

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

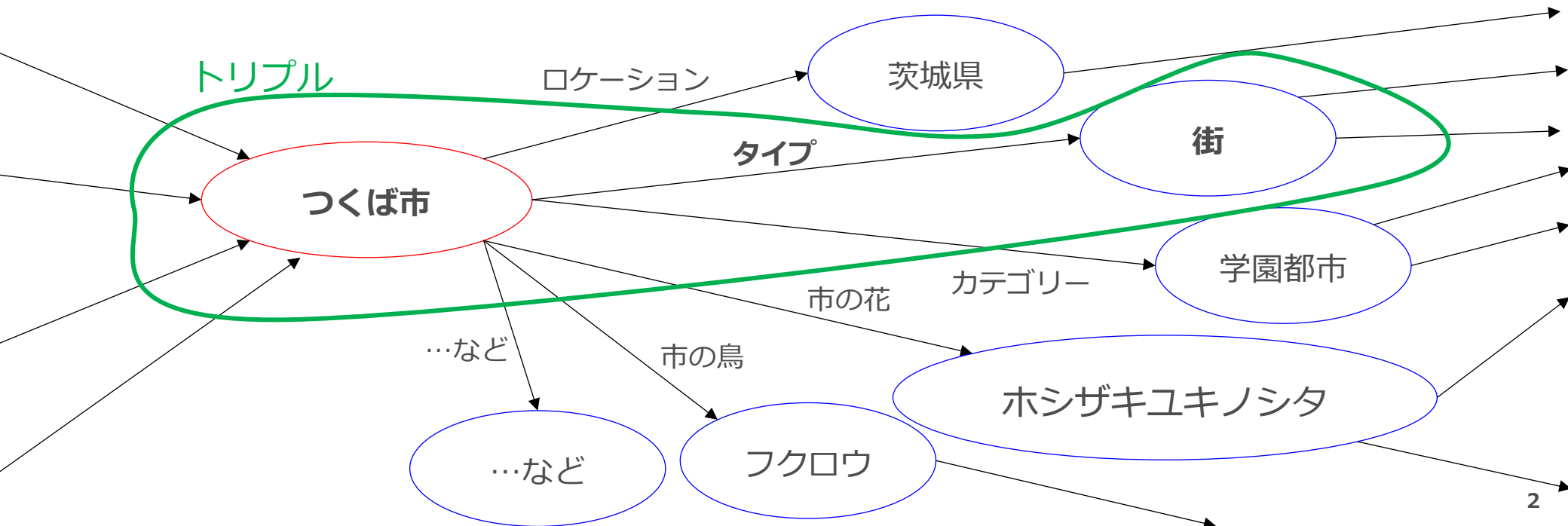
阿曾太郎，天笠俊之，北川博之

筑波大学

- 知識ベースとは
  - 様々な知識が蓄積されたデータベース
  - 人間や機械が利用することができる
- 代表的な知識ベース
  - DBpedia, Wikidata, YAGO, Freebaseなど

# 研究背景：Resource Description Framework(RDF)

- 知識ベースの記述にはRDFが用いられる
- 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



## 1. 問合せ言語SPARQL

- SPARQLの習得が必要

```
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>}
```

## 2. 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

– 例：“Tsukuba”

植物園	<a href="#">dbr:Tsukuba_Botanical_Garden</a>	Estación experimental de plantas medicinales de Tsukuba
	<a href="#">dbr:Sakura_Tsukuba</a>	Sakura Tsukuba
人？シンボル？	<a href="#">dbr:Japanese_cruiser_Tsukuba</a>	Tsukuba (Schiff, 1905)
クルーザー？	<a href="#">dbr:Tsukuba_University_of_Technology</a> <a href="#">wikidata:Q2738635ADF...0-832E-611E252166C3</a> <a href="#">wikidata:Q11272258S2...8-b762-c2f0a01193cb</a>	Tsukuba University of Technology
	<a href="#">dbc:Tsukuba...Ibaraki</a> <a href="#">dbc:Tsukuba-class_cruisers</a> <a href="#">dbr:Tsukuba_University</a>	Tsukuba, Ibaraki Tsukuba-class cruisers Tsukuba University
大学	<a href="#">wikidata:Q11602899SC...2-AFD3-4CCD524B36B2</a> <a href="#">wikidata:Q11251506</a> <a href="#">wikidata:Q11602882</a> <a href="#">wikidata:Q11272231</a>	Tsukuba club Tsukuba College of Technology つくばセンター

## 3. 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

# 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 item types (Place, Work, Species, etc.) and nationalities (Germany, United Kingdom, etc.). The 'Scientist' filter is highlighted in red.

**Screenshot 2:** The 'Scientist' filter is selected, and the results are refined to 8465. The 'Germany' filter is highlighted in red in the nationality section.

**Screenshot 3:** The 'Germany' filter is selected, and the results are refined to 485. The 'item type Scientist' and 'nationality Germany' filters are highlighted in red. The search results are displayed on the right, showing profiles for Carl Friedrich Gauss, Johann Friedrich Gmelin, and Carl Friedrich Philipp von Martius.

The interface includes a search bar, navigation links (First, Previous, Next, Last), and a 'Your Filters' section showing the current selection. The search results are displayed on the right side of the interface.

- 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

# 研究背景：関連研究の目的と課題

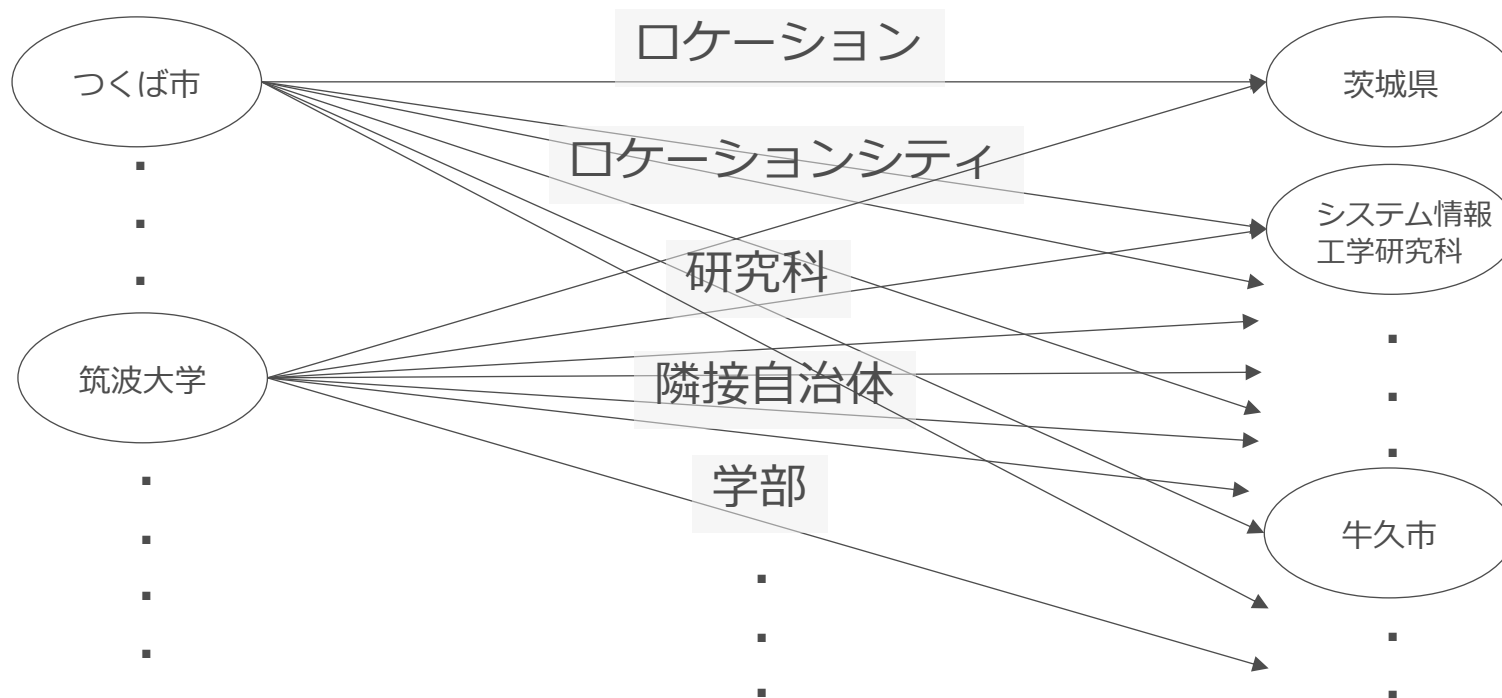
- 関連研究に共通する目的
  - エンティティの探索・発見を指向したシステム



- 課題：RDFの他の特徴に関しては探索しにくい
  - 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）

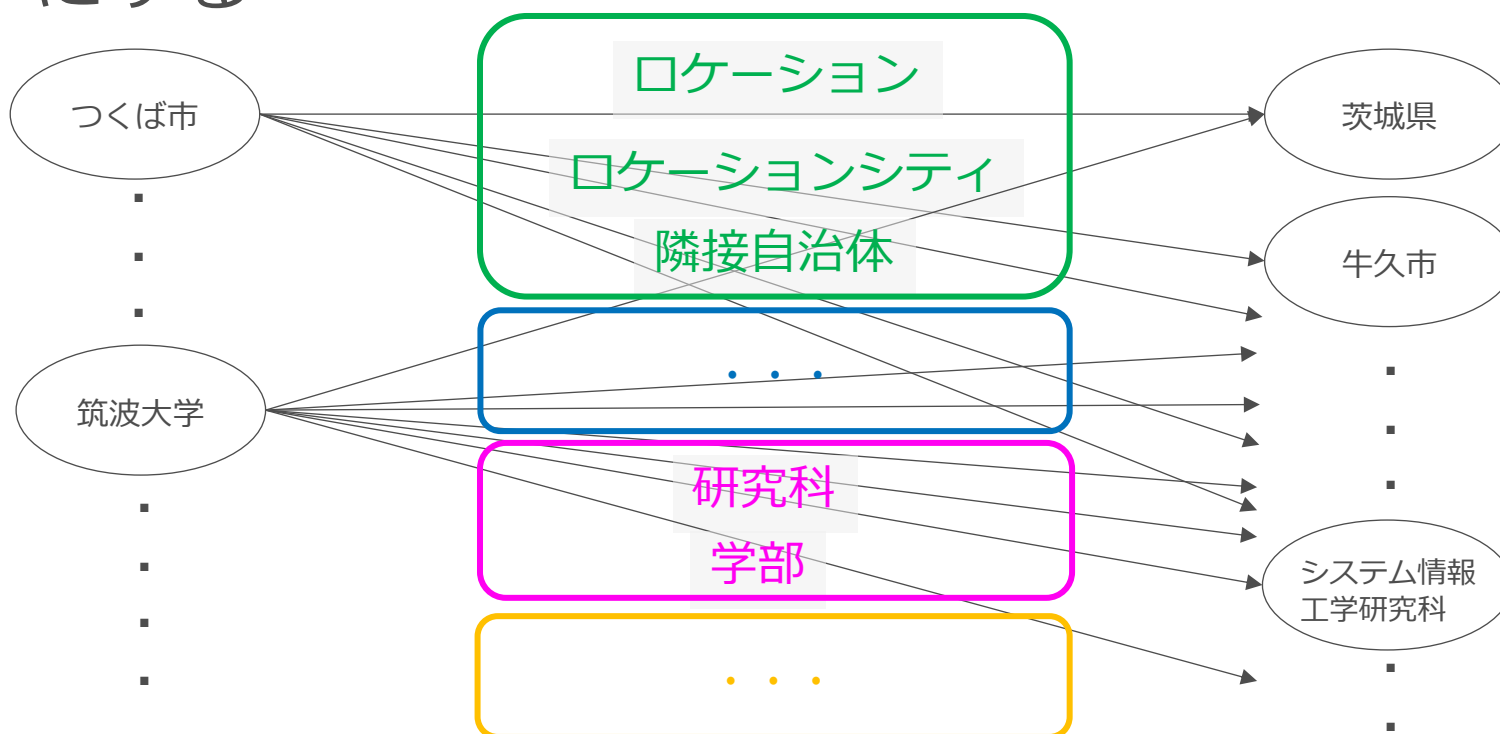
- 目的

- エンティティ間の多様な関係性（プロパティ）や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





- プロパティ指向のファセット検索システム
  - プロパティに関するファセットを生成し、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](#)]

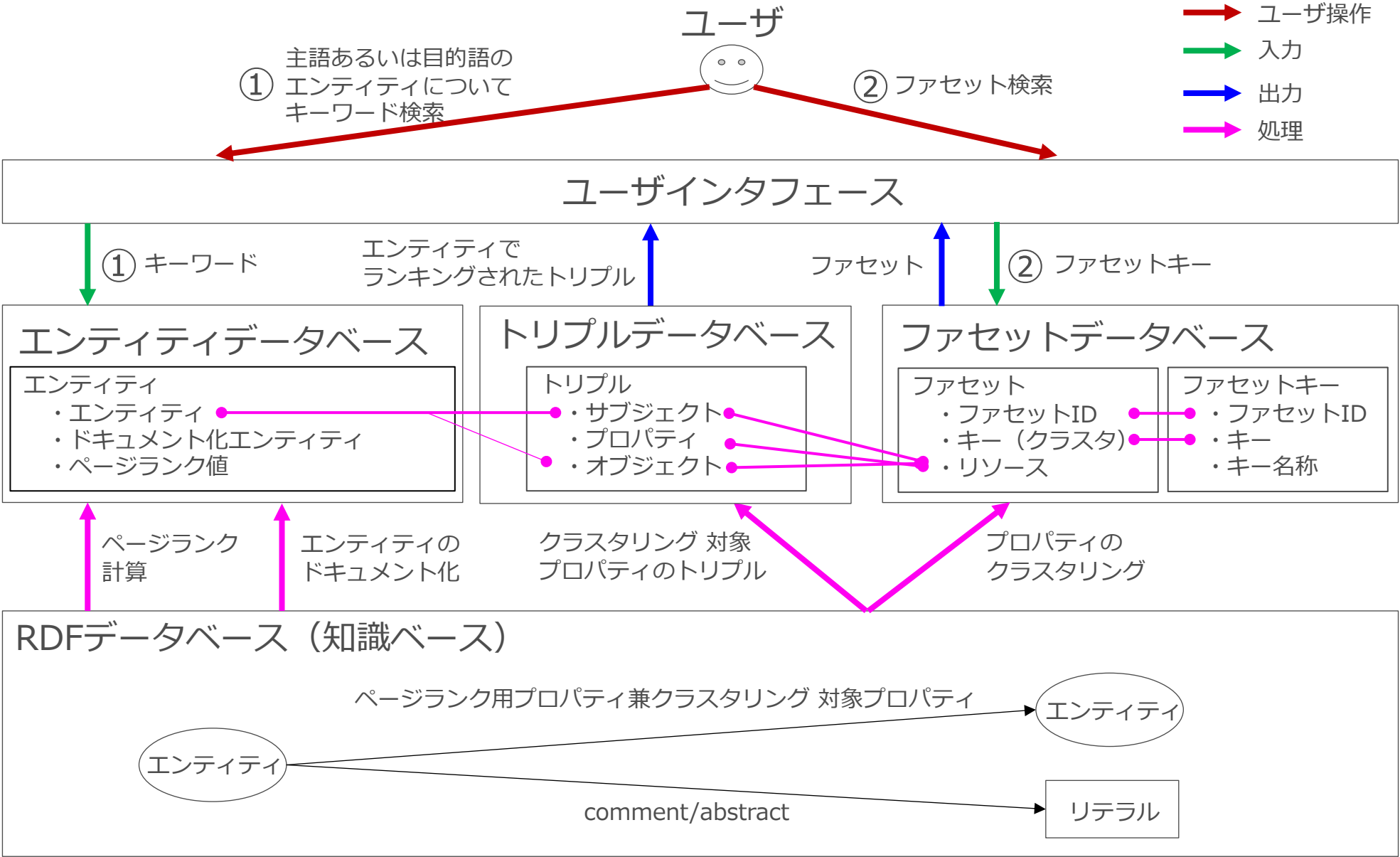
Object Type: [<http://www.w3.org/2002/07/owl#Thing>]

No.	Subject	Predicate	Object
0	<a href="#">Katsunori_Wakabayashi</a>	<a href="#">field</a>	<a href="#">Nanotechnology</a>
1D	<a href="#">Mitsutaka_Fujita</a>	<a href="#">field</a>	<a href="#">Nanotechnology</a>
2	<a href="#">Katsunori_Wakabayashi</a>	<a href="#">award</a>	<a href="#">Physical_Society_of_Japan</a>
3	<a href="#">Mitsutaka_Fujita</a>	<a href="#">award</a>	<a href="#">Physical_Society_of_Japan</a>

## 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. プロパティのクラスタリング 結果について目視による初期的な確認
  2. プロトタイプシステムによるプロパティファセットを用いた検索の確認（ユースケース）
- 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
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#sameSettingAs">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#sameSettingAs</a>	32
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#coparticipatesWith">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#coparticipatesWith</a>	26
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasLocation">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasLocation</a>	17
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isMemberOf">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isMemberOf</a>	7
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isClassifiedBy">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isClassifiedBy</a>	3
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasRole">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasRole</a>	3
<a href="http://dbpedia.org/ontology/location">http://dbpedia.org/ontology/location</a>	2
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isDescribedBy">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isDescribedBy</a>	2
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isPartOf">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isPartOf</a>	2
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isParticipantIn">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#isParticipantIn</a>	2
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasSetting">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasSetting</a>	1
<a href="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#conceptualizes">http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#conceptualizes</a>	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「ユースケース」：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	<a href="#">Harvard_University</a>	<a href="#">campus</a>	<a href="#">Urban_area</a>
1	<a href="#">Harvard_University</a>	<a href="#">city</a>	<a href="#">Cambridge, Massachusetts</a>
2	<a href="#">Harvard_University</a>	<a href="#">president</a>	<a href="#">Drew Gilpin Faust</a>
3	<a href="#">Harvard_University</a>	<a href="#">state</a>	<a href="#">Massachusetts</a>
4	<a href="#">Harvard_University</a>	<a href="#">type</a>	<a href="#">Research_university</a>
5	<a href="#">Harvard_University</a>	<a href="#">affiliation</a>	<a href="#">Association of American Universities</a>
6	<a href="#">Harvard_University</a>	<a href="#">affiliation</a>	<a href="#">Association of Independent Colleges and Universities in N</a>
7	<a href="#">Harvard_University</a>	<a href="#">affiliation</a>	<a href="#">National Association of Independent Colleges and Univers</a>
8	<a href="#">Harvard_University</a>	<a href="#">affiliation</a>	<a href="#">Universities Research Association</a>
9	<a href="#">Harvard_University</a>	<a href="#">athletics</a>	<a href="#">NCAA Division_I</a>
10	<a href="#">Princeton_University</a>	<a href="#">athletics</a>	<a href="#">Eastern Intercollegiate Volleyball Association</a>
11	<a href="#">Princeton_University</a>	<a href="#">athletics</a>	<a href="#">Ivy League</a>
12	<a href="#">Princeton_University</a>	<a href="#">athletics</a>	<a href="#">Middle Atlantic Intercollegiate Sailing Association</a>
13	<a href="#">Princeton_University</a>	<a href="#">athletics</a>	<a href="#">NCAA Division_I</a>
14	<a href="#">Princeton_University</a>	<a href="#">state</a>	<a href="#">New Jersey</a>
15	<a href="#">Princeton_University</a>	<a href="#">soundRecording</a>	<a href="#">Princeton_University_1</a>

## 予備実験②「ユースケース」：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	<a href="#">Harvard_University</a>	<a href="#">city</a>	<a href="#">Cambridge, Massachusetts</a>
1	<a href="#">Harvard_University</a>	<a href="#">affiliation</a>	<a href="#">Association of Independent Colleges and Universities in M</a>
2	<a href="#">Harvard_University</a>	<a href="#">affiliation</a>	<a href="#">National Association of Independent Colleges and Univers</a>
3	<a href="#">Harvard_University</a>	<a href="#">affiliation</a>	<a href="#">Universities Research Association</a>
4	<a href="#">Harvard_University</a>	<a href="#">affiliation</a>	<a href="#">Association of American Universities</a>
5	<a href="#">Harvard_University</a>	<a href="#">athletics</a>	<a href="#">NCAA Division I</a>
6	<a href="#">Harvard_University</a>	<a href="#">campus</a>	<a href="#">Urban area</a>
7	<a href="#">Harvard_University</a>	<a href="#">state</a>	<a href="#">Massachusetts</a>
8	<a href="#">Princeton_University</a>	<a href="#">athletics</a>	<a href="#">Eastern Intercollegiate Volleyball Association</a>
9	<a href="#">Princeton_University</a>	<a href="#">affiliation</a>	<a href="#">Association of American Universities</a>
10	<a href="#">Princeton_University</a>	<a href="#">affiliation</a>	<a href="#">Universities Research Association</a>
11	<a href="#">Princeton_University</a>	<a href="#">affiliation</a>	<a href="#">National Association of Independent Colleges and Univers</a>
12	<a href="#">Princeton_University</a>	<a href="#">city</a>	<a href="#">Princeton, New Jersey</a>
13	<a href="#">Princeton_University</a>	<a href="#">state</a>	<a href="#">New Jersey</a>
14	<a href="#">Princeton_University</a>	<a href="#">campus</a>	<a href="#">Suburb</a>
15	<a href="#">Princeton_University</a>	<a href="#">athletics</a>	<a href="#">NCAA Division I</a>
16	<a href="#">Princeton_University</a>	<a href="#">athletics</a>	<a href="#">Ivy League</a>

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/Region \(10\)](#)
- [http://dbpedia.org/ontology/Region \(10\)](#)
- [http://dbpedia.org/ontology/Region \(10\)](#)
- [http://www.w3.org/2002/07/owl#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