

Web サービスを対象とした検索意図推定とその応用

田麥 節治[†] 赤星 祐平[‡] 是津 耕司[‡] 木俣 豊[‡]

田中 克己[†]

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

[‡] 情報通信研究機構ユニバーサルコミュニケーション研究所

〒619-0289 京都府相楽郡精華町光台 3-5

E-mail: [†] {tamugi, tanaka}@dl.kuis.kyoto-u.ac.jp [‡] {akahoshi, zettsu, kidawara}@nict.go.jp

あらまし 本論文では、Web サービス検索においてユーザの要求に適合するサービスを見つけるために検索クエリから検索意図を推定する枠組みを提案する。利用者から入力される検索クエリから、どのような種類のサービスを利用したいかという意図推測を実現するため、クエリとサービスの種類の関係性を分析し、意図推定実現の枠組みを説明する。その際、クエリとサービスの種類の結びつけにおいては、検索連動型広告に着目する。

キーワード Web サービス, Web サービス検索, 意図推定

1. はじめに

近年のインターネットの普及に伴い、人々が Web 上で EC サイトや予約サイトなどに代表されるような Web サービスを利用する機会が増えている。ユーザが、利用したいサービスを実行するまでの一般的なプロセスとして、以下のような流れが考えられる。まず、ユーザは適切なクエリにより Web 検索を実行し、その検索結果の中から利用したいサービスを提供していると思われるページを選択する。次に、そのページ内の入力フォームを埋めることにより、サービスを実行する条件を指定する。その結果、そのサービスが、指定した条件のもと実行可能か実行不可能かが出力される。もし、そのサービスが実行不可能だった場合は、実行可能なサービスが見つかるまで、他のサービスに対して実行条件を繰り返し指定する必要がある。

例として、ユーザが京都のホテルを予約しようとする場合を考える。まず、このユーザは、「京都 ホテル 予約」などのクエリで検索を行う。次に、その検索結果の中から、ホテル予約サービスを提供していると思われるページを選択する。そして、そのページ内の入力フォームに従って、人数や宿泊日などの条件を送信し、その条件で空室があるかどうかの結果を得る。もし空室があれば、そこで予約を実行することで、ユーザの要求は満たされるが、もし空室がなかった場合、他のホテル予約サービスページにアクセスし、同じようにサービス実行の条件を指定する必要がある。

このように、現在の検索システムでは、京都のホテルを予約できるページを含むリストが結果として出力されるが、実際にそれらのページで提供されるサービスを介して予約を実行可能かどうかは、検索結果を得

る段階では分からないという問題がある。

そこで、本研究では、サービスをある実行条件のもと実行できるかどうか、つまり「今日、京都のホテルに泊まれるかどうか」や「在庫があって PORTER のかばんを買えるかどうか」などのような、サービスの実行条件と実行結果を含めた Web サービス検索の実現を目的とする。つまり、ユーザが入力した検索クエリから、サービス実行の条件とその条件のもとで実行可能なサービスを一気得られるような技術の基本概念を提案する。その技術は次の三つの要素から構成される。まず、ユーザが入力したクエリから、ユーザが意図したサービスの推定を行う。次に、クエリを構成する語などを用いてサービスの実行条件を指定し、サービスの実行結果を得る。最後に、得られた結果の中から実行可能であるもののみをフィルタリング・ランキングし、ユーザへ提示する。このような、従来のサービス検索を拡張した枠組みにより、ユーザは実行可能なサービスを検索することができるようになる。

本論文では、この三つの構成技術の第一段階である、ユーザの検索意図推定の枠組みを提案する。また、クエリから意図したサービスを推定するためには、クエリとサービスの関係性を分析する必要があるので予備調査を行う。その際、ユーザが入力したクエリとサービスを結びつけるものとして、検索連動型広告に着目する。この予備調査により、クエリの語の組み合わせとサービスの相関関係を示し、提案する意図サービス推定の手法の有用性を示す。

本論文の構成は以下になる。まず、1.1, 1.2 節で、本研究の基本概念と提案技術を述べる。次に、2 節で関連研究を述べ、3 節で提案技術の意図サービス

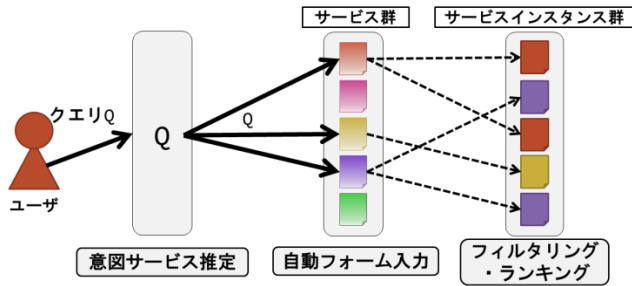


図 1: 実行可能サービスインスタンス検索

推定の部分について詳しく述べる。4 節で、意図サービス推定のための予備調査を行い、5 節でクエリとサービスの相関関係を分析する。最後に、6 節で考察、7 節でまとめを行う。

1.1. 実行可能サービスインスタンス検索

本節では、本研究が提案する、Web サービス検索の基本概念について述べる。

1.1.1. サービスインスタンス

サービスインスタンスとは、「サービスの実行条件」と「サービスの実行結果」の組を表す。ホテル予約を例に考えると、サービスの実行条件となるものは、ホテルの地域や宿泊人数、泊数などである。また、サービスの実行結果は「空室があり予約ができる」や「空室がなく予約ができない」などである。「京都のホテルで2月17日から大人二人で一泊」という条件で「空室があり予約ができる」という結果なら、この条件と結果の組が一つのサービスインスタンスとなり、これを「実行可能なサービスインスタンス」と呼ぶことにする。仮に空室がない場合でも、その条件と結果の組は一つのサービスインスタンスであるが、この場合は「実行不可能なサービスインスタンス」となる。

1.1.2. 実行可能サービスインスタンス検索

前節で述べたような、「実行可能なサービスインスタンス」を検索可能にするための、Web サービス検索の枠組みを提案する。

図 1 に、実行可能サービスインスタンス検索の枠組みを示す。提案手法は三つの構成技術からなる。まず、ユーザが入力したクエリをもとに、ユーザが意図したサービスの推定を行う。クエリとして入力された語の組み合わせから、適したサービスを推定する。次に、推定されたサービスの種類に当てはまるサービスが持つ各入力フォームに対して入力を行い、サービスの実行条件を指定する。その際、フォーム入力にはユーザがクエリとして入力した語を利用する。その結果得られるサービスインスタンスに対して実行可能かどうかでフィルタリングを行い、実行可能なものをランキングして出力する。このような処理の流れにより、ユーザが入力したクエリから、実行可能なサービスインス

タンスを検索できるようになると考えられる。次節で、これらの構成技術について述べる。

1.2. 構成技術

この節では、前節で述べた全体の枠組みを構成する技術のうち、自動フォーム入力とフィルタリング・ランキングについて簡単に説明する。なお、意図サービス推定については3節で詳しく述べる。

1.2.1. 自動フォーム入力

サービスインスタンス群を得るためには、各サービスが持つ入力フォームを、ユーザがクエリとして入力した複数の語を用いて埋める必要がある。そのために、各サービスの種類ごとに、サービスの型となるものを用意しておく。サービスの型には以下のような情報が含まれる。

- サービスの入力フォームの種類
- 各入力フォームが受付可能な値の種類
(例：地名、ブランド名、日付など)
- 入力フォームを実行した時の出力

例として「ホテル予約」のサービスについて考える。一般的に、ホテル予約サービスがもつ入力フォームには、「ホテルの地域」や「チェックイン」、「人数」などの種類がある。このそれぞれについて、入力フォームが受け付けることができる値は、「地名」、「日付」、「数字」というようになる。また、ホテル予約ではフォーム入力を実行した結果、「空室あり」か「空室なし」かが分かるようになっている。これらの情報をサービスの型として持つようなインデックスを利用することで、ユーザがクエリとして入力した語を、辞書などを利用して分析することで、どのフォームにどの語を入力すべきか判断できると考えられる。

1.2.2. フィルタリング・ランキング

自動フォーム入力の結果得られたサービスインスタンスのうち、実行可能なもののみをフィルタリングして抽出し、それをランキングして出力する必要がある。ランキングの要素として考えられるものとして、値段などが挙げられる。

2. 関連研究

2.1. クエリの分類

Broder[1]は、Web 検索のためのクエリを三つのタイプに分類できると述べている。その三つとは、インフォメーションクエリとナビゲーションクエリ、トランザクショナルクエリである。インフォメーションクエリとは、ユーザが検索クエリとして入力したワードが持つ意味や情報を得るために入力するクエリのタイプである。ナビゲーションクエリは、ユーザが検索クエリとして入力したワードが示す Web サイトを見つけるために入力するクエリのタイプであ

る．次に，トランザクショナルクエリとは，ユーザが検索クエリとして入力したワードによって示されるようなサービスを提供する Web ページを見つけるために入力するクエリのタイプである．つまり，Web サービス検索をする際，ユーザはこのトランザクショナルクエリを入力すると考えられる．本研究では，ユーザが入力するクエリはトランザクショナルクエリであることを前提としている．

2.2. トランザクショナルクエリの判別

Kang [2]は，クエリタイプの分別のために，リンクのアンカーテキストとリンク付けされたオブジェクトの種類の関係を用いて機械学習を行い，トランザクショナルクエリの判別を行っている．この研究では，まず，リンクのアンカーテキストに，リンク付けされたオブジェクトの種類をタグ付けする．オブジェクトの種類の例として，Site や Music， Service などがある．この種類により，リンク付けされたオブジェクトを「読む」のか「訪れる」のか「ダウンロードする」のかが分かる．そして，リンクのアンカーテキストから手がかりとなるようなワードを抽出し，そのキーワードとリンク付されたオブジェクトの種類をクエリのタイプ別のテストデータとして学習を行い，トランザクショナルクエリの判別を行っている．本研究では，トランザクショナルクエリの判別ではなく，トランザクショナルの種類判別を目的とする．つまり，ユーザの「サービスを実行したい」という意図ではなく，さらに一歩踏み込んだ「どの種類のサービスをどのように実行したい」という意図をクエリから推定することを目指す．

2.3. クエリの意図推定

Hu ら[3]は，ウィキペディアの構造を用いたユーザのクエリの意図推定の研究を行っている．

また，杉浦ら[4]は，ユーザが入力したクエリを，どの専門的な検索エンジンに振り分けるべきかを判別する研究を行っている．既存の Web ドキュメントから検索エンジンのトピックを学習する一方で，与えられたクエリに対してクエリ拡張を行い，そのトピックを認識する．そして，クエリとトピックが最も類似する検索エンジンにクエリを入力するという手法である．この研究は，捉え方によってはユーザが入力したクエリの意図を推定していると考えられる．その点では，本研究におけるユーザが入力したクエリから意図したサービスを推定するという点において関連していると考えられる．

本研究では，検索連動型広告から得られる知見を利用して意図サービス推定を行うという点でこれらの研究と異なると言える．

3. 意図サービス推定

本節では，1.1.2 節で述べた構成技術のうちの第一段階である，意図サービス推定について述べる．

3.1. 意図サービス推定の手法

1.1.2 節で述べたような Web サービス検索の枠組みを実現するためには，まずユーザが入力したクエリから意図したサービスを推定する必要がある．本研究では，ユーザが入力するクエリはトランザクショナルクエリであるということを前提とする．例えば，ユーザが「京都 ホテル」というクエリを入力したとする．この場合，ユーザの意図として，「京都のホテルを予約したい」や，「京都のホテルについて調べたい」というものが考えられるが，本研究ではこのクエリにおけるユーザの意図は前者であるとする．

意図サービス推定の手法を提案する．本研究では，ユーザが入力したクエリから，ユーザが「何に関して，どうして，どうなりたい」と考えているのかを判別することが意図サービス推定であると仮定する．例えば，京都のホテルの予約をとろうとしたユーザが入力したクエリが「京都 ホテル 予約」である場合，このクエリから「旅行業に関して，京都エリアのホテルの空室を検索して，予約をとりたい」という意図を判別することが意図サービス推定となる．

ここで，この意図サービス推定の手法の形式化を行う．ユーザが入力したクエリからユーザが「何に関して」「どうして」「どうなりたいか」を推定するというのを，より抽象的な表現で言い換えると，クエリから「サービスの種類」「ファンクション」「ファンクションの実行結果」を推定することであると言える．ファンクションとは，サービスがもつ機能であり，これによりユーザとのインタラクションをとる．サービスを提供する Web ページが入力フォームを持つものであるとすると，この三つ組を「We サービスページ」「フォームとフォーム入力値」「フォームの実行結果」ととらえることもできる．つまり，「京都 ホテル 予約」というクエリの例を考えると，「旅行業のサービスページで，ホテルの空室検索のフォームでエリア条件を京都に指定して，空室があれば予約をしたい」というようにクエリを解釈し，ユーザの意図を推定することができる．

このように，本研究では，ユーザが入力したクエリから，「サービスの種類」「ファンクション」「ファンクションの実行結果」を解釈することが出来ればユーザが意図したサービスインスタンスを推定することができるという仮定に基づき，ユーザの意図サービス推定を行う．

表 1 クエリ「京都」における検索連動型広告のリンク先サービスが持つ(T, F, R)の例
サービスの種類 T ファンクション $F1$ ファンクション $F2$ ファンクション $F3$ ファンクションの実行結果 R

旅行業	宿・ホテル検索	ツアー検索	航空券検索	プランを選び予約を行う
旅行業	ツアー検索	バス検索	レンタカー検索	詳細条件を指定しプラン選択
歯医者	-	-	-	-
賃貸業	資料請求	-	-	資料入手

表 2 クエリ「京都 東京」における検索連動型広告のリンク先サービスが持つ(T, F, R)の例
サービスの種類 T ファンクション $F1$ ファンクション $F2$ ファンクション $F3$ ファンクションの実行結果 R

旅行業	空席確認	バス予約	-	プランを選び予約を行う
旅行業	ツアー検索	バス検索	レンタカー検索	詳細条件を指定しプラン選択
旅行業	バス空席検索	フェリー検索	ツアー検索	プランを選び予約を行う
バス事業	バス検索	-	-	プランを選び予約を行う

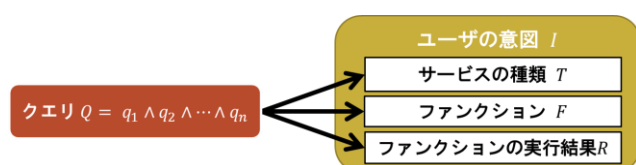


図 2: 意図サービス推定のモデル

3.2. 意図サービス推定のモデリング

ここで、クエリとユーザの意図の関係のモデリングを行う。ユーザが入力する n 個のワードからなるクエリ Q を $Q = q_1 \wedge q_2 \wedge \dots \wedge q_n$ と定義する。サービスの種類の集合 T を $T = \{t_1, t_2, \dots, t_k\}$ と定義する。サービスが持つファンクションの集合 F を $F = \{f_1, f_2, \dots, f_l\}$ と定義する。また、ファンクションの実行結果を $R = \{r_1, r_2, \dots, r_m\}$ と表す。これらを用いて、ユーザの意図 I を、 $I = (T, F, R)$ と定義する。ここで、意図推定関数を g で表すと、クエリの意図推定を、

$$I = g(Q)$$

と表すことができる。

図 2 に意図サービス推定のモデルを示す。ここで、クエリ Q の構成語 $q_1, q_2, \dots, q_n$ を T, F, R にマッピングできれば、その T, F, R から構成される I はそのクエリ Q から推定されるユーザの意図であると仮定する。この仮定が成り立つか検証を行うため、クエリ Q とサービスがもつ T, F, R の関係を分析する予備調査を行う。

4. 予備調査

この節では、クエリ Q とサービスがもつ (T, F, R) の関係を分析するために予備調査を行う。クエリとクエリから推定されるサービスを結びつけるものとして、

本研究では検索連動型広告に着目する。

4.1. 検索連動型広告

検索連動型広告とは、広告主が検索サービスに対してキーワードを購入することで、ユーザがそのキーワードで検索を行った時に広告を提示する仕組みである。提示される広告のリンク先ではサービスが提供されていることが多い。さらに、広告主はサービスを利用してもらう機会を増やすために、サービスに関係があり、かつ、検索時に入力される可能性が高いと考えられるキーワードを購入していることが多い。この仕組みを利用し、クエリ Q と検索連動型広告のリンク先のサービスが持つ (T, F, R) の関係を調べる。そこに関係性があれば、3.2.節で述べた仮定に基づき意図サービス推定のアルゴリズムを提案することができると言える。なお、今回の予備調査には、検索連動型広告として Google AdWords と Yahoo! スポンサーサーチの二つを用いる。

4.2. 予備調査の内容

クエリとサービスの関係を調べるために、検索連動型広告のリンク先のサービスが持つ「サービスの種類」「ファンクション」「ファンクションの実行結果」を調査する。この調査は手動で行い、通販や交通機関予約、旅行予約を意図した 32 種類のクエリを用いた。なお、得られた検索連動型広告の総数は 385 件であった。

ここで、「サービスの種類」「ファンクション」「ファンクションの実行結果」を、それぞれ Web ページのどこから抽出してきたかを説明する。「サービスの種類」は、主にサービス提供者の会社概要ページの事業内容や広告タイトルなどに含まれる語から抽出した。例えば、旅行業や EC 事業などがこれに当たる。「ファンクション」は、Web ページが持つ入力フォームの周辺テキストから、その入力フォームがどのような機能を持

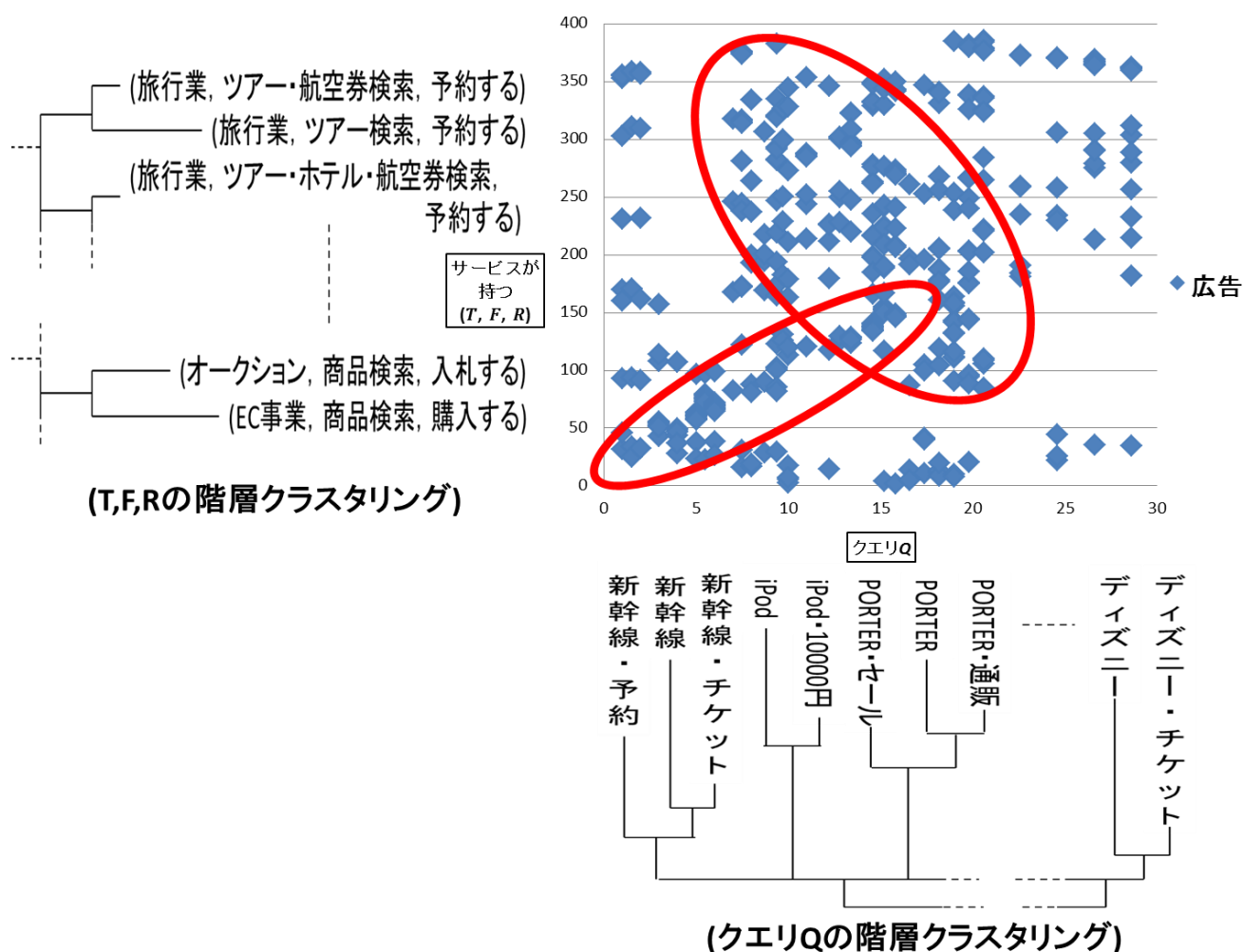


図 3: Q と (T, F, R) の相関関係

っているかを抽出する．例えば，空室検索や商品検索などがこれに当たる．「ファンクションの実行結果」に関しては，実際にフォームに何らかの条件を入力したあとにさらに必要となる操作，もしくはユーザが得る結果を調べる．例えば，“プランを選び予約を行う”や“商品を選んで購入する”などがこれに当たる．

4.3. 予備調査で得たデータの例

予備調査で得られたデータの一例として，「京都」をクエリとした場合の検索連動型広告のリンク先のサービスが持つ (T, F, R) を調査した結果の一部を表 1 に，「京都 東京」をクエリとした場合の結果の一部を表 2 に示す．

表 1 をみると，「京都」で検索した場合の検索連動型広告のリンク先のサービスの種類は，全体に占める割合では旅行業に関するものが多いが，歯医者や賃貸業などのサービスの広告も提示されていることが分かる．これは，「京都」というクエリで検索した場合に検索連動型広告で得られるサービスが，主に京都という地域内で提供されるものであることが原因であると考

えられる．

一方，表 2 の「京都 東京」で検索した場合の検索連動型広告のリンク先のサービスの種類は，旅行業やバス事業など，旅行関係のものが主となっている．これは，「京都 東京」というクエリで検索した場合に検索連動型広告で得られるサービスが，主に京都と東京の二地点間で提供されるものであることが原因であると考えられる．

これらの例から，クエリに一語を追加するだけでも，それらのクエリから推定されるサービスの種類には違いが生じるということが分かる．このようなデータを用いて，クエリとサービスが持つ (T, F, R) の関係を分析する．

5. Q と (T, F, R) の相関関係の分析

この節では，予備調査で得られたデータを利用して，クエリ Q とサービスの種類 T・ファンクション F・ファンクションの実行結果 R の相関関係の分析を行う．その準備として，予備調査に用いたクエリ Q と検索連動型広告が持つ (T, F, R) についてそれぞれ階層クラス

タ分析を行う。

5.1. クエリ Q の階層クラスタ分析

まず、クエリ Q に関して、各クエリを構成する各キーワードを要素とするようなキーワードベクトルを作成する。このキーワードベクトルを利用し、階層クラスタリングを行う。

5.2. (T, F, R) の階層クラスタ分析

次に、各検索連動型広告のリンク先のサービスが持つ (T, F, R) に関して階層クラスタ分析を行う。ここでは、各広告に対して、各広告が持つすべての (T, F, R) を索引語とするような特徴ベクトルを作成する。その特徴ベクトルを利用し階層クラスタリングを行うことによって広告のリンク先サービスが持つ (T, F, R) の分析を行う。

5.3. Q と (T, F, R) の相関関係

5.1.節で得られたクエリの分析結果と、5.2.節で得られた (T, F, R) の分析結果を用いて作成した Q と (T, F, R) の相関マップを図3に示す。クエリ Q の階層クラスタリングの結果を横軸に表し、検索連動型広告のリンク先のサービスが持つ (T, F, R) の階層クラスタリングの結果を縦軸に表している。

図3において、プロットされた点が密集している部分について、 Q と (T, F, R) の間に相関があると言える。また、相関係数の値は約 0.31 となっていた。

6. 考察

今回の予備調査は、4.2.節で述べたように手動で行ったが、これを自動化しデータの数を増やす必要がある。その際、 (T, F, R) のうち、 T と F については今回手動で調査したプロセスをそのまま自動化することが可能だと考えられる。一方、 R に関しては、今回のプロセスをそのまま自動化するのは難しいので、自動化のための工夫が必要である。例えば、Web サービスページや検索連動型広告などのタイトル中に含まれる“予約”や“購入”というような動詞に注目することで、ユーザがそのサービスを利用し、最終的に得られる結果を判別することができるのではないかと考えられる。このようなプロセスならば自動化することも可能である。

分析結果については、5.3.節で述べたように、クエリ Q とサービスが持つ (T, F, R) の間には相関が認められた。この結果から、 Q から (T, F, R) を推定することが可能だということが分かった。

しかし、どのようなクエリがどのような (T, F, R) に結びつくのかを具体的に推定するアルゴリズムの提案は、今後の課題である。このようなアルゴリズムを考える際には、図3においてプロット点が密集している部分に注目することが重要である。例えば、今回の予備調査では、ブランド名や商品名が含まれているクエ

リは、“EC 事業に関して、商品検索をして、商品を選んで購入する”という (T, F, R) に結びつくということが分かった。これは図3の相関マップにおいて、左下部の赤枠に囲まれた部分の例である。このようなデータを学習データとして機械学習をすることによって、新しい入力クエリから (T, F, R) を推定することができると考えられる。

7. まとめ

本論文では、Web サービス検索において、ユーザが入力した検索クエリから、実行可能なサービスインスタンスを得られるような技術の基本概念を提案した。さらに、その構成技術の第一段階であるクエリの意図サービス推定の手法を提案し、検索連動型広告を用いて予備調査を行い、その手法の有用性を示した。

今後の課題は、今回行った予備調査を自動化しデータの数を増やすことである。そのデータを学習データとして機械学習を行うことにより、新しい入力クエリから (T, F, R) を推定することが可能になると考えられる。また、3.2.節で述べたような、クエリ Q の構成語を T, F, R にマッピングする手法についても考えなければならない。

さらに、この意図サービス推定の応用として、1.1.節で述べたような実行可能サービスインスタンス検索の実現を目指す。そのためには、構成技術の第二段階である自動フォーム入力について検討する必要がある。なお、この分野は、Deep Web に関する研究[5][6]が関連すると考えられる。

謝辞 本研究の一部は、グローバル COE 拠点形成プログラム「知識循環社会のための情報学教育研究拠点」（拠点リーダー：田中克己）によるものです。ここに記して謝意を表します。

参考文献

- [1] Broder, A. : A Taxonomy of Web Search. SIGIR Forum, pages 3-10, 36(2), (2002).
- [2] In-Ho Kang. Transactional query identification in web search. In AIRS, 2005.
- [3] Jian Hu, Gang Wang, Fred Lochovsky, Jiantao Sun, and Zheng Chen. 2009. Understanding user's query intent with Wikipedia. In Proceedings of WWW-09, pages 471-480.
- [4] A. Sugiura and O. Etzioni. Query routing for web search engines: Architecture and experiments. In Proceedings of the Ninth Int'l World Wide Web Conf., 2000.
- [5] J. Madhavan, D. Ko, L. Kot, V. Ganapathy, A. Rasmussen, and A. Halevy. Google's Deep-Web Crawl. PVLDB, 1(2):1241-1252, 2008.
- [6] F. Liu, C. T. Yu, W. Meng, and A. Chowdhury. Effective keyword search in relational databases. In SIGMOD, pages 563-574, 2006.