

知識ベースに対するプロパティ指向のファセット検索システムに関する研究

阿曾 太郎[†] 天笠 俊之^{††} 北川 博之^{††}

[†] 筑波大学システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1

^{††} 筑波大学計算科学研究センター 〒305-8573 茨城県つくば市天王台 1-1-1

E-mail: [†]aso@kde.cs.tsukuba.ac.jp, ^{††}{amagasa,kitagawa}@cs.tsukuba.ac.jp

あらまし 知識ベースとは様々な知識が蓄積されたデータベースである。本研究では、その中でも RDF で記述された知識ベースに焦点を当てる。一般に、RDF による知識ベースのデータ構造は複雑であるため、専門知識を持たないユーザが簡単に検索を行うには、ファセット検索が有効であることが知られている。ファセットとはデータの切り口であり、ユーザはファセットの選択・解除を繰り返すことで対話的な検索を実行できる。本論文では、RDF の述語（プロパティ）について、主語と述語との構造的な関係に着目してクラスタリングすることで、プロパティに関するファセットを生成することを提案する。

キーワード 知識ベース, RDF, ファセット検索, クラスタリング

1 はじめに

知識ベースとは、様々な知識が蓄積されたデータベースである。そこには、世界に関する汎用的な知識から専門的な知識までが蓄積されている。代表的な知識ベースには、Wikipedia の情報を基にした DBpedia, Wikidata, YAGO などがある。人間や機械は知識ベースを使うことで、質問に答えたり、新たな知識を発見することができる。

知識ベースの記述には、Resource Description Framework (RDF) が用いられる。RDF とは、リソースに関する情報を記述する方法である。RDF では、Universal Resource Identifier (URI) で識別されるものすべてをリソースとして扱う。世の中のあらゆるエンティティは URI を付けることで、リソースとして記述することができる。あるリソースについての 1 つの情報は、主語 (Subject)、述語 (Predicate)、目的語 (Object) から構成される 3 つ組 (トリプル) のグラフ構造で記述される (図 1)。主語は情報を記述される対象のリソースを示し、述語は主語に関する情報のプロパティを定義する。そして、目的語には述語の対象となる値を格納する。主語と述語は URI で記述し、目的語は URI もしくは数値や文字列などのリテラルで記述する。

RDF で記述された知識ベースに対して検索を行うにはいくつかの方法がある。主要なもの 1 つは問合せ言語 SPARQL を用いた検索である。SPARQL の文法に従って、トリプルの条件を指定することで、情報を取り出すことができる。しかし、一般ユーザにとって、SPARQL 検索を行うハードルは高い。なぜなら、まず、SPARQL の文法を理解し習得する必要がある。そして、検索対象の知識ベースで定義されているプロパティやエンティティについて理解する必要があるためである。知識ベースは様々な種類のプロパティやエンティティが存在す

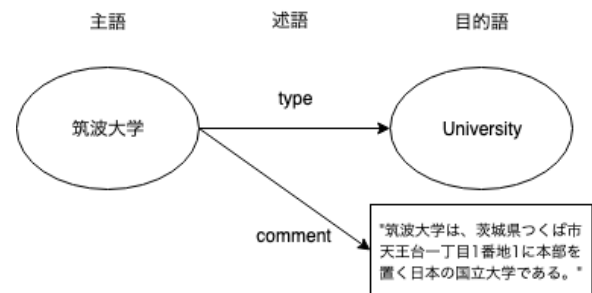


図 1 RDF の例

る複雑なグラフ構造になっているため、特に後者の理解は難しい。もう 1 つの方法として、キーワード検索がある。キーワード検索は、キーワードを入力することで検索が行えるため、前提知識が不要な検索手法である。結果には、キーワードに関するエンティティのランキングが返却される。しかし、様々な種類のエンティティが混在しているため、ユーザは欲しい情報を判断しづらい。また、ユーザはどのような種類のエンティティがあるのか把握していない場合や、そもそもどのような種類のエンティティが欲しいのかわかっていない場合もある。したがって、前提知識が不要という点で、キーワード検索は有効な手段だが、情報を整理して提示する必要があると考える。

こうした課題を解決する検索方法としてファセット検索がある。ファセット検索では、検索対象のエンティティを様々な切り口（ファセット）で絞り込む。ユーザは結果を確認し、ファセットを切り替えたり、組み合わせたりすることで、意図する結果を得るまで、対話的に検索を行うことができる。したがって、知識ベースのプロパティやエンティティの種類などの知識を持たない場合でも、検索を容易に実行できる。これまでに提案されてきた RDF の知識ベースに対するファセット検索システムの多くは、エンティティが持つプロパティをファセットと

して利用してきた。しかし、プロパティそれ自体も数多くの種類が存在しているため、必要となるプロパティ（ファセット）を見つけ出すことは簡単ではない。この課題を解決するには、数多くあるプロパティから興味のあるプロパティを見つけ易くすることが必要である。そのために、プロパティをその主語と目的語との関係性によってクラスタリングした結果をプロパティのファセットとして利用することを提案する。そして、興味のあるプロパティで関係付けられたエンティティ（トリプル）を検索できる、プロパティ指向の新しいファセット検索システムの開発を提案する。

本稿では、プロパティ指向のファセット検索システムの全体概要と、その機能要素であるプロパティファセットのためのプロパティのクラスタリングについて報告する。

2 前提知識

本節では、前提知識として、ファセット検索について説明する。

2.1 ファセット検索

ファセット検索とは、探索的検索における手法の1つである。その特徴は、検索対象のエンティティ集合を、様々な切り口（ファセット）によって絞り込むことで、意図するエンティティを発見しようとするところにある。例えば、E-CommerceのAmazonの商品検索サイトもファセット検索である。基本的なインターフェースとして図2を例にあげる。ユーザはキーワード検索によって、Extensionに初期の検索結果を、Transition Markerに初期のファセットを得る。その状態から、ファセットの選択や、選択したファセットの解除を対話的に繰り返して、検索結果を洗練させていく。

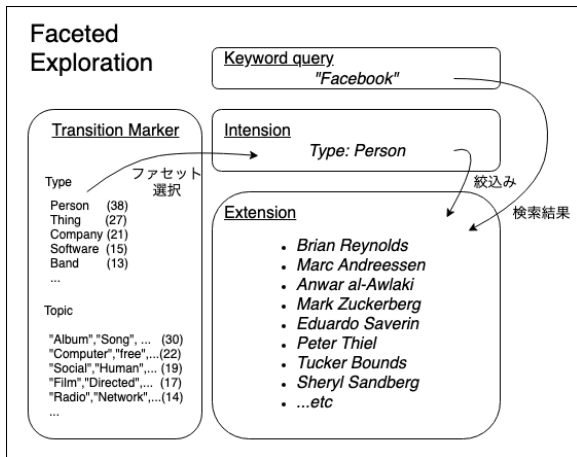


図2 ファセット検索のインターフェースイメージ

3 関連研究

3.1 RDFの検索結果に対するランキング

RDFの検索結果の有用性を向上させることを目的に、Ichinoseら[4]はDBpediaに対してPageRankを用いてエンティティを評価する手法を提案している。主語と目的語をノードと

し、述語をエッジとしてPageRankを計算し、エンティティの重要度を評価している。この研究の実験では、検索対象のエンティティが既知、あるいはSPARQLクエリにおいてトリプルの条件が設定されており、対象ユーザはSPARQLや対象データのDBpediaに関する知識を持っていることを前提としている。本研究では、エンティティの評価にはPageRankを用いるが、ユーザはSPARQLや知識ベースに関する知識を有していないケースを想定しているため、前提条件が異なる。

3.2 RDFに対するキーワード検索

RDFに対するキーワード検索を行う研究として、奥村ら[6]のObjectRankと適合フィードバックを用いた研究がある。検索対象となるエンティティはリテラルを目的語に持つトリプルを持つことを前提にして、エンティティをリテラルを含めたドキュメントとみなしている。本研究でもこの考え方に則る。しかし、検索対象となるエンティティの種類がシステム構築者などによってあらかじめ設定されることを前提とした手法であるため、知識ベースに含まれる多様なエンティティの全てに応える方法ではない。また、ユーザの検索意図に応える方法として、適合フィードバックを採用している点も、ファセット検索とはアプローチが異なる。

3.3 知識ベースに対するファセット検索

知識ベースに対するファセット検索の研究は数多く行われている。本稿の対象データセットであるDBpediaに関しては、Brunkら[3]が、DBpediaのオントロジーを利用して、階層的なタイプ情報をファセットとして選ばせるtFacetを提案している。Arenasら[1]は、知識ベースのYAGOに対してファセット検索を行うSemFacetを提案している。SemFacetでは、ファセットにRDFにおける目的語やプロパティを使用し、検索対象にエンティティ（URI）を設定している。エンティティに関する情報が疎である場合があることが課題とされるが、OWL2のオントロジーを用いて推論することによって解決することを提案している。Papadakisら[7]は、ファセットをランキングするHippalusというシステムを提案している。Hippalusでは、ユーザが検索プロセスの中でファセットを評価し、そのファセットの評価に基づいてファセットをランキングして返す。これにより、ユーザの好みに合わせたファセット検索が行えるとしている。Bastら[2]は、知識ベースの1つであったFreebaseのファセット検索を提案している。主な特徴は、ファセット検索の利便性を向上させるために、データセットそのものを編集したことにある。具体的には、データセットに含まれる冗長なエンティティやプロパティを削除や統一、タイプのタキソノミー（分類）の編集などを行なっている。Wikidataに関しては、Moreno-Vegaら[5]が大規模な知識ベースに対するクエリの高速化を目指したGraFaを提案している。

上記の研究では、既存のプロパティをエンティティのファセットとして利用し、エンティティを検索することを目的としている。本研究では、まず関心のあるプロパティを見つけ出すことを重視する。そのために、プロパティに関するファセットを生

成する。そして、そのプロパティファセットに基づいて、関係付けられているエンティティの検索を可能にするシステムを目指している点異なる。

4 提案手法

知識ベースに対するプロパティ指向のファセット検索システムとして、図3のシステム概要を提案する。プロパティ指向のファセット検索システムの目的は、興味のあるエンティティをキーワード検索し、関心のあるプロパティ集合をプロパティファセットから選択することで、それらのプロパティで関係付けられたエンティティ（トリプル）を見つけることである。ユーザの操作とシステムで実行される処理は次の通りである。

(1) ユーザはキーワード検索を実行する。システムは、キーワードをエンティティのドキュメントに含むエンティティの部分集合を生成し、それらが主語または目的語に位置付けられているトリプル集合を主語あるいは目的語のランク値降順で返却する。また、返却したトリプル集合に含まれるプロパティについて、プロパティファセットを提示する。

(2) ユーザはプロパティファセットから興味のあるファセットキーを選択する。システムは、選択されたファセットキーに含まれるプロパティを持つトリプル集合を返却する。

1におけるキーワード検索の対象は、エンティティを説明するリテラルとエンティティを1つのドキュメントとみなしたドキュメント化されたエンティティである。本システムのポイントは、プロパティファセットのような柔軟なファセット生成を可能とするために、知識ベースから独立してデータベースを設計することにある。以降で、RDF データベース、エンティティデータベース、ファセットデータベースを説明する。

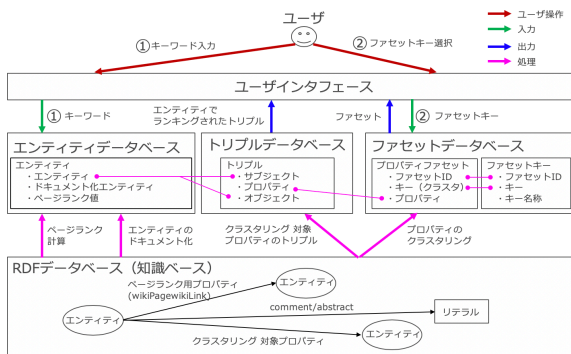


図3 システム概要

4.1 RDF データベース

RDF データベースは、キーワード検索とファセット検索に必要な知識ベースのデータを格納する。本稿では、エンティティのランク値を計算するためのRDFデータ、ファセットを生成するためのRDFデータ、検索対象となるドキュメント化されたエンティティを生成するためのRDFデータ、の3種類に大別できる。

4.2 エンティティデータベース

エンティティデータベースは、エンティティを主キーとして、検索対象のドキュメント化されたエンティティ、ランク値をタプルとしたテーブルを持つ。RDF データベースのエンティティのランク値を計算するためのデータをPageRankで処理し、結果をエンティティとランク値をタプルとしたテーブルに格納する。また、ドキュメント化されたエンティティを生成するためのRDFデータに対して、主語のエンティティと目的語のリテラルを1つのドキュメントとする処理を行う。結果として、エンティティとドキュメント化されたエンティティをタプルとしたテーブルに格納する。2つのテーブルを結合演算し、エンティティテーブルを作成する。

4.3 トリプルデータベース

クラスターリング対象のプロパティが述語であるトリプルをタプルとしたテーブルを持つ。なお、目的語にリテラルを含むトリプルは除いている。

4.4 ファセットデータベース

ファセットデータベースは、テーブル“プロパティファセット”とテーブル“ファセットキー”を持つ。プロパティファセットは、ファセットの種類を示す番号、ファセットの内容を示すキー、プロパティのURIをタプルとして持つ。ファセットキーは、ファセットの種類を示す番号、ファセットキー、キーの内容を示すラベルをタプルとして持つ。ファセットの種類を示す番号は、ファセットの種類の拡張に対応するためである。また、ファセットキーは、プロパティのクラスに該当する。本稿では、プロパティファセットを群平均法による階層型クラスターリングによって生成した。プロパティ間の距離は、Jaccard係数を変換したJaccard距離である。Jaccard係数は、各プロパティの主語に関するJaccard係数と目的語に関するJaccard係数の平均値である。この手法を適用した理由として、プロパティはエンティティとエンティティ（またはリテラル）の関係性として機能するため、プロパティ間で同じ主語や目的語を共有している可能性があるからである。別の方法として、RDFスキーマで定義されるdomain（主語に期待されるエンティティのクラス）やrange（目的語に期待されるエンティティのクラスやリテラルのデータ型）も考えられるが、あくまでも期待値であり実際には異なるクラスのエンティティに関係付けられていることもあるため、実際のデータに基づいてクラスターリングを行うことが有効と考えた。

このように、知識ベースから独立してデータベースを設計することで、実際のRDFデータの状態に合わせて必要なファセット生成の処理を柔軟に実行することができる。

5 予備実験

プロパティ指向のファセット検索システムの機能要素となるプロパティファセットの生成とその結果について目視による初期的な確認を行うことを目的として、次の2種類のプロパティのクラスターリング結果についての確認を行なった。

- (1) 上位プロパティによるクラスタリング
- (2) 階層型クラスタリングによるクラスタリング

また、想定するユーザ操作に従って、階層型クラスタリングによって生成したプロパティファセットによる検索結果の確認を行った。

5.1 クラスタリング

5.1.1 上位プロパティによるクラスタリング

知識ベースの語彙の体系に基づいて、プロパティをクラスタリングする。プロパティは、RDF スキーマと呼ばれる語彙を定義する基本的な仕組みによって、そのプロパティの性質が定義されている。その中で、上位プロパティを定義するために使用される `rdfs:subPropertyOf` を利用して、各プロパティを上位プロパティでクラスタリングした。

5.1.2 階層型クラスタリングによるクラスタリング

実際のデータの関係性に基づいて、プロパティをクラスタリングする。トリプル集合から、プロパティ間の主語に関する Jaccard 係数と目的語に関する Jaccard 係数の平均値を計算し、群平均法による階層型クラスタリングによって、主語と目的語との関係が類似するプロパティをクラスタとしてまとめた。

5.2 実験環境

提案手法を Python 3.7.3 で実装し、Intel(R) Core™i7-7700 3.60 GHz CPU, 32 GB RAM を搭載した Ubuntu 18.04.3 LTS で実験を行った。

5.3 データセット

本実験では、DBpedia 2016-10¹ の Page Links のデータセットの約 0.4%を取得し、検索対象のエンティティとした。エンティティの数は 449,339 個である。そして、同じ DBpedia 2016-10 の Mappingbased Objects のデータセットから検索対象のエンティティが含まれるトリプルを抽出し、クラスタリング対象のデータセットとした。トリプルに含まれるプロパティの種類数は 617 個である。

5.4 上位プロパティによるクラスタリング 結果

上位プロパティを使うと、617 個のプロパティのうち、615 個のプロパティを 41 個のクラスタにまとめることができた。2 個のプロパティは、上位プロパティが定義されていないため、クラスタから漏れてしまった。図 4 は 41 個のクラスタから、プロパティ数が 10 個以上となったクラスタの一覧である。約 4 割のプロパティが、2 つの上位プロパティにまとめられていることがわかる。それらのクラスタには、多様な領域に関するプロパティがメンバーとして含まれていることを確認した。

また、図 5 は、“hasPart” クラスタに含まれるプロパティの一覧である。組織、都市、食物など、様々な領域に関するプロパティが含まれていることがわかる。

5.5 階層型クラスタリングによるクラスタリング 結果

階層型クラスタリングでは、617 個のプロパティを 128 個の

上位プロパティ (クラスタ)	プロパティ数
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#coparticipatesWith	163
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#sameSettingAs	112
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasLocation	57
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasParticipant	37
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasCommonBoundary	29
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isMemberOf	26
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isParticipantIn	26
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isPartOf	24
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isClassifiedBy	21
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isLocationOf	19
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasPart	13

図 4 上位プロパティによるクラスタリング 結果の一部（プロパティ数が 10 以上のクラスタの一覧）

プロパティ
http://dbpedia.org/ontology/committeeInLegislature
http://dbpedia.org/ontology/frazioni
http://dbpedia.org/ontology/ingredient
http://dbpedia.org/ontology/largestCity
http://dbpedia.org/ontology/largestSettlement
http://dbpedia.org/ontology/mainOrgan
http://dbpedia.org/ontology/majorIsland
http://dbpedia.org/ontology/openingTheme
http://dbpedia.org/ontology/part
http://dbpedia.org/ontology/politicalPartyInLegislature
http://dbpedia.org/ontology/politicalPartyOfLeader
http://dbpedia.org/ontology/subregion
http://dbpedia.org/ontology/youthWing

図 5 “hasPart” クラスタのプロパティ一覧

クラスタを生成した。クラスタの生成はデンドログラムを確認し、分割の閾値を 0.995 として実行した。図 6 は、プロパティ数が 10 個以上となったクラスタの一覧で 19 個ある。クラスタの名称は、クラスタ内のプロパティの内容を確認して便宜的に付けた。クラスタのサイズは最大でも 51 となり、上位プロパティによるクラスタリングと比較して、クラスタのサイズの分散は小さくなっていることは明らかである。クラスタ内のプロパティが関係する領域も、クラスタの名称通り、一定の傾向があることを確認した。図 7 を具体例として提示する。一定の領域に属するプロパティの集合となっていることが確認できる。

クラスタ	プロパティ数
61：人に関するプロパティ	51
87：音楽に関するプロパティ	38
69：企業に関するプロパティ	34
22：地理に関するプロパティ	23
39：航空機に関するプロパティ	20
55：政治に関するプロパティ	19
65：河川に関するプロパティ	18
86：創作物に関するプロパティ	17
42：カーレースに関するプロパティ	15
28：スポーツチームや選手に関するプロパティ	14
68：大学に関するプロパティ	13
2：器官に関するプロパティ	13
79：分類学に関するプロパティ	12
4：ゴルフ選手に関するプロパティ	11
49：地域に関するプロパティ	11
57：国に関するプロパティ	11
78：自動車に関するプロパティ	10
46：競走馬に関するプロパティ	10
60：人文科学に関するプロパティ	10

図 6 階層型クラスタリングによるクラスタリング 結果の一部（プロパティ数が 10 以上のクラスタの一覧）

1 : <https://wiki.dbpedia.org/downloads-2016-10>

プロパティ
http://dbpedia.org/ontology/trainer
http://dbpedia.org/ontology/sire
http://dbpedia.org/ontology/raceHorse
http://dbpedia.org/ontology/race
http://dbpedia.org/ontology/honours
http://dbpedia.org/ontology/grandsire
http://dbpedia.org/ontology/damsire
http://dbpedia.org/ontology/dam
http://dbpedia.org/ontology/breeder
http://dbpedia.org/ontology/billed

図 7 “競走馬に関するプロパティ” クラスタのプロパティ一覧

5.6 プロパティファセットによる検索結果

実験に関して想定するユーザは次の通りである。

- (1) 知識ベースに関する専門知識を持たない
- (2) “Arab Spring” について調べる

5.6.1 キーワード検索結果

キーワード検索の結果、主語のエンティティのドキュメントに “Arab Spring” を含むトリプルを 373 個、目的語のエンティティのドキュメントに “Arab Spring” を含むトリプルを 245 個を得た。全トリプルに含まれるプロパティの種類数は 68 個だった。図 8 はその一部の抜粋である。

主語のエンティティに “Arab Spring” を含む		
subject	property	object
http://dbpedia.org/resource/Bahrain	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Hamad_bin_Isa_Al_Khalifa
http://dbpedia.org/resource/Bahrain	http://www.w3.org/2000/01/rdf-schema#seeAlso	http://dbpedia.org/resource/Bahrain
http://dbpedia.org/resource/Bahrain	http://www.w3.org/2000/01/rdf-schema#seeAlso	http://dbpedia.org/resource/List_of_universities
http://dbpedia.org/resource/Bahrain	http://www.w3.org/2000/01/rdf-schema#seeAlso	http://dbpedia.org/resource/Peninsula_Shield_Force
http://dbpedia.org/resource/Bahrain	http://dbpedia.org/ontology/language	http://dbpedia.org/resource/Arabic_language
http://dbpedia.org/resource/Bahrain	http://dbpedia.org/ontology/capital	http://dbpedia.org/resource/Manama

目的語のエンティティに “Arab Spring” を含む		
subject	property	object
http://dbpedia.org/resource/Karim_Fakhrai	http://dbpedia.org/ontology/deathPlace	http://dbpedia.org/resource/Bahrain
http://dbpedia.org/resource/Karim_Fakhrai	http://www.w3.org/2000/01/rdf-schema#seeAlso	http://dbpedia.org/resource/Bahrain
http://dbpedia.org/resource/Arab_Banking_Corporation	http://dbpedia.org/ontology/location	http://dbpedia.org/resource/Bahrain
http://dbpedia.org/resource/Al-Nasr_Bahraini_newspaper	http://dbpedia.org/ontology/headquarter	http://dbpedia.org/resource/Bahrain
http://dbpedia.org/resource/Zakariya_Rashid_Hassan_al-Ashiri	http://dbpedia.org/ontology/stateOfOrigin	http://dbpedia.org/resource/Bahrain

図 8 “Arab Spring” のキーワード検索結果

5.6.2 キーワード検索結果に対応するプロパティファセット

キーワード検索結果に対応するプロパティファセットは、図 9 の通りである。なお、この結果は図 6 で提示したファセットキーのみに絞っている。この結果を踏まえて、ユーザは政治に関するプロパティを選択すると仮定する。

ファセットキー	プロパティ数
61 (人に関するプロパティ)	34
69 (企業に関するプロパティ)	13
57 (国に関するプロパティ)	6
55 (政治に関するプロパティ)	4
87 (音楽に関するプロパティ)	2
68 (大学に関するプロパティ)	2
86 (創作物に関するプロパティ)	1
60 (人文科学に関するプロパティ)	1

図 9 キーワード検索結果に対応するプロパティファセット

5.6.3 ファセット検索結果

政治に関するプロパティのキーを選択した検索結果は、図 10 の通りである。主語のエンティティのドキュメントに “Arab Spring” を含むトリプルを 10 個、目的語のエンティティのドキュメントに “Arab Spring” を含むトリプルを 11 個得た。“Arab Spring” に関係する国や政党の指導者や思想を記述したトリプルであることがわかる。

主語のエンティティに “Arab Spring” を含む		
subject	property	object
http://dbpedia.org/resource/Bahrain	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Hamad_bin_Isa_Al_Khalifa
http://dbpedia.org/resource/Bahrain	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Khalifa_bin_Salman_Al_Khalifa
http://dbpedia.org/resource/Bahrain	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Salman_bin_Hamad_bin_Isa_Al_Khalifa
http://dbpedia.org/resource/Qatar	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Abdullah_bin_Hamad_bin_Khalifa_Al_Thani
http://dbpedia.org/resource/Qatar	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Abdullah_bin_Nasser_bin_Khalifa_Al_Thani
http://dbpedia.org/resource/Qatar	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Tamim_bin_Hamad_Al_Thani
http://dbpedia.org/resource/Muslim_Brotherhood	http://dbpedia.org/ontology/ideology	http://dbpedia.org/resource/Pan-Islamism
http://dbpedia.org/resource/Muslim_Brotherhood	http://dbpedia.org/ontology/ideology	http://dbpedia.org/resource/Sunni_Islam
http://dbpedia.org/resource/Muslim_Brotherhood	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Mohamed_Radie
http://dbpedia.org/resource/Muslim_Brotherhood	http://dbpedia.org/ontology/speaksPerson	http://dbpedia.org/resource/Gehad_el-Haddad

目的語のエンティティに “Arab Spring” を含む		
subject	property	object
http://dbpedia.org/resource/Bangladesh_Jamaat-e-Islami	http://dbpedia.org/ontology/ideology	http://dbpedia.org/resource/Islamism
http://dbpedia.org/resource/Green_Algeria_Alliance	http://dbpedia.org/ontology/ideology	http://dbpedia.org/resource/Islamism
http://dbpedia.org/resource/Jamaat-e-Islami	http://dbpedia.org/ontology/ideology	http://dbpedia.org/resource/Islamism
http://dbpedia.org/resource/Jamaat-e-Islami_Pakistan	http://dbpedia.org/ontology/ideology	http://dbpedia.org/resource/Islamism
http://dbpedia.org/resource/Jamaat-e-Islami	http://dbpedia.org/ontology/ideology	http://dbpedia.org/resource/Islamism
http://dbpedia.org/resource/Lamma_Party_(Egypt)	http://dbpedia.org/ontology/ideology	http://dbpedia.org/resource/Islamism
http://dbpedia.org/resource/Hamas	http://dbpedia.org/ontology/internationalAffiliation	http://dbpedia.org/resource/Muslim_Brotherhood
http://dbpedia.org/resource/Jamaat-e-Islami_Pakistan	http://dbpedia.org/ontology/internationalAffiliation	http://dbpedia.org/resource/Muslim_Brotherhood
http://dbpedia.org/resource/Muslim_Brotherhood_in_Egypt	http://dbpedia.org/ontology/internationalAffiliation	http://dbpedia.org/resource/Muslim_Brotherhood
http://dbpedia.org/resource/Sunni_Islam	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Bashar_al-Assad
http://dbpedia.org/resource/Armenian_National_Congress	http://dbpedia.org/ontology/leader	http://dbpedia.org/resource/Levon_Ter-Petrosyan

図 10 “Arab Spring” のキーワード検索結果に対するファセット検索結果

6 まとめと今後の課題

本稿では、プロパティ指向のファセット検索システムの全体概要について提案し、その機能要素となるファセット生成のためのクラスタリングに関する予備実験について報告した。予備実験では、プロパティの主語と目的語に関する Jaccard 係数を利用した階層型クラスタリングによって、一定程度関係する領域でプロパティをまとめられることを確認できた。また、上位プロパティを利用したクラスタリングの場合は、同一クラスターであっても関係する領域の異なるプロパティの集合になることを確認できた。

今後の課題として、これらのクラスタリングを使った検索のユーザビリティについて調査を行う。また、そのためにプロパティ指向のファセット検索システムの実装についても進める必要がある。

7 謝辞

2019 年度共同研究 (SKY 株式会社) (CPE01017) 「データエンジニアリングの知見の応用による SKYSEA Client View のログ及び資産情報の処理の高速化・軽量化・高度化」の助成を受けたものです。

文献

- [1] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, Dmitriy Zheleznyakov, and Ernesto Jimenez-Ruiz. Semfacet: Semantic faceted search over yago. In *Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion*, pp. 123–126, New York, NY, USA, 2014. ACM.
- [2] Hannah Bast, Florian Baurle, Björn Buchhold, and Elmar

- Haußmann. Easy access to the freebase dataset. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pp. 95–98, New York, NY, USA, 2014. ACM.
- [3] Sören Brunk and Philipp Heim. tfacet: Hierarchical faceted exploration of semantic data using well-known interaction concepts. In *DCI@INTERACT*, 2011.
 - [4] Shiori Ichinose, Ichiro Kobayashi, Michiaki Iwazume, and Kouji Tanaka. Ranking the results of dbpedia retrieval with sparql query. In *JIST*, 2013.
 - [5] José Moreno-Vega and Aidan Hogan. Grafa: Scalable faceted browsing for rdf graphs. In Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, and Elena Simperl, editors, *The Semantic Web – ISWC 2018*, pp. 301–317, Cham, 2018. Springer International Publishing.
 - [6] 奥村彩水, 天笠俊之, 北川博之. リンク構造解析を用いた linked open data に対するキーワード検索. In *DEIM Forum*, 2016.
 - [7] Panagiotis Papadakos and Yannis Tzitzikas. Hippalus: Preference-enriched faceted exploration. In *EDBT/ICDT Workshops*, 2014.