

DEIM2020

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

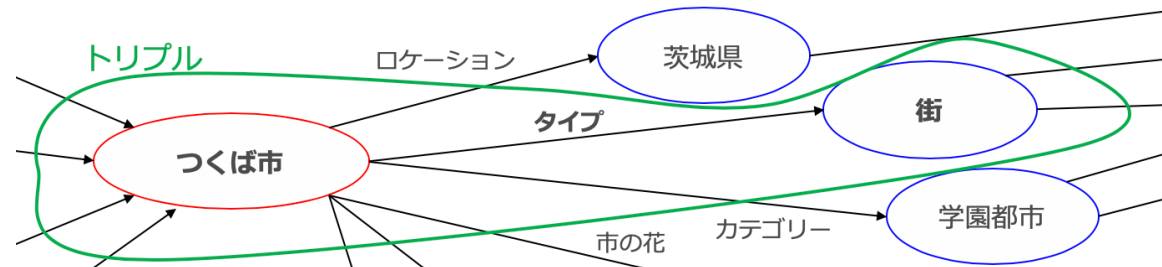
筑波大学

- 知識ベース

- 様々な知識が蓄積されたデータベース
- 代表的な知識ベース
 - DBpedia, Wikidata, YAGO, Freebaseなど

- Resource Description Framework (RDF)

- リソース(エンティティ)に関する情報を記述する方法
- 主語, 述語, 目的語の3つ組 (トリプル) で記述されたグラフデータ



- 検索の意図が明確な場合

- 問合せ言語SPARQL

```
select distinct * where
{?film <http://dbpedia.org/ontology/director> ?who.
?film <http://dbpedia.org/ontology/starring> ?who.
?film <http://dbpedia.org/ontology/writer> ?who.
?who a <http://dbpedia.org/ontology/Comedian>}
```

- 検索の意図が曖昧な場合

- キーワード検索

- 検索結果には様々なエンティティが含まれ、情報が整理されていない

- ファセット検索

- 検索結果に関する切り口（ファセット）を提供
 - 専門知識を持たないユーザも検索しやすい

– 例：“Tsukuba”	
植物園	dbr:Tsukuba_Botanical_Garden Estación experimental de plantas medicinales de Tsukuba
人？シンボル？	dbr:Sakura_Tsukuba Sakura Tsukuba
クルーザー？	dbr:Japanese_cruiser_Tsukuba Tsukuba (Schiff, 1905)
大学	dbr:Tsukuba_University_of_Technology Tsukuba University of Technology wikidata:Q273863SADF...0-832E-611E252166C3 wikidata:Q11272258S2...8-b762-c270a01193d0 dbc:Tsukuba_Ibaraki Tsukuba, Ibaraki dbc:Tsukuba-class_cruisers Tsukuba-class cruisers dbr:Tsukuba_University Tsukuba University wikidata:Q11602899SC...2-AFD3-4CCD524B36B7 wikidata:Q11251506 Tsukuba club wikidata:Q11602882 Tsukuba College of Technology wikidata:Q11272231 つくばセンター

研究背景：ファセット検索システム

ユーザはデータに関する切り口（ファセット）を選択・解除しながら，検索結果を探索する

The figure illustrates the DBpedia faceted search interface through three sequential screenshots, numbered 1, 2, and 3, showing how filters are applied to refine search results.

Screenshot 1: The initial search results page for 'Scientist' (8465 results). The left sidebar shows various facets, with 'Scientist (8465)' highlighted. The main content area shows 'Your Filters: No Filters selected'.

Screenshot 2: The 'Scientist' facet is selected, and the results are refined to 8465. The left sidebar shows the 'nationality' facet, with 'Germany (485)' highlighted. The main content area shows 'Your Filters: item type Scientist'.

Screenshot 3: The 'Germany' facet is selected, and the results are further refined to 485. The left sidebar shows the 'nationality' facet, with 'Germany (485)' highlighted. The main content area shows 'Your Filters: item type Scientist, nationality Germany'. The search results are displayed as a list of items, including 'Carl Friedrich Gauss' and 'Johann Friedrich Gmelin'.

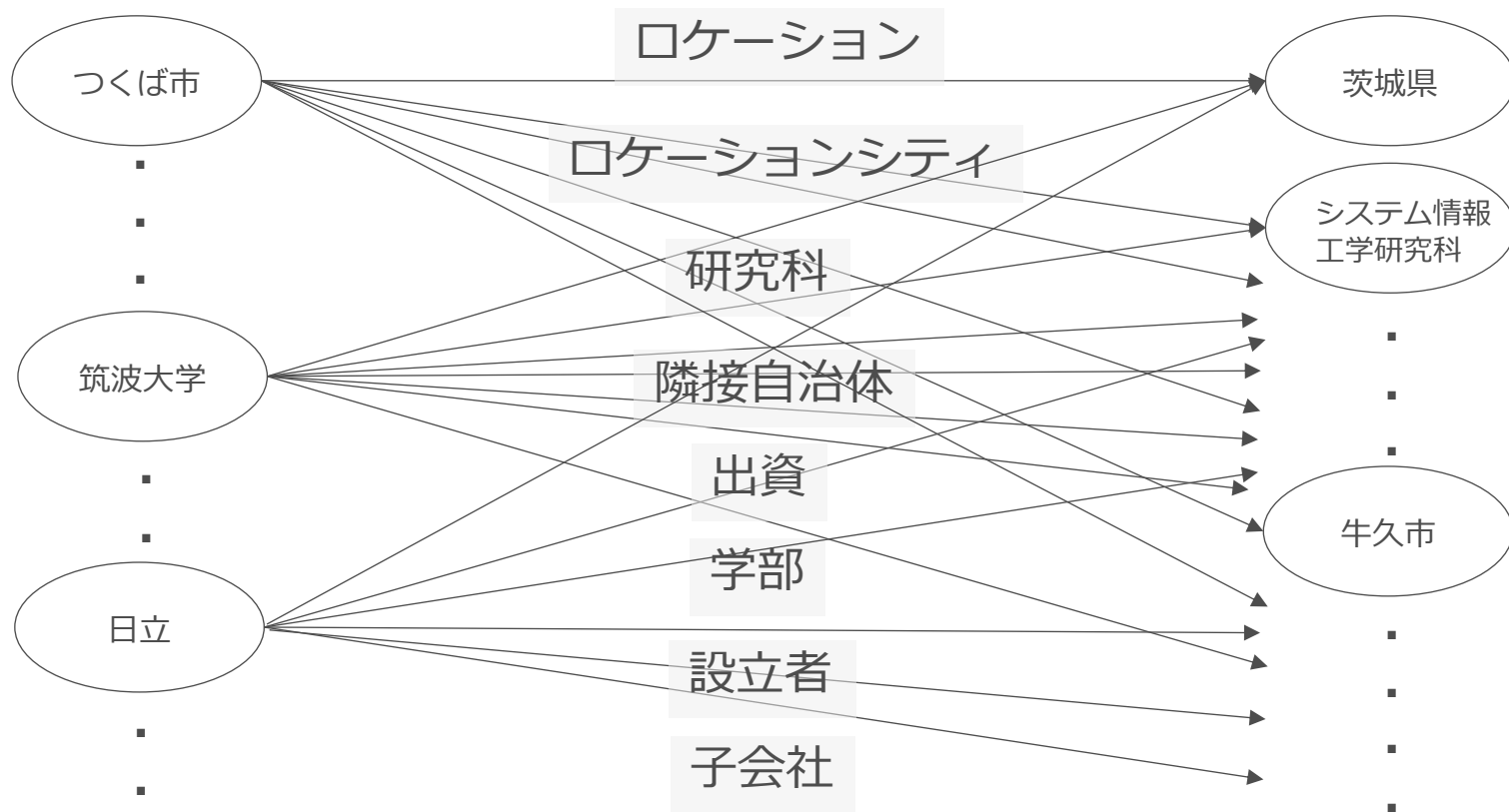
Red dashed boxes and arrows highlight the selected facets and the resulting search results. A red label '検索結果 (エンティティ一覧)' (Search Results (Entity List)) points to the list of results in Screenshot 3.

研究背景：関係性に着目したエンティティ関係の検索ニーズ

- 例：企業間の関係を調べたい

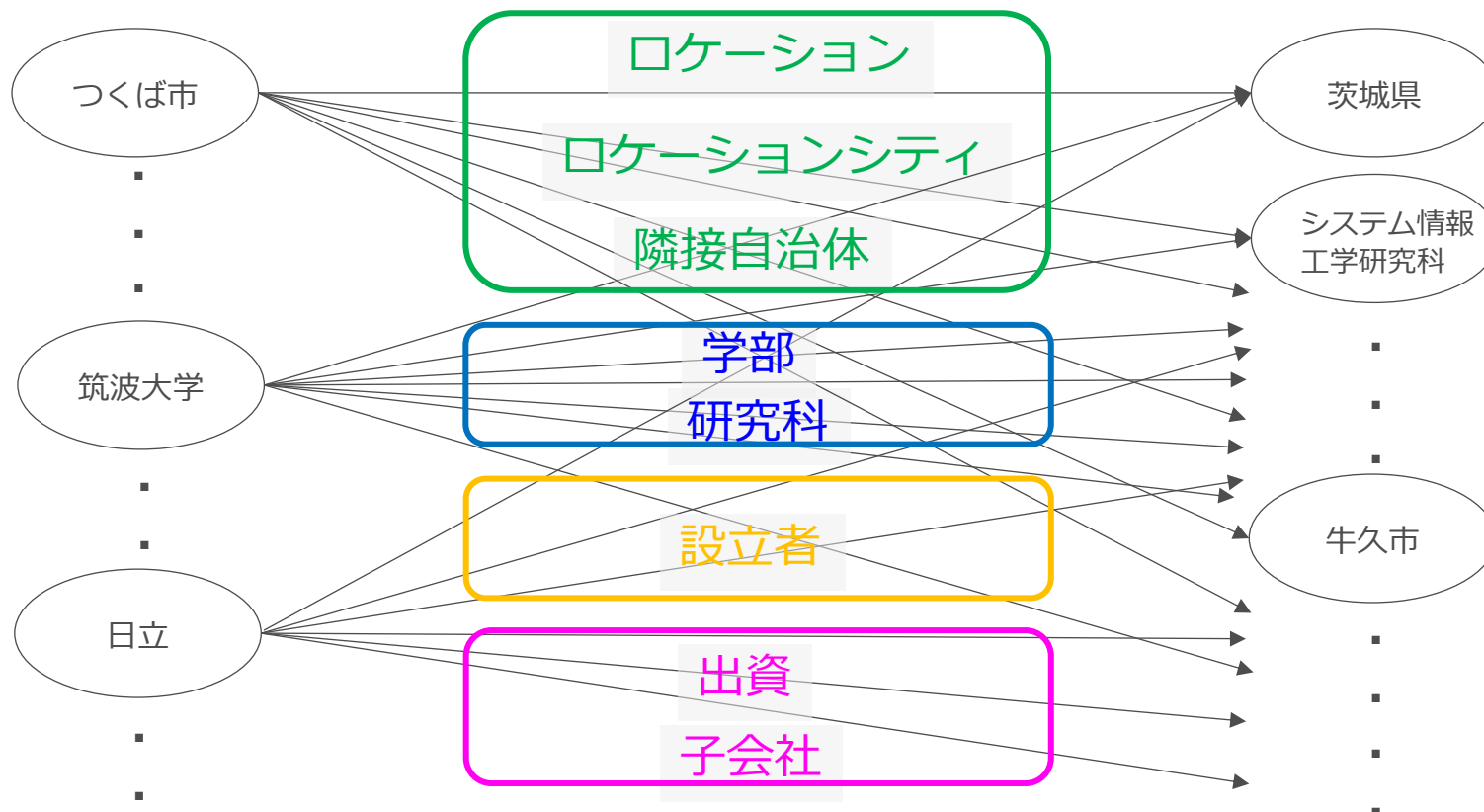
- 企業間の関係とは？

- 多くの関係性（DBpediaで約1300個）から関心のある関係性を見つけることは難しい



研究背景：プロパティのクラスタリング

- 類似するプロパティをまとめ、ファセット化することで、関心のあるプロパティを探索し易くする
 - プロパティの階層構造を利用する
 - プロパティの主語、目的語との関係を利用する



- 目的
 - エンティティ間の多様な関係性（プロパティ），それらに基づくエンティティ関係（トリプル）を探索できるようにする
- 提案手法
 - プロパティ指向のファセット検索システム
 - プロパティのクラスタリング 結果をファセットとして利用する
 - エンティティ間の関係性（プロパティ）に着目して，関係付けられているエンティティの探索を可能にする

- tFacet [Brunk et al. 2011]
 - DBpediaのオントロジーを利用した階層的なタイプ情報をファセットとして利用する
- Hippalus [Papadakos et al. 2014]
 - 検索プロセスの中でユーザがファセットを評価し、評価に基づいてファセットをランキングする
- GraFa [Moreno-Vega et al. 2017]
 - Wikidataのような巨大な知識ベースに対して、転置索引とマテリアライズドクエリによる検索の高速化を図る

関連研究の目的と課題

- 関連研究に共通する目的
 - エンティティの探索・発見を指向したシステム



- 課題：エンティティ関係は検索しにくい
 - エンティティ間の多様な関係性（プロパティ）
 - エンティティ関係（トリプル）

- 目的

- エンティティ間の多様な関係性（プロパティ），それらに基づくエンティティ関係（トリプル）を探索できるようにする

- 提案手法

- プロパティ指向のファセット検索システム
 - プロパティのクラスタリング 結果をファセットとして利用する
 - エンティティ間の関係性（プロパティ）に着目して，関係付けられているエンティティの探索を可能にする

提案手法：プロトタイプシステムのユーザインターフェース

ProFacet

A Keyword: Tsukuba on ☒ Subject ☐ Object

B Subject Type: <http://dbpedia.org/ontology/Scientist> (2)

Predicate Type: [academics](#) (2)

Object Type: <http://www.w3.org/2002/07/owl#Thing> (2)

Search

Search Condition

C Keyword: "Tsukuba" on "Subject"

Subject Type: [<http://dbpedia.org/ontology/Scientist>]

Predicate Type: [[academics](#)]

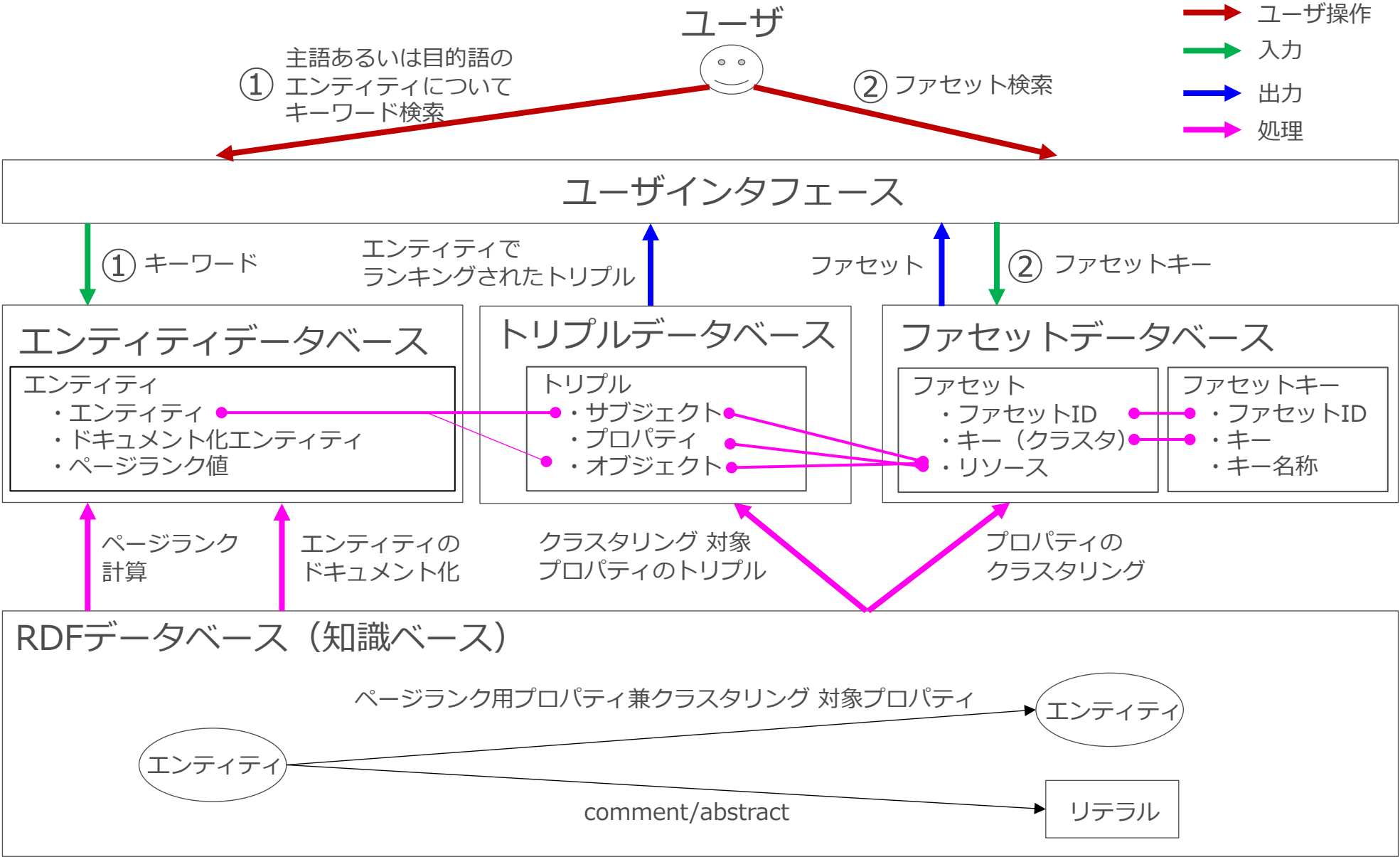
Object Type: [<http://www.w3.org/2002/07/owl#Thing>]

No.	Subject	Predicate	Object
0	Katsunori_Wakabayashi	field	Nanotechnology
1D	Mitsutaka_Fujita	field	Nanotechnology
2	Katsunori_Wakabayashi	award	Physical_Society_of_Japan
3	Mitsutaka_Fujita	award	Physical_Society_of_Japan

機能

- A) キーワード入力
- B) ファセット選択
- C) 検索状態
- D) 検索結果

提案手法：システム概要



- 役割
 - エンティティやプロパティのファセットに関するテーブルを格納すること
- プロトタイプシステムの場合
 - エンティティのファセット
 - エンティティのクラス情報
 - プロパティのファセット
 - プロパティのクラスタリング 結果

- アイデア

- 実データにおけるプロパティの主語，目的語の重複度合いを類似度として，似ているプロパティをまとめる

- 手法

- プロパティの主語，目的語に関するJaccard係数の群平均法による階層型クラスタリング

1. プロパティについて 2 つのJaccard係数行列を生成する
 - 主語に関するJaccard係数行列
 - 目的語に関するJaccard係数行列
2. 2つのJaccard係数行列の成分の平均値の行列をプロパティに関する類似度行列とする
3. 群平均法による階層型クラスタリング を適用する

予備実験：全体概要

- 目的
 - 提案システムの概念実証（Proof of Concept）
- 予備実験
 1. プロパティをクラスタリング し, 類似するプロパティでまとめられているかを確認した
 2. プロトタイプシステムでプロパティファセットを用いた検索が機能することを確認した
- データセット
 - DBpedia 2016-10の5クラスに属すエンティティを記述するトリプル

	<i>Astronaut</i>	<i>Company</i>	<i>Politician</i>	<i>Scientist</i>	<i>University</i>	<i>Total</i>
Number of Subjects	635	54570	16986	23423	20214	115828
Number of Predicates	11	29	54	31	28	101
Number of Objects	1536	117567	27829	51067	39461	219968
Number of Triples	5157	349860	78733	172382	113587	719719

- 概要

- 2種類のクラスタリング 結果を比較した
 1. 上位プロパティによるクラスタリング（比較手法）
 2. 階層型クラスタリング（提案手法）

- 上位プロパティによるクラスタリング

- 各プロパティはRDFスキーマという語彙を定義する基本的な枠組みによってプロパティの性質が定義されている
- 上位プロパティを示す“`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 する

予備実験①「プロパティのクラスタリング」：結果（比較手法）

上位プロパティによるクラスタリング 結果

Clusters	Number of Properties
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#sameSettingAs	32
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#coparticipatesWith	26
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasLocation	17
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isMemberOf	7
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isClassifiedBy	3
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasRole	3
http://dbpedia.org/ontology/location	2
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isDescribedBy	2
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isPartOf	2
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isParticipantIn	2
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasSetting	1
http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#conceptualizes	1

同種の意味合いだが、名前空間の異なるプロパティが混在している

最大クラスターの "sameSettingAs" の要素

appointer	predecessor	monarch	runningMate
athletics	president	nominee	service
child	primeMinister	owner	spouse
citizenship	principal	owningCompany	subsidiary
education	provost	parent	successor
incumbent	rector	parentCompany	superintendent
influencedBy	relation	parentOrganisation	viceChancellor
keyPerson	relative	partner	vicePresident

“同じ状況に参加している2つのエンティティ間の関係”を示す“sameSettingAs”と左の要素の関係性を直感的には理解しにくい

予備実験①「プロパティのクラスタリング」：結果（提案手法）

階層型クラスタリングの結果

clusters	Number of properties	clusters	Number of properties
company	20	director	2
academics	15	ethnicity	1
university	11	deathCause	1
professionals	11	appointer	1
personalInformation	7	language	1
military	6	dean	1
politics	4	provost	1
governance	4	vicePresident	1
relationship	4	officerInCharge	1
organization	3	depiction	1
soundRecording	2	partner	1
principal	2		

クラスタの名称は、要素数が1の場合は、その要素の名称をそのまま用いた。複数の場合は、各要素の内容を確認して便宜的に付けた。

“University”と“Academics” クラスタの要素

大学のような機関・組織のエンティティに対して使用されるプロパティと、学者のような人間のエンティティに対して使用されるプロパティでま
とまっていることがわかる

Properties in "University" cluster	Properties in "Academics" cluster
affiliation	academicAdvisor
athletics	almaMater
campus	award
chancellor	birthPlace
city	citizenship
country	deathPlace
head	doctoralAdvisor
sport	doctoralStudent
state	field
viceChancellor	influenced
differentFrom	influencedBy
	knownFor
	nationality
	notableStudent
	residence

- 概要

- プロトタイプシステムを用いて、プロパティのファセットを用いた検索が機能することを確認する

- ユースケース

- 主語のエンティティに対して“Ivy League”でキーワード検索を行う
- 検索結果を得た後、エンティティ間の関係性に着目して、さらに探索を行う

予備実験②「プロパティファセットによる検索」：キーワード検索

- 26個のプロパティを含む359個のトリプルに対して、8個のプロパティファセットのキーが生成された

ProFacet

Keyword: on ☒ Subject ☐ Object

Subject Type: Predicate Type: Object Type:

****Search Condition****

Keyword: "Ivy League" on "Subject"

Number: 359 records

No.	Subject	Predicate	Object
0	Harvard_University	campus	Urban_area
1	Harvard_University	city	Cambridge, Massachusetts
2	Harvard_University	president	Drew Gilpin Faust
3	Harvard_University	state	Massachusetts
4	Harvard_University	type	Research_university
5	Harvard_University	affiliation	Association of American Universities
6	Harvard_University	affiliation	Association of Independent Colleges and Universities in N
7	Harvard_University	affiliation	National Association of Independent Colleges and Univers
8	Harvard_University	affiliation	Universities Research Association
9	Harvard_University	athletics	NCAA Division_I
10	Princeton_University	athletics	Eastern Intercollegiate Volleyball Association
11	Princeton_University	athletics	Ivy League
12	Princeton_University	athletics	Middle Atlantic Intercollegiate Sailing Association
13	Princeton_University	athletics	NCAA Division_I
14	Princeton_University	state	New Jersey
15	Princeton_University	soundRecording	Princeton_University_1

予備実験②「プロパティファセットによる検索」：ファセット検索

- ファセットキー“University”を選択した結果，大学関係のエンティティに関係付けられやすい10個のプロパティを持つ215個のトリプルに絞られた

ProFacet

Keyword: on ☒ Subject ☐ Object

Subject Type: Predicate Type: Object Type:

****Search Condition****

Keyword: "Ivy League" on "Subject"

Subject Type: [-] Predicate Type: [university]

Number: 215 records

No.	Subject	Predicate	Object
0	Harvard_University	city	Cambridge, Massachusetts
1	Harvard_University	affiliation	Association of Independent Colleges and Universities in M
2	Harvard_University	affiliation	National Association of Independent Colleges and Univers
3	Harvard_University	affiliation	Universities Research Association
4	Harvard_University	affiliation	Association of American Universities
5	Harvard_University	athletics	NCAA Division I
6	Harvard_University	campus	Urban area
7	Harvard_University	state	Massachusetts
8	Princeton_University	athletics	Eastern Intercollegiate Volleyball Association
9	Princeton_University	affiliation	Association of American Universities
10	Princeton_University	affiliation	Universities Research Association
11	Princeton_University	affiliation	National Association of Independent Colleges and Univers
12	Princeton_University	city	Princeton, New Jersey
13	Princeton_University	state	New Jersey
14	Princeton_University	campus	Suburb
15	Princeton_University	athletics	NCAA Division I
16	Princeton_University	athletics	Ivy League

ユーザはさらに目的語のエンティティのタイプを選ぶ等して探索を行える

- [http://dbpedia.org/ontology/Thing \(10\)](#)
- [http://dbpedia.org/ontology/AdministrativeRegion \(8\)](#)
- [http://dbpedia.org/ontology/Organisation \(6\)](#)
- [http://dbpedia.org/ontology/Country \(4\)](#)
- [http://dbpedia.org/ontology/Hospital \(3\)](#)
- [http://dbpedia.org/ontology/OfficeHolder \(3\)](#)
- [http://dbpedia.org/ontology/Scientist \(1\)](#)
- [http://dbpedia.org/ontology/Publisher \(1\)](#)
- [http://dbpedia.org/ontology/TelevisionShow \(1\)](#)
- [http://dbpedia.org/ontology/Town \(1\)](#)
- [http://dbpedia.org/ontology/Building \(1\)](#)
- [http://dbpedia.org/ontology/BasketballLeague \(1\)](#)

- まとめ

- エンティティ間の関係性に着目してエンティティの関係を探索するというプロパティ指向のファセット検索システムを提案し、そのプロトタイプを開発した
- 実データの状態に基づいてプロパティをクラスタリングして生成したプロパティファセットが機能することを確認した

- 今後の課題

- ユーザ調査を実施し、エンティティの関係を探索するというニーズにおいて、提案システムが既存のファセット検索システムに比べて有効であることを確認する