

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

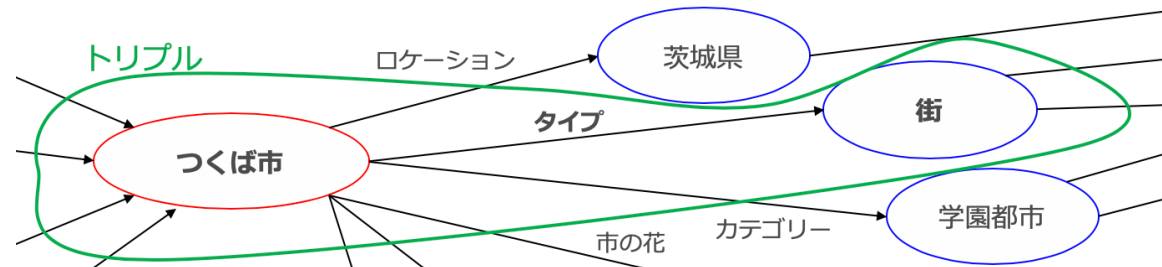
筑波大学

- 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	<a href="#">dbr:Tsukuba_Botanical_Garden</a>	Estación experimental de plantas medicinales de Tsukuba
人？シンボル？	<a href="#">dbr:Sakura_Tsukuba</a>	Sakura Tsukuba
クルーザー？	<a href="#">dbr:Japanese_cruiser_Tsukuba</a>	Tsukuba (Schiff, 1905)
大学	<a href="#">dbr:Tsukuba_University_of_Technology</a> <a href="#">wikidata:Q273863SADF...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> <a href="#">wikidata:Q11602899SC...2-AFD3-4CCD524836BZ</a> <a href="#">wikidata:Q11251506</a> <a href="#">wikidata:Q11602882</a> <a href="#">wikidata:Q11272231</a>	Tsukuba, Ibaraki Tsukuba-class cruisers Tsukuba University Tsukuba club Tsukuba College of Technology つくばセンター

# 研究背景：ファセット検索システム

ユーザはファセットで絞り込んだ結果を確認し、  
どのようなエンティティがあるのか探索できる



Current Query:

Keyword:

Tsukuba

Matching documents: 118

Showing top 50 results

Properties:

ファセット

- ☐ instance of (108 results)
- ☐ country (52 results)
- ☐ located in the administrative territorial entity (21 results)
- ☐ published in (16 results)
- ☐ topic's main category (9 results)
- ☐ sex or gender (8 results)
- ☐ occupation (7 results)
- ☐ country of citizenship (6 results)
- ☐ headquarters location (6 results)
- ☐ follows (6 results)
- ☐ followed by (6 results)

検索結果

(エンティティ一覧)

**Tsukuba**

city in Japan



**University of Tsukuba**

university in Ibaraki Prefecture, Japan

**Category:University of Tsukuba**

Wikimedia category

**Tsukuba Express** TX

Railway line



# 研究背景：エンティティ関係の検索ニーズ

## 例：組織間の関係を調べたい

### Properties:

- + instance of (108 results)
- + country (52 results)
- + located in the administrative territorial entity (21 results)
- + published in (16 results)
- + topic's main category (9 results)
- + sex or gender (8 results)
- + occupation (7 results)
- + country of citizenship (6 results)

ファセット

?

プロパティは数百から数万種類ある

類似プロパティや表記揺れなどもある

“組織間の関係性”を示すようなプロパティを見つけることは難しい

検索結果  
(エンティティ一覧)

Tsukuba

city in Japan



University of Tsukuba

university in Ibaraki Prefecture, Japan

category: University of Tsukuba

Wikimedia category

Tsukuba Express TX

Railway line

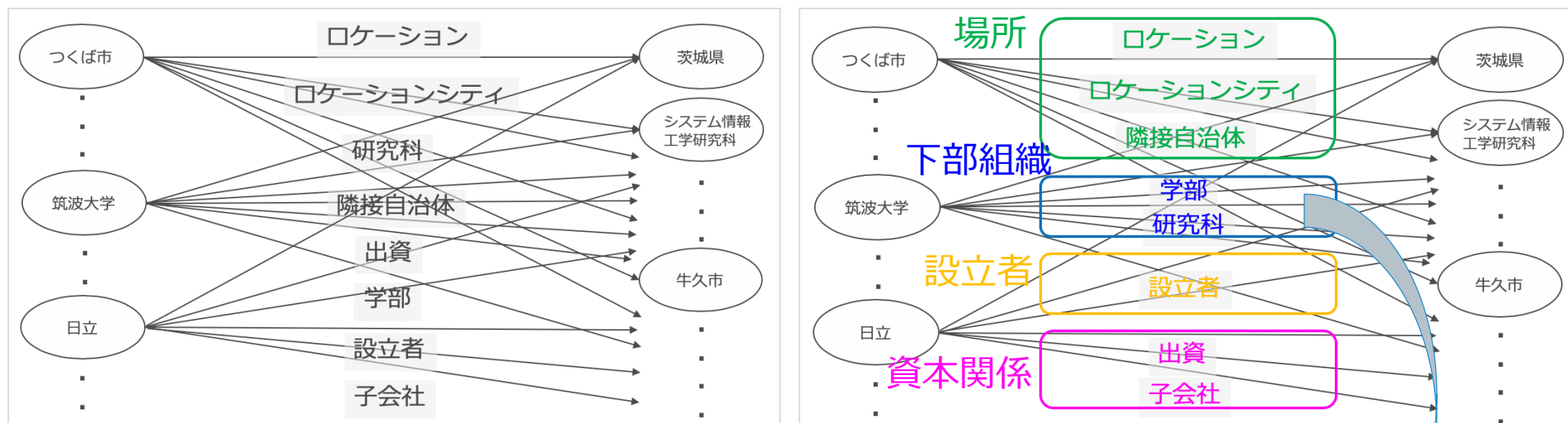


エンティティ指向の検索の場合、  
検索結果のエンティティがどのエンティティと関係するのか一見してわからない

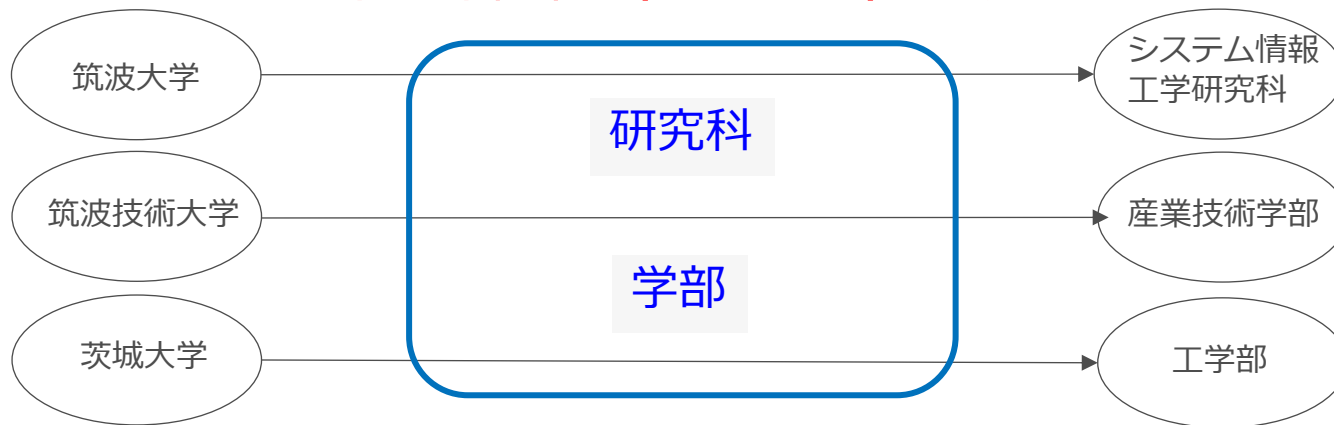
?

# 研究背景：エンティティ関係の検索ニーズに対するアイデア

類似するプロパティでまとめ、  
見つけやすくする



エンティティ関係（トリプル）を提示する



- 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
- 提案システム
  - 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し，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

→エンティティの探索・発見を指向したシステムであるため、本研究の目的とは異なる



## ProFacet

### A)キーワード入力

Keyword:  on ☒ Subject ☐ Object

### B)ファセット選択

Subject Type:

▾

Predicate Type:

▾

Object Type:

▾

Search

### C)検索状態

#### \*\*Search Condition\*\*

Keyword: "Tsukuba" on "Subject"

Subject Type:

[http://dbpedia.org/ontology/Scientist]

Predicate Type:

[academics]

Object Type:

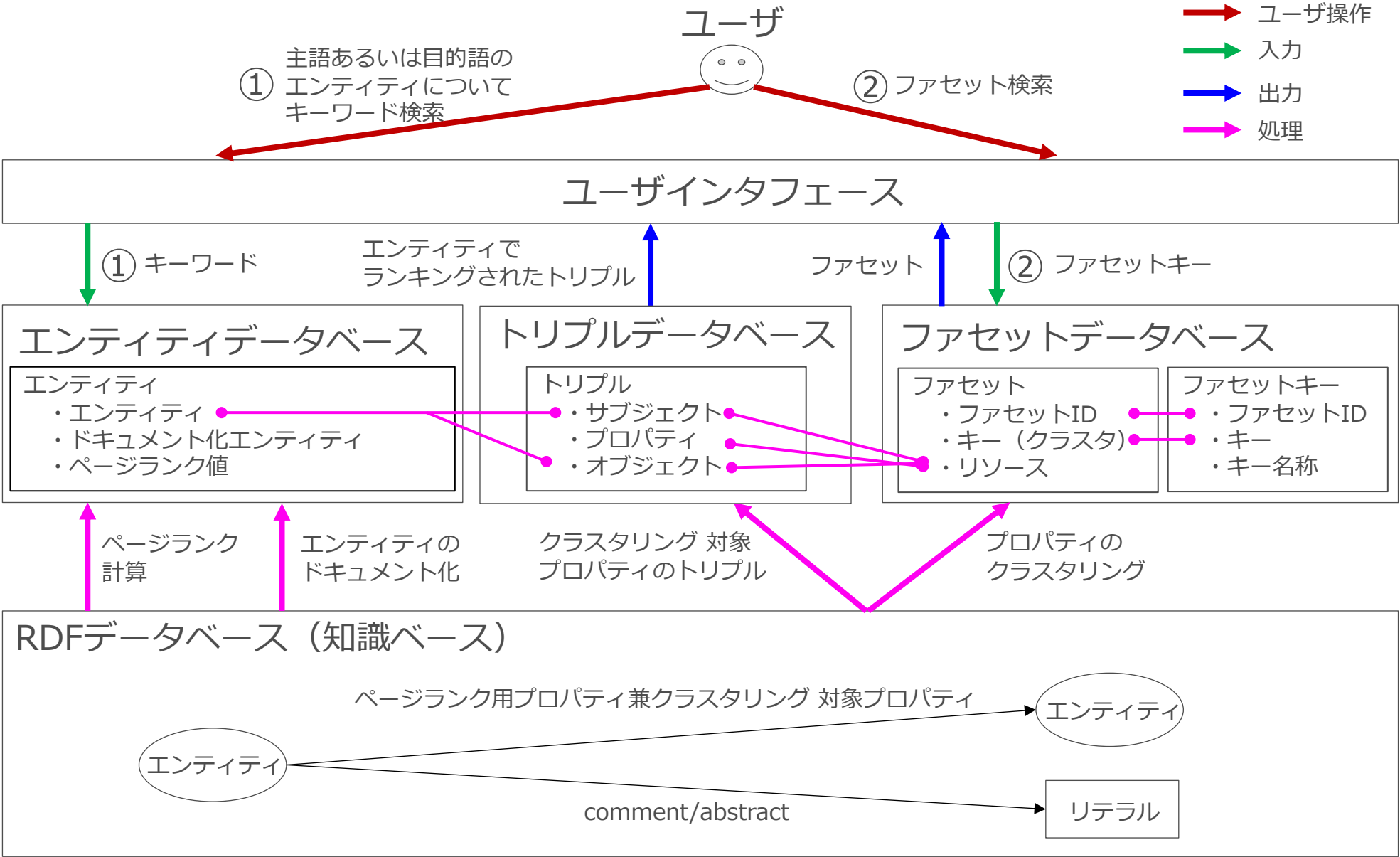
[http://www.w3.org/2002/07/owl#Thing]

Number: 4 records

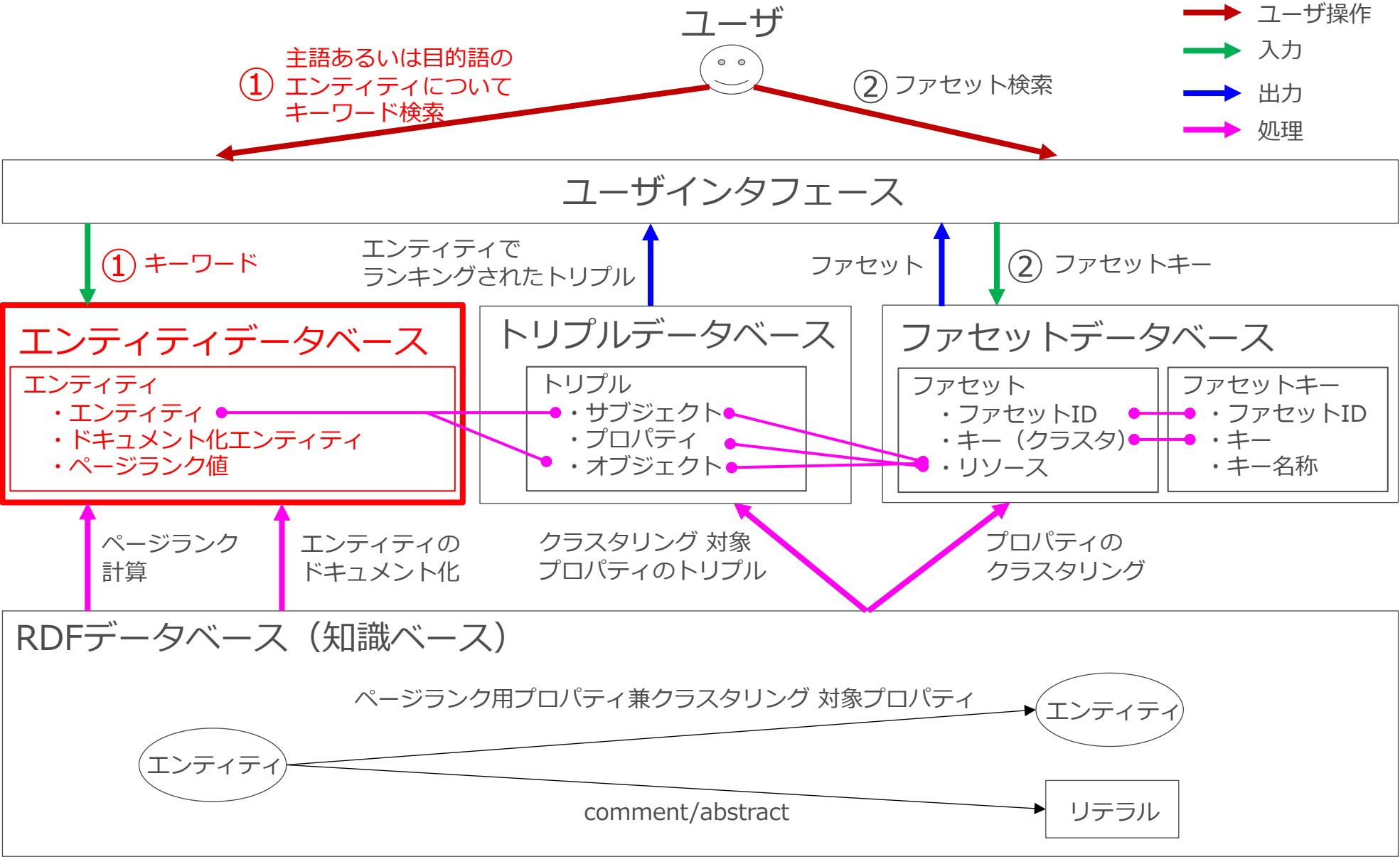
### D)検索結果

No.	Subject	Predicate	Object
0	<a href="#">Katsunori Wakabayashi</a>	<a href="#">field</a>	<a href="#">Nanotechnology</a>
1	<a href="#">Mitsutaka Fujita</a>	<a href="#">field</a>	<a href="#">Nanotechnology</a>
2	<a href="#">Mitsutaka Fujita</a>	<a href="#">award</a>	<a href="#">Physical Society of Japan</a>
3	<a href="#">Katsunori Wakabayashi</a>	<a href="#">award</a>	<a href="#">Physical Society of Japan</a>

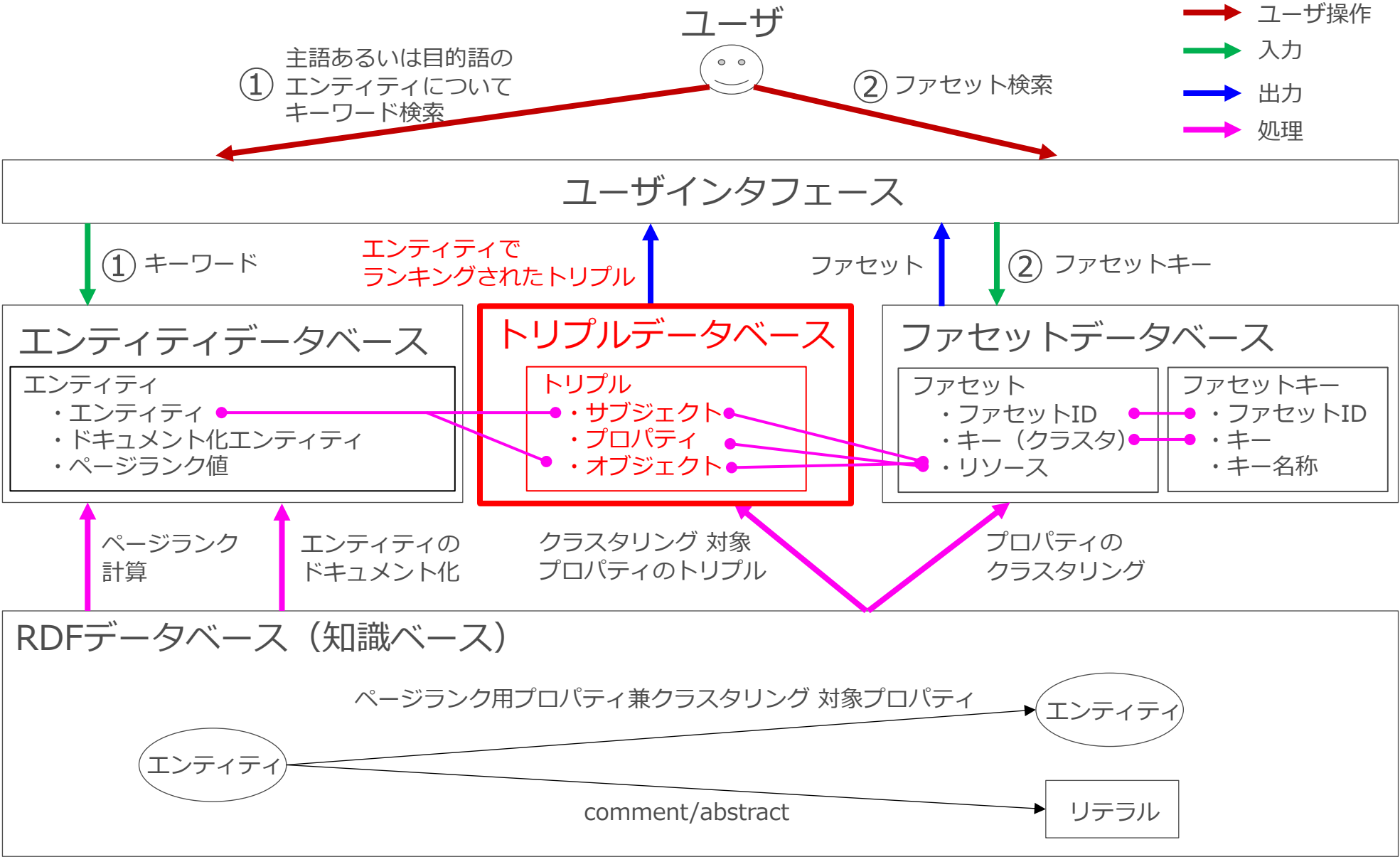
# 提案システム：概要



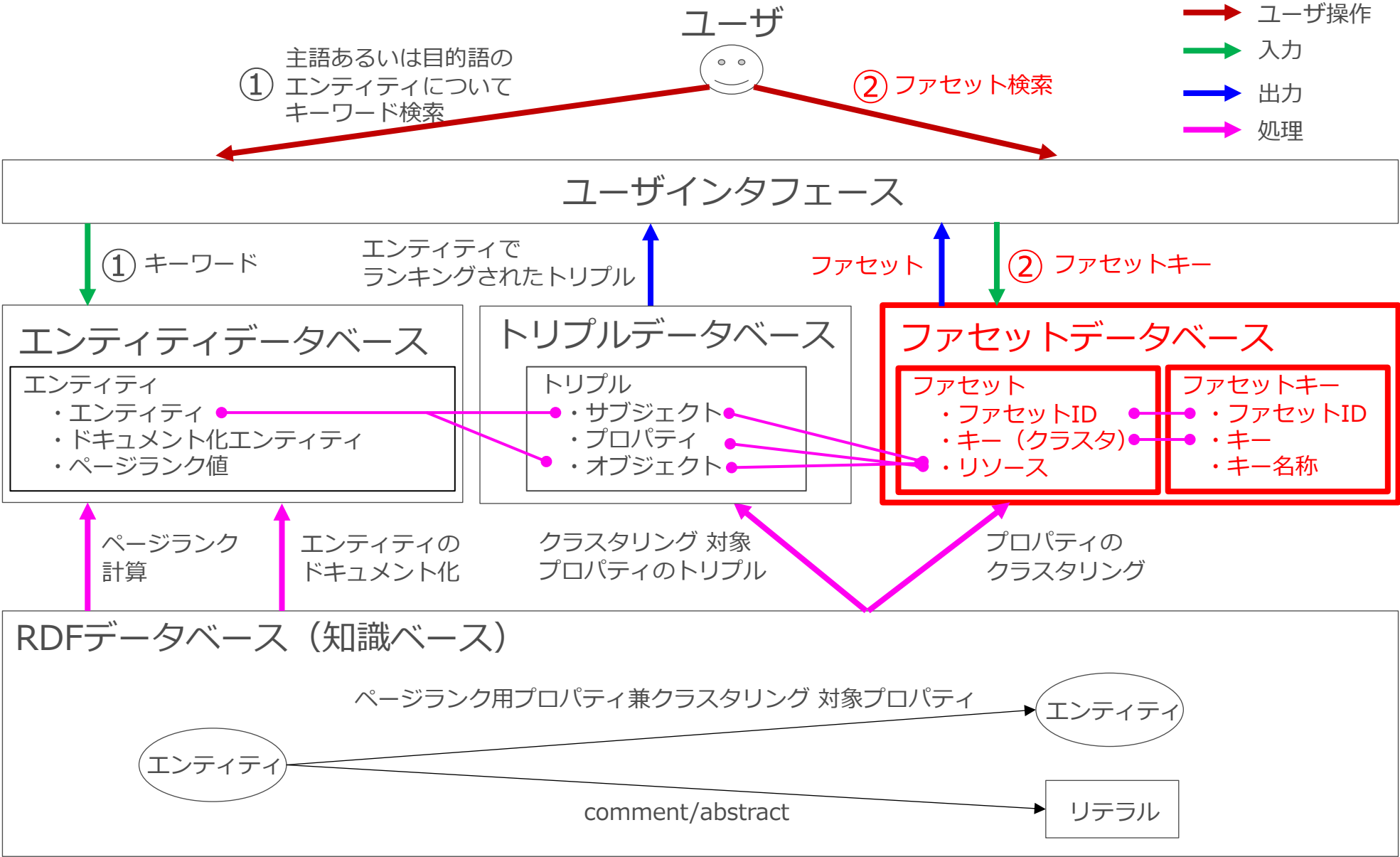
# 提案システム：概要



# 提案システム：概要



# 提案システム：概要



- 目的
    - 元の知識ベースにはない柔軟なファセット検索を可能にするため
  - 提案システムの場合
    - エンティティのファセットテーブル
      - エンティティのクラス情報
    - プロパティのファセットテーブル
      - プロパティのクラスタリング 結果
- } 元の知識ベースにはない情報

- アイデア

1. RDFスキーマで定義されるプロパティ階層構造を利用する

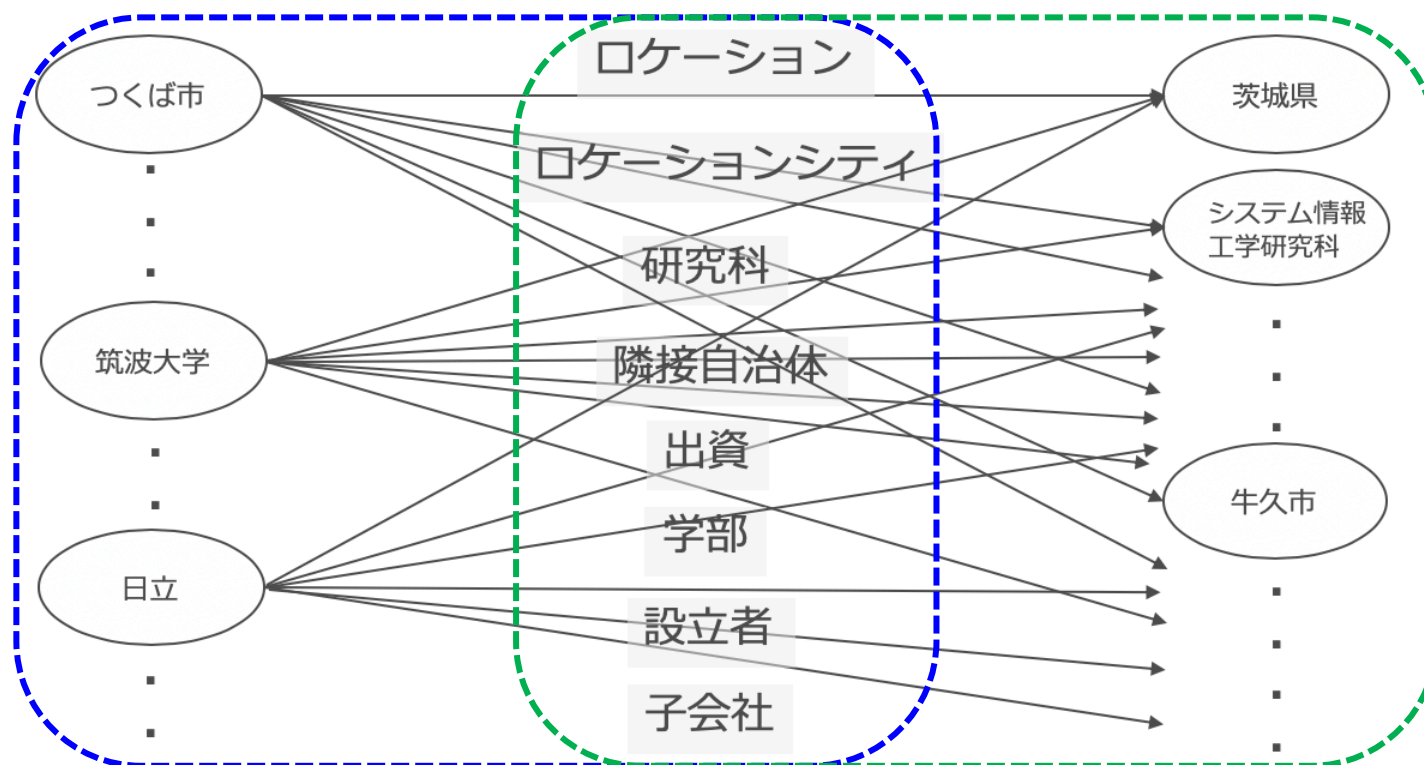
- 上位プロパティを示す“`rdfs:subPropertyOf`”で、各プロパティを上位プロパティでまとめる
- RDFスキーマ定義は期待値であり、定義通りにはなっていない場合もある.
- RDFスキーマ定義の理解が必要になる

2. 実データを基にクラスタリング する

- さまざまなクラスタリング 方法が考えられる
- 実データの状態を反映したクラスタリング となる

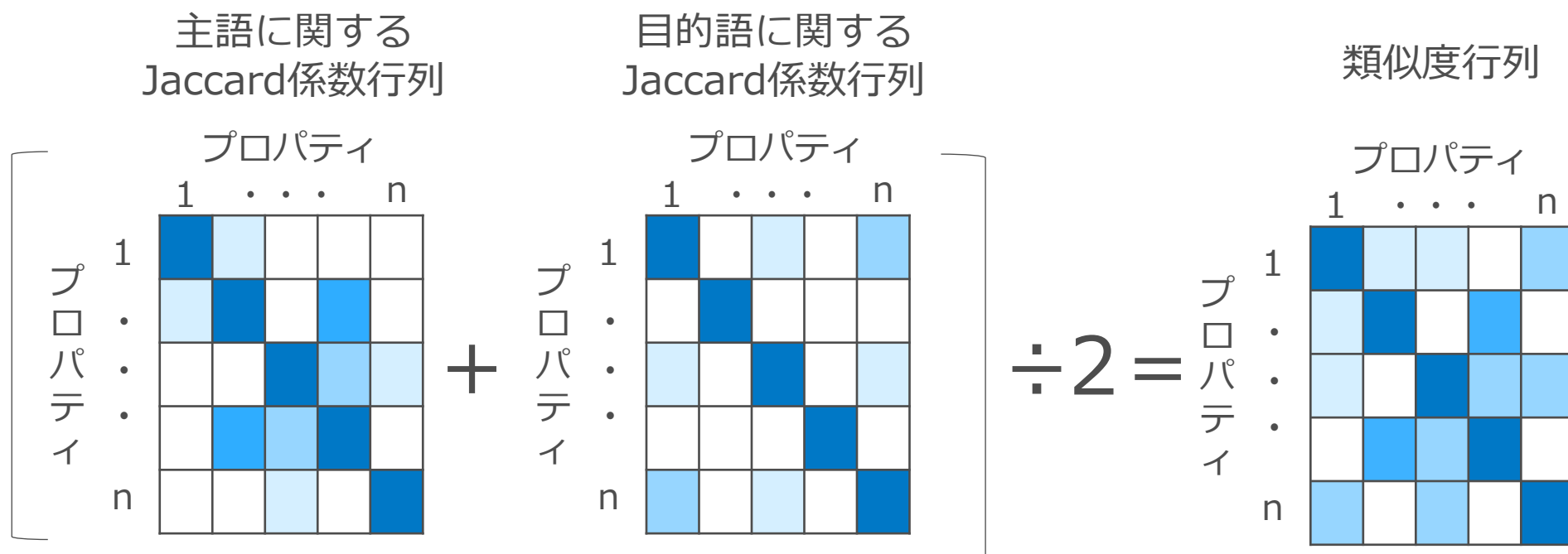
# 提案システム：実データを基にしたクラスタリング

- プロパティの主語, 目的語の重複度合いを類似度としたクラスタリング
  - プロパティの性質によって, プロパティの**主語**, **目的語**は決まるという仮説





- 群平均法による階層型クラスタリング
  - － 類似度の高い要素からまとめるシンプルなアルゴリズム
  - － プロパティの類似度は下記の通り計算した



# 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、直感的に類似するプロパティでまとまっ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. プロパティ階層構造を利用したクラスタリング

- 上位プロパティによるクラスタリング

- 2. 実データに基づいたクラスタリング

- 階層型クラスタリング

## 予備実験①：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

名前空間の異なる  
プロパティが混在  
している

## 予備実験①：上位プロパティによるクラスタリング 結果（続き）

### 最大クラスタの "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

### "sameSettingAs" の定義

- “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

→直感的には、定義と実際のプロパティ要素の関係性を理解することは難しい

## 予備実験①：階層型クラスタリングの結果

<i>clusters</i>	<i>Number of properties</i>
company	20
academics	15
university	11
professionals	11
personalInformation	7
military	6
politics	4
governance	4
relationship	4
organization	3
soundRecording	2
principal	2

<i>clusters</i>	<i>Number of properties</i>
director	2
ethnicity	1
deathCause	1
appointer	1
language	1
dean	1
provost	1
vicePresident	1
officerInCharge	1
depiction	1
partner	1

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。

## 予備実験①：階層型クラスタリングの結果（続き）

### “University”と“Academics” クラスタの要素

#### *Properties in "University" cluster*

affiliation

athletics

campus

chancellor

city

country

head

sport

state

viceChancellor

differentFrom

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  
ロパティでまとまっている

#### *Properties in "Academics" cluster*

academicAdvisor

almaMater

award

birthPlace

citizenship

deathPlace

doctoralAdvisor

doctoralStudent

field

influenced

influencedBy

knownFor

nationality

notableStudent

residence

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

- 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種類のプロパティを含む359個のトリプルに対して、8種類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

**8 種類**

**26種類**

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「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きやすい10種類のプロパティを含む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

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

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>

10種類

- 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
- 他のクラスタリング 手法の適用・評価