

関連性を考慮した WebAPI 検索方式の検討

WebAPI Search Schemes Based on Relativity Information

井上 紘希† Koki Inoue
 小坂 隆浩‡ Takahiro Koita

1. はじめに

1.1 背景

インターネットの普及に伴い、様々な Web サービスへのアクセスが可能になった。その中で、Web ビジネスを展開する企業が情報の共有化による Web サービスの向上を目的に WebAPI を公開するようになった。WebAPI が公開されることで、WebAPI を利用して多くの Web サービスの中から、任意の Web サービス同士を組み合わせる新しい 1 つの Web サービスを作り出すマッシュアップという技術が登場し、注目を集めている。マッシュアップを用いることで、開発環境の構築コストやサービスの開発コストを抑えて、新たな Web サービスを開発することができる。その有用性から、現在では、多くの Web サービスでマッシュアップが用いられている。マッシュアップによって Web サービスは新しい価値を持つことが可能となる。このためマッシュアップが盛んに行われるようになり、マッシュアップは広く利用されるようになった。マッシュアップの例を図 1 に示す。例として挙げているのは、Google Maps¹⁾と食べログの WebAPI を組み合わせ、地図上にお店の情報を表示するというものである。このようにマッシュアップを行うことで、異なる Web サービスの機能を組み合わせるサービスを作ることが可能になる。



図 1 マッシュアップの例

1.2 目的

本研究では、WebAPI 検索を対象とし、ユーザが組み合わせる WebAPI を容易に検索できる環境の実現を目的とする。WebAPI 検索とは、マッシュアップを行う際に伴う作業で、ユーザが望んでいる機能を持った Web サービスの WebAPI をインターネット上から検索する作業のことである。しかし、インターネット上に数多く存在する WebAPI の中から、ユーザが仕様を理解し、目的に合う WebAPI の組み合わせを探し出すことは、WebAPI の入出力に関する情報の組み合わせを考慮しながら検索する必要があり、ユーザの負担となる。さらに、WebAPI の関連性を示す情報がないため、マッシュアップに利用するユーザの目的に合う WebAPI を数多くの WebAPI の中から選択した後、ユーザの目的に合い、かつ入出力情

報が一致するものを、もう一度数多くの WebAPI の中から検索する必要があり、WebAPI の組み合わせを検索することは困難な作業となっている。

WebAPI 検索は今後の WebAPI の増加に伴いさらに困難な作業となり、マッシュアップを行うユーザの大きな負担となっていく。ユーザの負担を少しでも減らすためには、WebAPI の関連性を取得し、組み合わせ検索を行う対象となる WebAPI を絞ることが必要になる。本研究では、目的の実現に必要な WebAPI の関連性を取得するため、既存の WebAPI 検索手法の現状分析を行い、関連性を考慮した WebAPI 検索手法を検討する。

2. 既存の WebAPI 検索手法

本研究での分析対象となる、既存の WebAPI 検索手法と、その検索手法が抱える問題について述べる。WebAPI 検索手法は大きく分けて検索エンジンを使用した検索と WebAPI 情報ポータルサイトを使用した検索の 2 つがある。検索エンジンを使用した検索は検索の範囲が広く、WebAPI 発見までの手順が明確ではないので、情報の精度やユーザの負担の面から、本研究では、WebAPI 検索に特化した、WebAPI 情報ポータルサイトを使用した検索手法を既存の検索手法とする。

2.1 ProgrammableWeb

ProgrammableWeb²⁾は、WebAPI の検索のみに特化した、世界最大の WebAPI 情報ポータルサイトである。ユーザは ProgrammableWeb を利用することで、検索エンジンを用いて WebAPI を検索する場合に比べ、WebAPI の検索を行うことが容易となった。そのため、ユーザは WebAPI 検索を行う際、積極的に ProgrammableWeb を用いている。

ProgrammableWeb に登録されている WebAPI 情報には、カテゴリや企業情報、WebAPI 名がタグ付けされており、マッシュアップ情報には開発者名やマッシュアップに利用された WebAPI などがタグ付けされている。WebAPI に関する情報を検索する際には、キーワードや登録されているタグ情報をもとにして検索が可能である。WebAPI に関する情報は WebAPI を提供している企業が任意で記載する。更に、WebAPI を提供している企業のサイトの URL も登録されているため、WebAPI を利用可

Google Maps: Highlights	
Summary	Mapping services
Category	Mapping
Tags	mapping places viewer display
Protocols	JavaScript
Data Formats	XML, VML, JSON, KML
API home	https://developers.google.com/maps/
This is 1 of	107 APIs by Google

図 2 ProgrammableWeb の情報例

†同志社大学大学院 理工学研究科, Graduate School of Science and Engineering, Doshisha University

‡同志社大学 理工学部, Faculty of Science and Engineering, Doshisha University

録されている全ての WebAPI の平均で見ると、接続数は約 1.2 個となる。この結果は、ProgrammableWeb 上に隠れた情報であり、ProgrammableWeb に登録されている WebAPI の関連性の偏りと、登録されている WebAPI 全体の関連性が希薄であることを示している。

3.2 WebAPI の媒介中心性

図 4 の WebAPI の関連性のネットワークを対象に、各 WebAPI について、媒介中心性を計算した。媒介中心性とは、ネットワークにおいてそのノードを通過しないと他のノードに到達できない度合、つまり、ある点がある他の 2 点を結ぶ最短経路である度合である。これにより、接続している WebAPI の数と、WebAPI の関連性のネットワークにおける重要性の 2 つを考慮した、WebAPI の関連性の強さを計算することができる。得られた媒介中心性の結果と、ProgrammableWeb に登録されているマッシュアップ数上位 10 個とを比較したものを表 2 に示す。

表 2 マッシュアップ登録数と媒介中心性の比較

順位	ProgrammableWeb マッシュアップ登録数	媒介中心性
1	Google Maps	Google Maps
2	Twitter	Twitter
3	Youtube	Youtube
4	Flickr	Facebook
5	Amazon PA	Flickr
6	Facebook	Amazon PA
7	Twilio	PayPal
8	Last.FM	Last.FM
9	Ebay	Twilio
10	Google Search	Ebay

この結果で注目すべき点は、媒介中心性で第 7 位にランクインしている PayPal⁴⁾ という WebAPI である。PayPal は、ProgrammableWeb のマッシュアップ登録数は 37 個で、マッシュアップ登録数ランキングでは 59 位の WebAPI である。この結果により、Programmable 上のマッシュアップ登録数によるランキングの表示だけでは、WebAPI の関連性において重要な WebAPI が、人気の高い WebAPI に埋もれてしまい、検索することが困難な状態を示している。

3.3 WebAPI の近接中心性

図 4 のグラフデータを対象に、各 WebAPI について、近接中心性を計算した。近接中心性とは、ネットワークにおいて、そのノードが接続している他のノードにどれだけ近い距離にあるのかを表す度合である。近接中心性を計算することによって、接続している WebAPI との関連がどれだけ強いのかを計算することができる。得られた近接中心性の結果と、ProgrammableWeb に登録されているマッシュアップ数上位 10 個とを比較したものを表 3 に示す。表 3 で注目すべき点は、Gephi で計算した近接中心性で第 8 位にランクインしている del.icio.us⁵⁾ という

表 3 マッシュアップ登録数と近接中心性の比較

順位	ProgrammableWeb マッシュアップ登録数	近接中心性
1	Google Maps	Google Maps
2	Twitter	Twitter
3	Youtube	Youtube
4	Flickr	Flickr
5	Amazon PA	Facebook
6	Facebook	Amazon PA
7	Twilio	Last.FM
8	Last.FM	Del.icio.us
9	Ebay	Ebay
10	Google Search	Google Search

WebAPI と、ProgrammableWeb のマッシュアップ登録数で第 7 位にランクインしている Twilio⁶⁾ という WebAPI である。del.icio.us は ProgrammableWeb のマッシュアップ登録数第 12 位の WebAPI である。近接中心性を計算したことによって、少しではあるが、関連性を持つ WebAPI との関連性の強い WebAPI を掘り起こすことに成功している。また、Twilio の近接中心性の順位は第 30 位である。この結果により、ProgrammableWeb のマッシュアップ登録数によるランキングの表示だけでは、強い関連性で結びついている WebAPI を検索することが困難な状態になっていることを示している。

3.4 WebAPI のグループ分け

図 4 の WebAPI の関連性のネットワークを対象にモジュラリティを計算し、類似した性質をもった WebAPI をグループ化した。モジュラリティとは、ネットワークやグラフの解析に用いられる効果関数で、ネットワークからモジュールやコミュニティへの分割の質を定量化するものである。これにより分類したグループの内、最も大きな 4 つのグループを図 5 に示す。

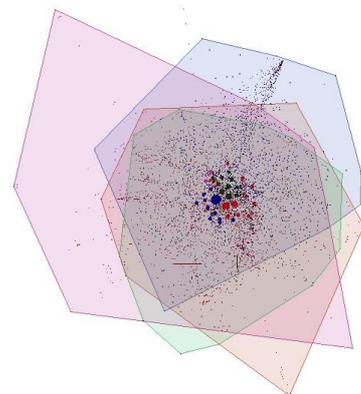


図 5 WebAPI のグループ化

直線で囲まれた部分が分類されたグループを表している。このグループは、関連性が強い WebAPI 群をまとめたもので、マッシュアップに使用されている WebAPI の組み合わせの内、一方を使用した他の WebAPI との組み合わせで使用されている WebAPI ような、間接的な関連

性を持つ WebAPI も含まれている。この結果、WebAPI を 32 個のグループに分類することができた(その内上位 4 個のグループだけで総 WebAPI 数の 80%以上を占めている)。分類したグループ毎に、WebAPI の接続数、媒介中心性、近接中心性を計算することができるので、より細かく WebAPI の関連性を計算することが可能である。分類したグループの内、最も大きなグループである、Google maps を中心としたグループの詳細をまとめたものを表 4 に示す。

表 4 Google maps 中心グループの詳細表

グループ内ノード数	351(全体の 27%)	
グループ内接続数	2214(全体の 16%)	
WebAPI 名	接続数	順位
Google maps	500	1
GeoNames	176	2
Google AdSense	164	3
グループ内平均	12.6	
WebAPI 名	媒介中心性	順位
Google maps	0.579	1
Google App Engine	0.054	2
Google Homepage	0.05	3
グループ内平均	0.004	
WebAPI 名	近接中心性	順位
Google maps	0.769	1
Google AdSense	0.549	2
GeoNames	0.543	3
グループ内平均	0.426	

表 4 の媒介中心性で第 3 位の Google Homepage⁷⁾は、接続数が 64 個でグループ内第 42 位の WebAPI であり、接続数だけではわからない、WebAPI の関連性において重要な WebAPI を掘り起こせている。また、このグループ内では、接続数と近接中心性の順位はほぼ等しい結果となった。

ProgrammableWeb 上に存在する、カテゴリやタグによる分類に加えて、関連するグループの要素を追加すれば、さらに関連性のある WebAPI の分類が可能になる。これもまた、ProgrammableWeb を分析しなければ出てこない、隠れた情報である。

4. 検索手法の検討

ProgrammableWeb を分析した結果、2 つの問題点が明らかになった。

1 つ目の問題点は、WebAPI の関連性の数に偏りがあり、WebAPI 全体の関連性が希薄になっているということである。この問題を解決するためには、ProgrammableWeb に登録されているマッシュアップでの WebAPI の組み合わせ以外の情報を付加することが必要である。例えば、登録されたマッシュアップ情報と分析によるグループ分けの関連性の取得で、WebAPI 全体の関連性の質を向上させることで解決が可能である。

2 つ目の問題点は、人気の高い WebAPI に隠れて、WebAPI の関連性において重要、または強い関連性を持つ WebAPI が埋もれてしまい、検索が困難になっているということである。この問題を解決するためには、Gephi で計算した WebAPI の媒介中心性と近接中心性を考慮した、WebAPI の関連性の取得が必要である。登録されている WebAPI の組み合わせと、WebAPI の媒介中心性・近接中心性を用いて重み付けをして関連性を取得することで、解決が可能である。

本研究で明らかにした ProgrammableWeb の問題点を考慮し、ProgrammableWeb 上に存在するマッシュアップやタグの情報と、分析により得られる媒介中心性、近接中心性、グループの情報を組み合わせた、WebAPI 推薦システムを提案する。キーワードとして与えられた WebAPI やタグを基にして、それに高い関連性を持つ WebAPI の組み合わせを自動で検索し推薦するシステムが実現できれば、容易な WebAPI の組み合わせ検索を支援することが可能になる。推薦システムのアルゴリズムの中に、Gephi での分析結果を反映させることで、ProgrammableWeb 上の隠れた情報を考慮した推薦を行うことができ、推薦結果の質を向上させることができる。今後、推薦システムの実装に関しては、Wenxing Xu⁸⁾らの研究や、Buqing Cao⁹⁾らの研究を参考にして進めていく予定である。

5. まとめ

本研究では、マッシュアップにおける WebAPI の組み合わせ検索を容易に行える環境の実現を目的として、既存の WebAPI 検索手法である、ProgrammableWeb の現状分析を行った。分析の結果、WebAPI の関連性に偏りがあり、全体的に関連性が希薄であることと、人気の高い WebAPI に重要な関連性を持つ WebAPI が隠れてしまっているという問題点を明らかにした。本研究で明らかにした問題点を解決するために、ProgrammableWeb と Gephi による分析で得られた情報を組み合わせた推薦システムの実装を行うことが、今後の課題となる。

文献

- [1] Google Maps: <https://developers.google.com/maps/>.
- [2] Programmable web: <http://www.programmableweb.com/>.
- [3] Gephi: <http://oss.infoscience.co.jp/gephi/gephi.org/>
- [4] PayPal: <https://developer.paypal.com/>
- [5] Del.icio.us: <https://delicious.com/developers#title1>
- [6] Twilio: <http://www.twilio.com>
- [7] Google Homepage: <https://developers.google.com/gadgets/>
- [8] Wenxing Xu, Jian Cao, Liang Hu, Jie Wang, Minglu Li: A Social-Aware Service Recommendation Approach for Mashup Creation, Web Services (ICWS), 2013 IEEE 20th International Conference on, pp.107-114, 2013.
- [9] Buqing Cao, Jianxun Liu, Mingdong Tang, Zibin Zheng, Guangrong Wang: Mashup Service Recommendation Based on User Interest and Social Network, Web Services (ICWS), 2013 IEEE 20th International Conference on, pp.99-106, 2013.