This is because the task search method removes tasks from task sets. When useless tasks are removed, the task search method wins advantage, but if useful tasks are removed, its completeness will be worse.

6.2.2 Evaluating Purity

I made solution sets for each queries in order to evaluate purity and diversity. These solution sets focus on SUBTYPE-OF tasks because purity and diversity are based on SUBTYPE-OF relations as I described in Section 3.3.

Purity stands for the degree of consistence of a set of tasks; Namely, when a task set contains many tasks that are classified to variety of SUBTYPE-OF tasks then its purity declines. When you evaluate a task set’s purity, you must consider which SUBTYPE-OF tasks they belong. Most query tasks have multiple SUBTYPE-OF tasks, however, the number of SUBTYPE-OF tasks is ambiguous; For example, how many SUBTYPE-OF tasks the query “Clean a room” have? you may make up with “Clean a bathroom”, “Clean a nursery” and “Clean a kitchen”, though, it is not enough. How about “Clean a Japanese-style room?” How about “Clean a Chinese-style room ?” Eventually, the number of SUBTYPE-OF tasks is uncountable. Therefore, I made the solution set by the task search results and evaluated the methods relatively.

An original query task is also a SUBTYPE-OF task itself; therefore, when the query “部屋 掃除する (Clean a room)” is given, SUBTYPE-OF tasks results from the query become like this.

- Clean a bathroom
- Clean a room when you move
- Clean a room

SUBTYPE-OF tasks are not exclusive; therefore, sometimes they overlap each other on a task. For example, “Clean a room when you move” connects SUBTYPE-OF relation to the SUBTYPE-OF tasks “Clean a bathroom” and “Clean a room when you move”.

I labeled the task sets as a belonging of a SUBTYPE-OF task by counting

the SUBTYPE-OF tasks; Table 6 and Table 7 represents the results. Task sets are labeled based on the most dominant SUBTYPE-OF task set T_s. I defined $TaskType\,(T)$ that the most natural SUBTYPE-OF task for a task set T. I define S as the set of SUBTYPE-OF task sets, therefore, $TaskType$ is chosen by the equation below.

$$TaskType\,(T) = \arg\max_{T_s \in S} |T \cap T_s| \tag{2}$$

A task set's purity is scored by $TaskType\,(T)$ and tasks in the task set. I define $Children\,(t)$ that tasks belong to the SUBTYPE-OF task t. The purity score of a task set T is calculated by the equation below.

$$Purity\,(T) = \frac{|Children\,(TaskType\,(T))\,|}{|T|} \tag{3}$$

The purity scores in Table 6 and Table 7 are the average score of ten task sets per the queries.

Most purity scores of my task search method are higher than the ones of the baseline; Especially, the scores of the query "ビリヤード 優勝する (Win on cue sports)" are contrasting. My task search method scores 0.52, on the other hand, the baseline method scores 0.29. This is because my task search method removes useless tasks thereby condensing the task sets. Moreover, my task search method adds lower score to commercial sites and higher score to official sites so that it avoids noises.

6.2.3 Evaluating Diversity

Diversity is evaluated to a list of task sets, hence I added scores to the task search result list. Diversity also is based on SUBTYPE-OF relations; therefore, I used the solution set described above. Diversity scores on Table 6 and Table 7 are the number of SUBTYPE-OF tasks at ten rankings. Comparing with the task search method and the baseline, they are not different about diversity. This is because the original tasks are almost same; Furthermore, frequently appearing tasks tend to be similar.

To sum up the results of the experiment, my task search can be described as below.

- Maintains completeness.
- Improves purity
- Maintains diversity.

6.3 Discussion

The results of the experiments indicates that the task search method in superior to the baseline in purity. However, in the case of some queries it does not succeed in purity and other evaluation measures. To sum up, queries that are suitable for the task search method are they below.

- 部屋 掃除する (Clean a room)
- ビリヤード 優勝する (Win on cue sports)
- 保育園 入園させる (Join a kindergarten)

On the other hand, queries that are suitable for the baseline method are these following items.

- 胃もたれ 防止する (Prevent stomach problems)

The reason why there are suitable queries and not suitable queries can be described as below. For some queries, such as "部屋 掃除する (Clean a room)", the best ways to achieve the goal are written frequently, In these cases, when you perform these tasks, you can accomplish tasks (almost) completely. This is a suitable query case. Meanwhile, for queries that have useless tasks on many webpages, tasks found by the proposed method become meaningless ones; Namely, the task search method amplifies noises. Webpages searched by the query "部屋 掃除する (Clean a room)" contained many useless tasks, thereby the results were worse.

In addition, there are some queries that resulted in useless tasks only both in the proposed method and the baseline. Here is the list of the queries leaded useless tasks.

- クレー射撃 体験する (Try clay target shooting)
- ビリヤード 優勝する (Win on cue sports)

When I searched with these queries in an ordinary web search engine, I did not find good tasks to accomplish the goal in the top rankings. In other words, Bing does not offer good tasks for such queries. This problem occurs because

only a few people write about how to try clay target shooting and how to win on cue sports. Clay target shooting and cue sports are both unpopular sports; therefore it is difficult to find good tasks.

Table 5: Task Search Results

Rank	Proposed method	Baseline
1	新聞 使う 全体 拭く 拭き する 不動産業者 確認する 契約 確認する 説明書通り 利用する 換気 充分する きれい する 状態チェック 済ませる 額 差し引く 管理者 連絡する 確認 する 賃貸借契約書 チェックする 換気 する	きれい する 額 差し引く 状態チェック 済ませる 明細 貰う 新聞 使う 説明書通り 利用する 確認 する 管理者 連絡する 契約 確認する 様 する 拭き する 賃貸借契約書 チェックする 掃除 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 掃除 する 引越し業者 決める ハタキ 利用する 汚れ ためる きれい する 洗剤 かける フキン 拭く マスク 活用する 雑巾がけ する 商品内容 確認する 物 置く 洗剤 使う	運気 乱す きれい する リビング ある 状態 整理する 世話 する 植物 育てる 優先 片付ける そこ 片付ける リビング 置く 掃除 する 分別 行う 数 減らす
3	植物 育てる 掃除 する 世話 する 運気 乱す 状態 整理する リビング ある 優先 片付ける きれい する 数 減らす 分別 行う リビング 置く そこ 片付ける	空間 散らかる 公園デビュー する 工作 する 空間 想像する モテ護身術 マスターする 身 持つ クリスマスイヴ 掃除する ゴミ 出す 入浴剤 解消する これから 書く 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

Table 6: Evaluation of Proposed method

Query	Completeness	Sum of completeness	Purity	Diversity
部屋 掃除する	3.0	17.5	0.35	3
ビリヤード 優勝する	2.0	8.25	0.52	3
家庭菜園 始める	3.0	14.5	0.40	5
保育園 入園させる	3.0	14.25	0.59	3
花粉症 対策する	3.0	16.75	0.41	2
犬 育てる	2.5	6.75	0.25	3
胃もたれ 防止する	2.75	13.75	0.43	3
クレー射撃 体験する	2.25	4.75	0.14	2

Table 7: Evaluation of Baseline

Query	Completeness	Sum of completeness	Purity	Diversity
部屋 掃除する	2.5	15.0	0.28	3
ビリヤード 優勝する	1.5	5.75	0.29	3
家庭菜園 始める	3.0	15.75	0.37	5
保育園 入園させる	2.75	11.75	0.34	2
花粉症 対策する	3.0	17.75	0.41	2
犬 育てる	3.0	10.5	0.22	2
胃もたれ 防止する	3.0	14.75	0.48	3
クレー射撃 体験する	2.25	6.5	0.14	1

Chapter 7 Conclusion

In this paper I proposed the task search method based on generalization and specialization. This method scores tasks by their frequency, however, it focuses on relations of hyponymy and entailment thereby finds tasks based on ontology.

I defined a task as an action that will achieve the whole goal or a part of the goal. The purpose of this research is to find tasks that make the user accomplish the goal. Hence, I analyzed tasks based on their goals and actions.

I found that three types of relations are important to task search. I named these relations as follows.

- INSTANCE-OF relation: If you perform t_1, the goal, which would be achieved when t_0 is performed, will be wholly achieved. In this case, t_1 connects to t_0 with INSTANCE-OF relation.
- PART-OF relation: Task t_1 is consisting of a task t_0, which is an sequential task. In this case, t_1 connects to t_0 with PART-OF relation.
- SUBTYPE-OF relation: If you perform t_1, the goal, which would be achieved when t_0 is performed, will be partly achieved. In this case, t_1 connects to t_0 with SUBTYPE-OF relation.

All tasks are connected by these relationships. When a user perform task search, these relations are important to find and evaluate tasks.

Furthermore, I defined three evaluation measures for task search as below.

- Completeness
- Purity
- Diversity

Completeness is evaluated on each tasks. Purity is evaluated on task sets. Diversity is evaluated on a list of task sets. These evaluation measures are scored based on the relationships I described above, especially SUBTYPE-OF relations.

Moreover, I proposed a query expansion method for task search based on sponsored search. I showed that query expansion based on sponsored search improves completeness of task search. I performed a experiment that represents the advantages of the query expansion method.

Finally, I implemented the task search system and performed a experiment. The experiment demonstrated that the task search method is excellent in terms of purity. Moreover, this method removes meaningless noises, although, it does not decrease completeness and diversity.

Acknowledgments

I appreciate the feedback offered by supervisor Professor Katsumi Tanaka. My research adviser Professor Masatoshi Yoshikawa gives insightful comments and suggestions. I have had the support and encouragement of my research adviser Professor Kazutoshi Sumiya. Associate Professor Hiroaki Ohshima's meticulous comments were an enormous help to me. I received generous support from Assistant Professor Makoto Kato. Associate Professor Takehiro Yamamoto gives me constructive comments and warm encouragement. I would like to express my sincere gratitude to Satoshi Ohshige, Komei Morii, Hikaru Takemura and Hiroki Komurasaki for participating in the experiment. I thank ALAGIN Forum for permission to use the entailment dictionaries. I would also like to express my gratitude to Ken Kaizu for permission to use the dictionary he generated. Finally, I would like to thank all the members of TANAKA Laboratory.

References

[1] Andrei Broder, Marcus Fontoura, Vanja Josifovski, and Lance Riedel. A semantic approach to contextual advertising. In *Proc. of SIGIR*, pages 559–566, 2007.

[2] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *In Research and Development in Information Retrieval*, pages 335–336, 1998.

[3] Cristian Danescu-Niculescu-Mizil, Andrei Z. Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. Competing for users' attention: on the interplay between organic and sponsored search results. In *Proc. of WWW*, pages 291–300, 2010.

[4] Tim Finin, Yun Peng, R. Scott, Cost Joel, Sachs Anupam Joshi, Pavan Reddivari, Rong Pan, Vishal Doshi, and Li Ding. Swoogle: A search and metadata engine for the semantic web. In *In Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, pages 652–659. ACM Press, 2004.

[5] Chikara Hashimoto, Kentaro Torisawa, Kou Kuroda, Masaki Murata, and Junichi Kazama. large-scale verb entailment acquisition from the web. In *conference on empirical methods in natural language processing (emnlp2009), pages 1172-1181*, 2009.

[6] Ahmed Hassan and Ryen 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.

[7] Bernard J. Jansen and M Resnick. Examining searcher perceptions of and interactions with sponsored results. In *Workshop on Sponsored Search Auctions*, 2005.

[8] Rosie Jones and Kristina Lisa Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proc. of CIKM*, pages 699–708, 2008.

[9] Shinzato Keiji, Shibata Tomohide, Kawahara Daisuke, and Kurohashi Sadao. Tsubaki: An open search engine infrastructure for developing in-

formation access methodology. *IPSJ*, 52(12):12p, dec 2011.

[10] T. KUDO. Mecab : Yet another part-of-speech and morphological analyzer. *http://mecab.sourceforge.net/*, 2005.

[11] Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proc. of WWW*, pages 521–530, 2007.

[12] Asuka Sumida, Naoki Yoshinaga, and Kentaro Torisawa. Hyponymy relation acquisition from hierarchical layouts (in Japanese). In *Journal of Natural Language Processing Vol.16 No.3*, 2009.

[13] Idan Szpektor and Ido Dagan. Learning entailment rules for unary templates. In *In Proceedings of the 22nd International Conference on Computational Linguistics (COLING2008)*, 2008.

[14] 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.

[15] 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.

[16] Setsuji Tamugi, Koji Zettsu, Yutaka Kidawara, and Katsumi Tanaka. Service information search and ranking by eligibility (in Japanese). In *IPSJ SIG DBS*, 2012.

[17] Dwi H. Widyantoro and John Yen. A fuzzy ontology-based abstract search engine and its user studies. In *IN THE 10TH IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS*, pages 1291–1294. IEEE Press, 2001.

[18] Ichiro Yamada, Chikara Hashimoto, Jong-Hoon Oh, Kentaro Torisawa, Kow Kuroda, Stijn De Saeger, Masaaki Tsuchida, and Jun'ichi Kazama. Generating information-rich taxonomy from wikipedia. In *Proc. of IUCS-2010*, 2010.

[19] 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.

[20] Takayuki Yumoto. Hierarchizing how-to information based on correspondence of procedure (in Japanese). In *DEIM2013*, 2013.

[21] Ruoyan Zhang, I. Budak Arpinar, and Boanerges Aleman-meza. B.: Automatic composition of semantic web services. In *Intl. Conf. on Web Services, Las Vegas NV*, pages 38–41, 2003.