

# Web サービス検索のための操作カテゴリ分類手法の提案

片淵聡<sup>†a)</sup> 石川冬樹<sup>††</sup> 鄭顕志<sup>†</sup> 高橋竜一<sup>†</sup> 深澤良彰<sup>†</sup> 本位田真一<sup>††</sup>

<sup>†</sup>早稲田大学大学院基幹理工学研究科情報理工専攻 〒169-8555 東京都新宿区大久保 3-4-1

<sup>a)</sup> Email: s.katafuchi@fuka.info.waseda.ac.jp

<sup>††</sup>国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

## 1. はじめに

Web サービスとは、XML やHTTP などのインターネット標準技術を用いてネットワーク経由で利用可能なWeb アプリケーションであり、情報を取得するのによく用いられる。Web サービスは機能の集合体であり、各機能はプログラムの部品として組み込むことができる。アプリケーションの開発において、処理の一部をWeb サービスによって実現することで、システムの開発コストを低減することが可能である。

プログラマがWeb サービスを利用する際は、インターネット上に公開されているWeb サービスの中から要求するWeb サービスを人手によって検索する必要がある。検索において、有効な方法にカテゴリ分類がある。しかし、既存のカテゴリ分類手法では、操作が異なる機能やシステムの振舞いを修正しなければならないような機能の篩い分けができないので、結果として検索コストやシステムの振舞いの修正コストが増大してしまう。本論文では、Web サービスの機能の操作内容に対するカテゴリ分類手法を提案する。

## 2. 既存手法

現在、膨大な数のWeb サービスが様々なサービス提供者によってXmethods[1]などのWeb サービスデータベースに登録されている。要求するWeb サービスを効率よく検索するための有効な手段としてカテゴリ分類が挙げられる。

ASSAM[2]などでは、サービスをドメインカテゴリに自動分類するための手法を提案している。

ドメインカテゴリ分類を用いた検索では、対象ドメインごとにWeb サービスを分類することで、アプリケーションドメインによる効果的なWeb サービスの絞り込みが可能となる。

しかし、既存のドメインカテゴリ分類では、Web サービス(ドメイン)の分類を行っており、Web サービスの構成要素となる1つ1つの機能(操作)の分類ができないので、より粒度の細かい機能に関しては効率的な絞り込みができないという問題がある。

また、既存手法では、入出力のデータ構造による分類ができないので、入出力のデータ構造が要求と異なり、システムを実現するために振舞いを修正しなければならないような機能を検索から弾くことができないという問題がある。

表1: カテゴリと操作種別名との対応表

操作種別	入力	参照
``検索``	あり	あり
``情報取得``	なし	あり
``変換``	あり	なし
該当無し	なし	なし

## 3. 提案手法

本研究では、操作内容のカテゴリ分類手法を提案する。まず、組込む機能によってはシステムの振舞い修正コストが増えてしまうという問題を解決するために、本研究では、入出力のデータ構造が異なる機能をカテゴリによって区別する。入出力のデータ構造が異なる機能を分類することで、組み込む際にシステムの振舞いを修正する必要が出てくる機能を検索結果から弾くことを目的とする。そのために本論文では、入力の数と出力の数による分類を行う。具体的には、「単一」の値か「複数」の値かによって分類する。結果として、Web サービスを組み込んだ際におけるシステムの振舞いの修正コストを低減することが可能となる。

次に、操作内容に関する絞り込みができないという既存手法の問題点を解決するために、操作内容に対してのカテゴリ分類手法を提案する。操作内容によって分類することでWeb サービスの検索コストの低減を目的とする。本論文では、``検索``、``情報取得``、``変換``の3つの操作種別にWeb サービスの機能を分類する。

- ``検索``: 「Web サービスが持つ情報に対して制限をかけてその一部を出力として取得する」
- ``情報取得``: 「Web サービスが持つ情報に対して制限をかけずにそのまま出力として取得する」
- ``変換``: 「入力の値に対して変換処理を行う」

Web サービスの機能をこれら3種類に分類するために、その機能の入力と参照データの有無によって分類することを提案する。参照データとは、Web サービスが持っている情報のことを指し、``検索``・``情報取得``のグループと``変換``のグループを区別するために用いる。表1に、操作種別名に対する入力及び参照データの有無の対応を示す。参照データの有無によって``検索``・``情

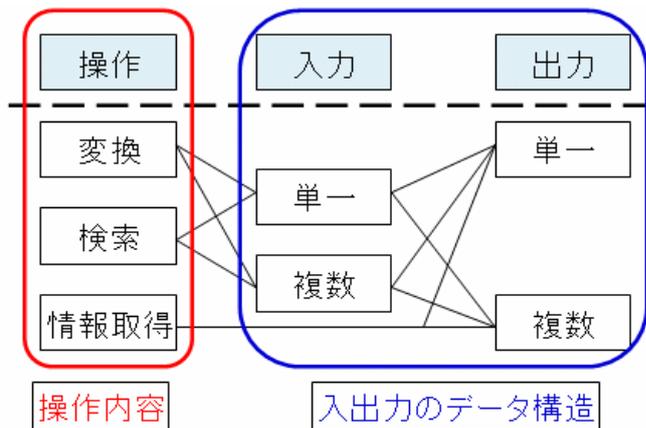


図 1 : 提案手法の最終モデル

報取得'のグループと'変換'のグループを分類し、入力の有無によって'検索'と'情報取得'を分類する。入力も参照データも用いずに出力だけ返す機能は、サービスとしての役割を果たしていないので「入力なし・参照データなし」のカテゴリに分類される機能は存在しない。結果として操作種別によって分類することにより、Web サービス利用者は要求するサービスの機能を、直感的な操作名によって効率的に検索することが可能となる。

提案手法の最終的なモデルは図 1 に示すとおりになる。

#### 4. 評価・考察

検索の絞り込み精度を、既存手法のみの検索と、提案手法も併せた検索とで比較評価する。具体的な指標としてAccuracy(不正解であるゴミの少なさを表す指標)とRecall(正解の取り逃がしの少なさを表す指標)、そしてAccuracyとRecallの総合評価するための指標としてF値を用いる。F値は以下の式で表すことができる。

$$F = \frac{2}{\frac{1}{Accuracy} + \frac{1}{Recall}} \quad (1)$$

評価には、「操作種別に偏りが無い」という性質を持つ金融ドメインに属する25個の機能と、「'検索'カテゴリに操作が偏っている」という性質を持つ天気ドメインに属する35個の機能を用意し、各々6つの要求に対してAccuracyとRecallを測り、その平均を取った。

表2と表3に、金融及び天気ドメインにおける、各要求に対するAccuracyとRecallの平均、及びそれらの値から求めたF値を示す。表2と表3から、両ドメインにおいて提案手法を用いたことで、Accuracyが11~24%高くなったことがわかる。しかし、Recallは既存手法のみと比べて減少した。これは、「convertToDollars(fromCountry,value)」と

表 2 : 金融ドメインにおける Accuracy,Recall,F 値

手法	Accuracy	Recall	F値
既存	0.087	1.000	0.160
既存+提案	0.325	0.875	0.474

表 3 : 天気ドメインにおける Accuracy,Recall,F 値

手法	Accuracy	Recall	F値
既存	0.033	1.000	0.060
既存+提案	0.137	0.916	0.238

「convertCurrency(fromCountry,toCountry,value)」のように、出力の数が違っていても共にプログラマの要求を満たす場合が存在し、提案手法によって片方を検索結果から弾いてしまったからである。それでも、検索の総合評価値であるF値に関しては、提案手法を用いなかった場合と比べて、約0.18~0.31高くなったという結果が得られた。この結果から、提案手法を用いることで検索の絞り込み精度が向上することがわかる。

また、金融ドメインと天気ドメインを比較した時、「操作種別に偏りが無い」金融ドメインの方が、「'検索'カテゴリに操作が偏っている」天気ドメインよりAccuracy及びF値が高いという結果が得られた。これは、「'検索'カテゴリで検索してもほとんど機能が絞り込まれない天気ドメインと違って、金融ドメインではどの操作種別においても効果的に絞り込みが可能だからである。

#### 5. まとめ

本研究では、操作内容を分類するためのカテゴリを提案した。提案手法では、3種類の操作種別と入出力の数によってWebサービスの機能の分類を行った。操作内容によって分類することで、プログラマが要求しないWebサービスを検索から弾き、結果としてプログラマが要求するWebサービスの機能を効率よく探し出すことができる。

今後の課題としては、操作種別による分類の有効性を向上するために、「フィルタリング」や「ソーティング」といった、より詳細な操作種別の分類を行うことがまず挙げられる。また、「購入」や「予約」といった出力は返さないが副次的な「効果」をもたらすサービスについての分類を行うことも課題として挙げられる。

#### 参考文献

- [1] XMethods. <http://www.xmethods.net>
- [2] Andreas Hes, Eddie Johnston and Nicholas Kushmerick.  
 "ASSAM: A Tool for Semi-automatically Annotating Semantic Web Services"  
 International Semantic Web Conference, 2004.