



# Stella



# NAVER 통합 Knowledge Graph



김현호 이주현 전희선 NAVER SEARCH

# CONTENTS

0. 소개
1. 통합 Knowledge Graph 시작
2. NAVER DB 통합을 위한 과정
3. How
4. Applications



# 소개

**Daum** | **세종**

통합검색 | 백과사전 | 뉴스 | 어학사전 | 지도 | 블로그 | 브런치

**관련** 세종시 세종대왕 법무법인 세종 세종특별자치시 세종 아파트 세종 코로나 세종 부동산 (주) 세종 세종시청 세종텔레콤 세종 주가 세종 날씨 세종공업 세종

**백과사전**

**세종**  
왕위를 물려받은 셋째 아들 우리나라 사람치고 **세종**을 모르는 사람은 없을 것이다. 보노라면 누구나 **세종**이 미남형의 얼굴에다 인자하다는 느낌을 받게 된다. **세종**(재위 1418~50)은 33년 동안 왕위에 있으면서 바쁜 정무에...  
이미화의 인물한국사

**세종**  
조선전기 제4대(재위:1418~1450) 왕. | 개설 재위 1418~1450, 본관은 전주(全州), **세종**, 자는 원정(元正), 태종의 셋째아들이며, 어머니는 원경왕후 민씨(元敬王后閔氏)이다. 비는 심은(沈溫)의 딸 소헌왕후(昭憲王后)이다. 1408년(태종 8) ...  
한국민족문화대백과사전

**세종**  
태종의 마스터 플랜에 의해 탄생한 왕 **세종**은 1397년(태조 6)에 태종과 원경왕후로 태어났으며, 이름은 도(禔), 자는 원정(元正)이다. 위로는 14년간 세자의 자리에 양녕대군과 **세종**의 세자 책봉 후 불가에 귀의한 효령대...  
조선국왕전

**N** | **세종**

통합 | 뉴스 | 이미지 | VIEW | 지식IN | 인플루언서 | 동영상 | 쇼핑 | 어학사전 | 지도

**지식백과** | 어린이백과 | 학생백과

**세종** 世宗  
조선 제4대 왕

조선전기 제4대(재위:1418~1450) 왕. [개설] 재위 1418~1450. 본관은 전주(全州). 이름은 이도(李祹), 자는 원정(元正). 태종의 셋째아들이며, 어머니는 원경왕후 민씨(元敬王后閔氏)이다. 비는 심은(沈溫)의 딸 소헌왕후(昭憲王后)이다. 1408년(태종 ...  
한국민족문화대백과

**세종** 世宗  
조선왕조 제4대 왕(재위 1418~1450). 인재를 고르게 등용하여 이상적 유교정치를 구현하였다. **세종**대에 훈민정음이 창제되고 측우기와 같은 과학 기구가 제작되는 등 백성들의 생활에 실질적으로 도움이 되는 문화 정책이 추진되었다. 이름은 도...  
두산백과

**이전왕, 조선 제3대 태종**

**다음왕, 조선 제5대 문종**

**세종 다른 뜻**

- 요리 **별기에 농부들이 마시던 맥주**
- 인물 **금나라 제5대 황제**
- 인물 **오대 후주 제2대 황제**
- 인물 **신라의 화랑**
- 인물 **가락국왕 김구해의 장남**

지식백과 더보기 →

**Google** | **세종**

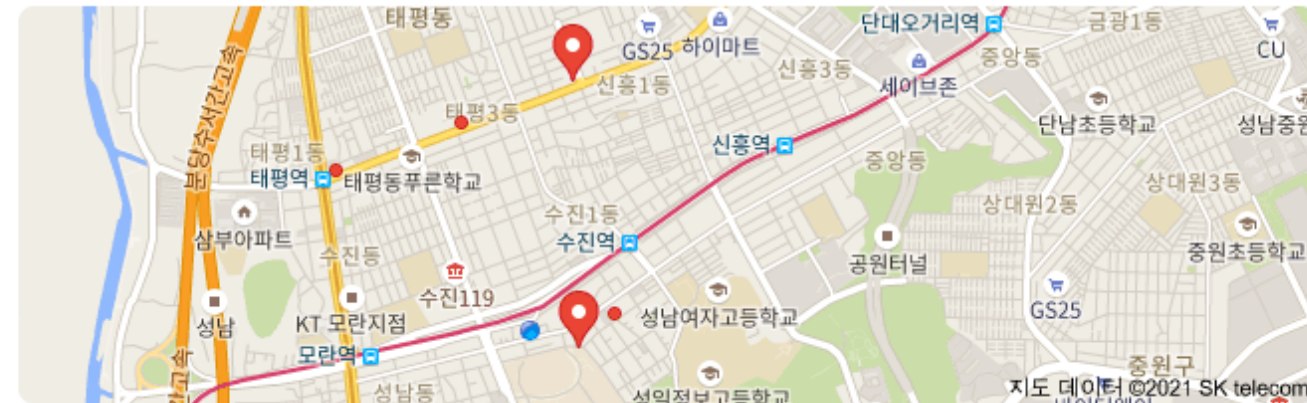
검색결과 약 71,200,000개 (0.98초)

[https://ko.wikipedia.org/wiki/조선\\_세종](https://ko.wikipedia.org/wiki/조선_세종)  
**조선 세종 - 위키백과, 우리 모두의 백과사전**  
세종(世宗, 1397년 5월 7일 ~ 1450년 3월 30일)은 조선의 제4대 국왕(재위 : 1418년 9월 9일 ~ 1450년 3월 30일)이다. 세종의 업적에 대한 존경의 의미를 담아 ...  
사망지: 조선 한성부 영응대군 사저 안동별궁 | 출생지: 조선 한성부 준수방  
시호: 장헌영문예무인성명효대왕(莊憲英文睿武仁聖明孝大王) | 휘: 이도(李祹) (웃소매 도)  
생애 · 치세 · 최후 · 가족 관계

<https://www.sejong.go.kr>  
**세종특별자치시 - 행정수도 세종**  
세종특별자치시,시민주권 특별자치시 행정수도 세종.

<https://namu.wiki> | 세종(조선)  
**세종(조선) - 나무위키:대문**  
세종 이도는 조선이 건국된 해인 1392년에서 5년이 지난 1397년에 태어났다.[9] 고려 왕씨 왕조의 군인으로 지내다가 역성혁명으로 나라를 뒤엎고 왕위에 올랐던 ...  
2021. 3. 16. · 업로드: KBS역사저널 그날  
세종(조선)/창작물 · 김기창 · 소헌왕후 · 성군

**세종**



# 소개



음악 DB

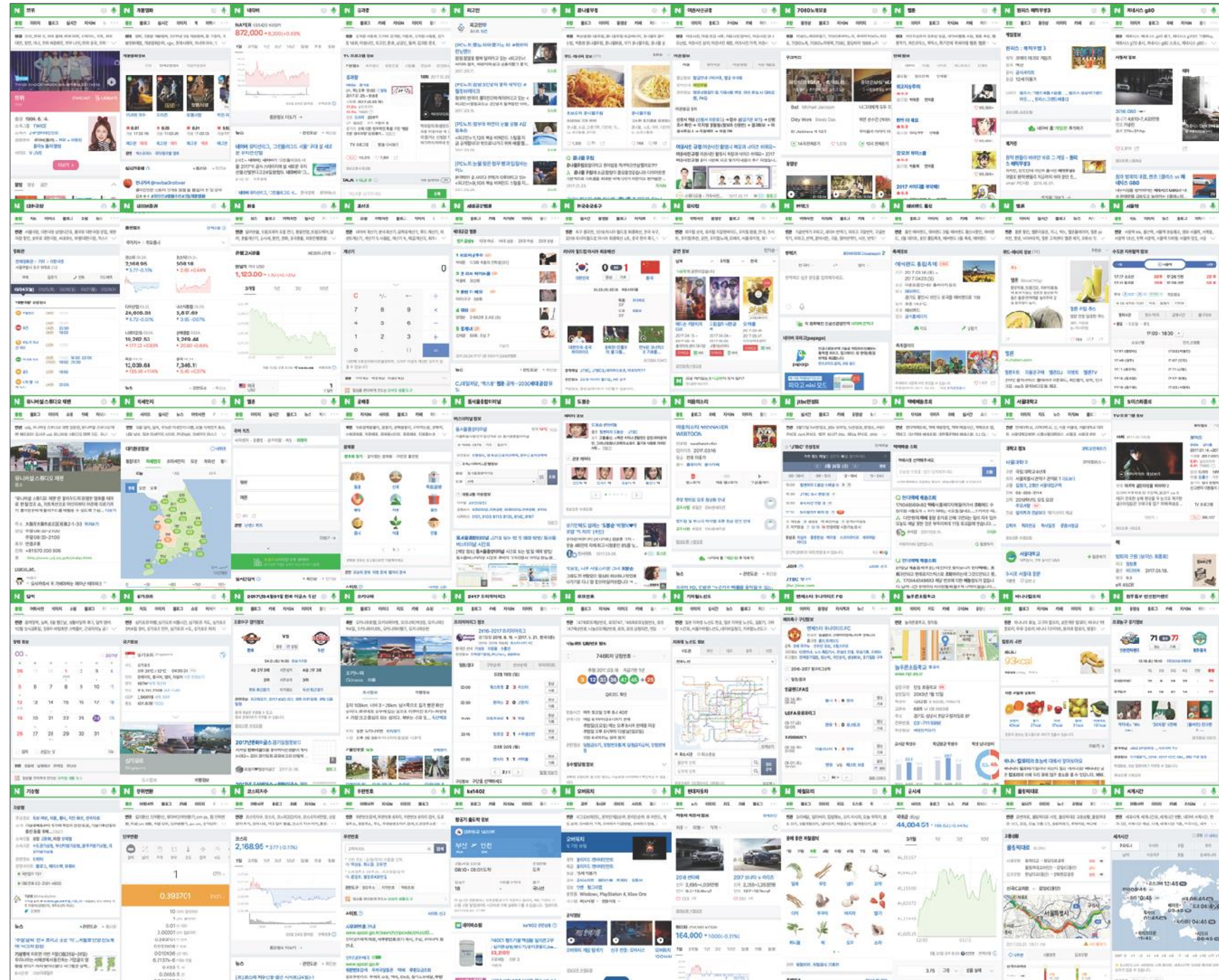
인물 DB

방송 DB

뉴스 문서

# 1. 시작

# <2019 DEVIEW 중>



DEVIEW 2019

40여개 주제 250여개 컬렉션으로 서비스 중.

통합검색 전체 질의어 중 약 30% 대응 중.

# 1.1 통합 Knowledge Graph 시작

## {JSON}

JavaScript Object Notation

```
{
  "userID" : 1,
  "isActive": false,
  "identity":
  {
    "lastName" : "Doe",
    "firstName" : "John",
    "birthDate" : "01/02/1975",
    "socialSecurityNumber" : "331-80-5651"
  },
  "address":
  {
    "number" : "01",
    "street" : "2068 Thomas Street",
    "city" : "Roselle",
    "zipCode" : "60172",
    "state" : "IL"
  },
  "contact":
```

```
"data": {
  "city": "Paris",
  "state": "Ile-de-France",
  "country": "France",
  "location": {
    "type": "Point",
    "coordinates": [2.351666, 48.859425]
  }
}
```

position = 1

`["data"]["location"]["coordinates"][1] = 48.859425`

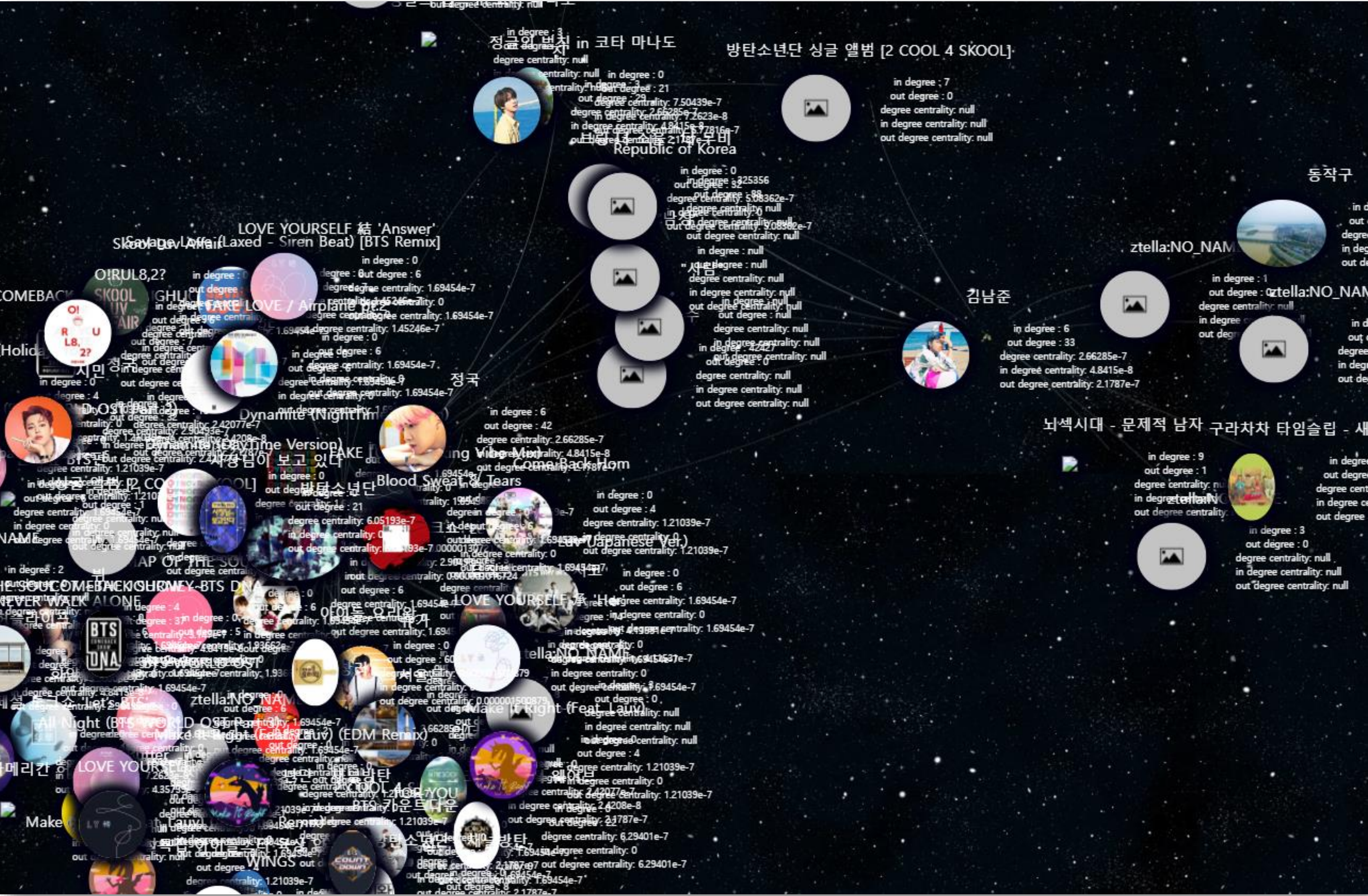
The diagram illustrates the path to a specific value in a nested JSON object. Red boxes highlight the keys 'data', 'location', and 'coordinates'. Red arrows point from these keys down to the corresponding parts of the path expression `["data"]["location"]["coordinates"][1]`. A red arrow also points from the second element of the coordinate array, `48.859425`, to the `[1]` index in the path, with the text `position = 1` next to it.

# 1.1 통합 Knowledge Graph 시작





# 1.2 Stella, Knowledge Graph



## 2. NAVER DB 통합을 위한 과정

## 2.1 과정

### Ontology

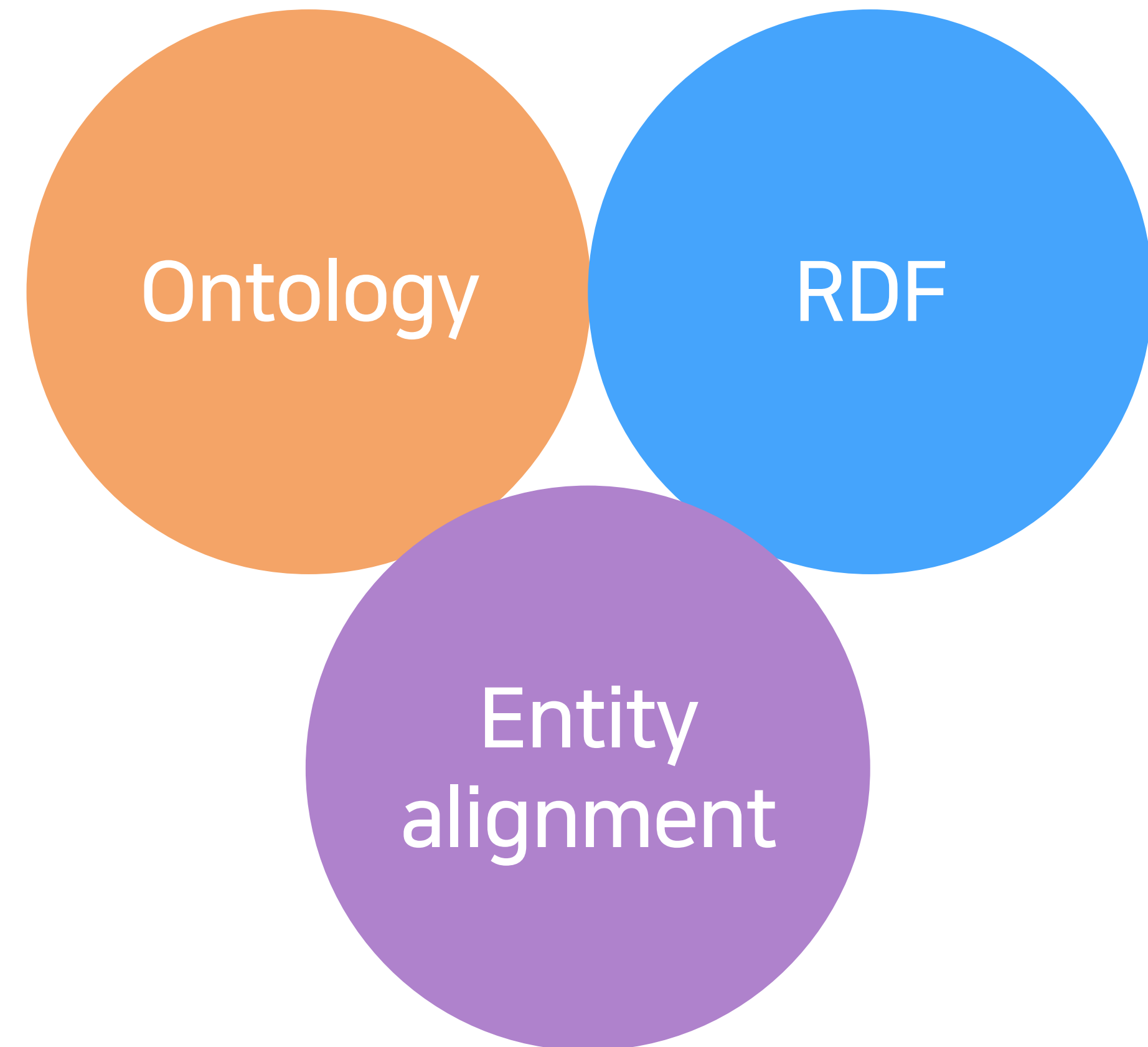
- Taxonomy (class hierarchy)
- Relations, Restrictions

### RDF 형태로 변환

- Resource Description Framework
- URI (Uniform Resource Identifier)

### Entity Alignment

- 같은 Entity를 찾아서 연결
- DB Linkage / Entity Alignment / Entity Resolution



# 3. Ontology

# 3.1 Ontology

## Ontology

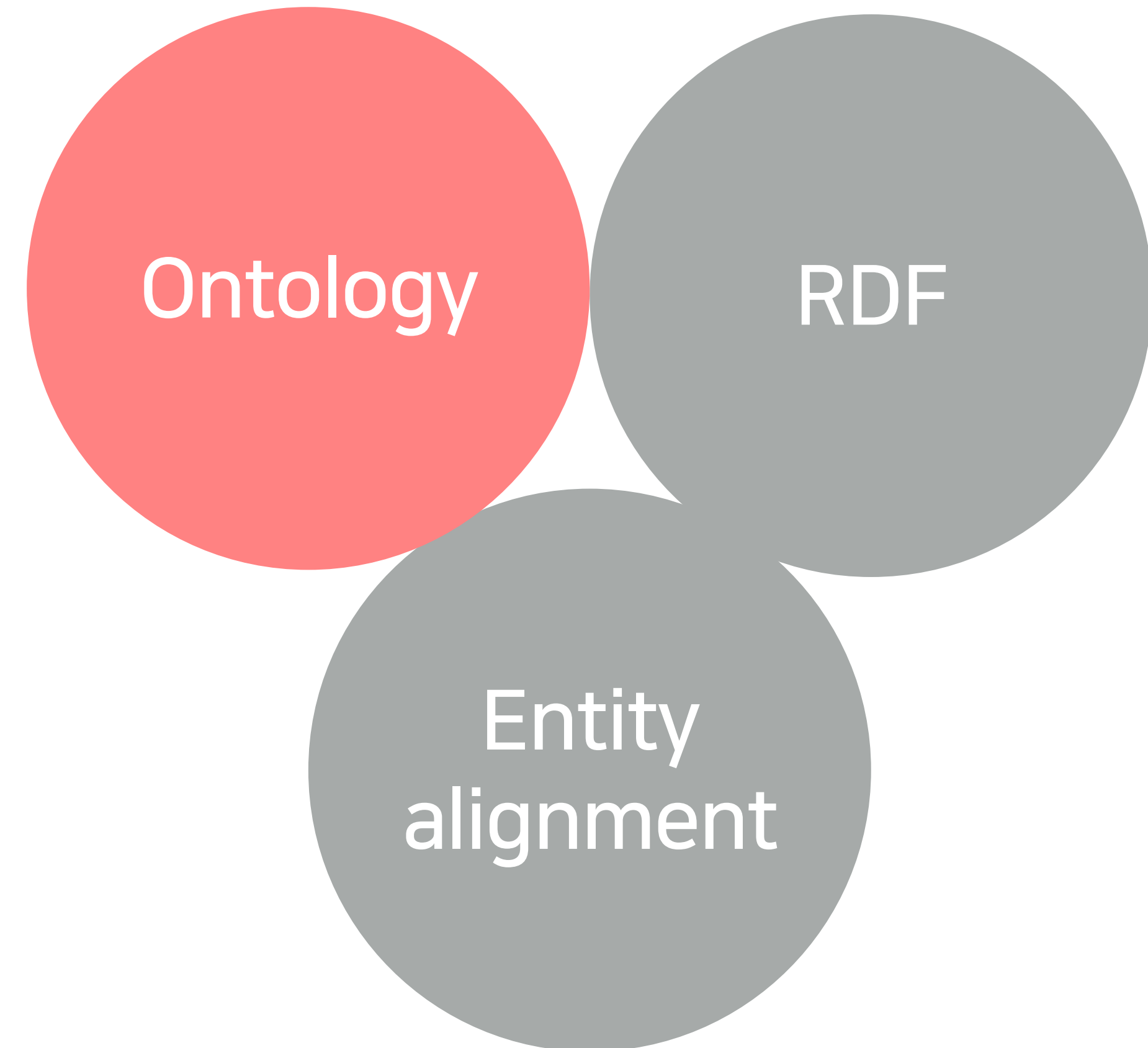
- Taxonomy (class hierarchy)
- Relations, Restrictions

## RDF 형태로 변환

- Resource Description Framework
- URI (Uniform Resource Identifier)

## Entity Resolution

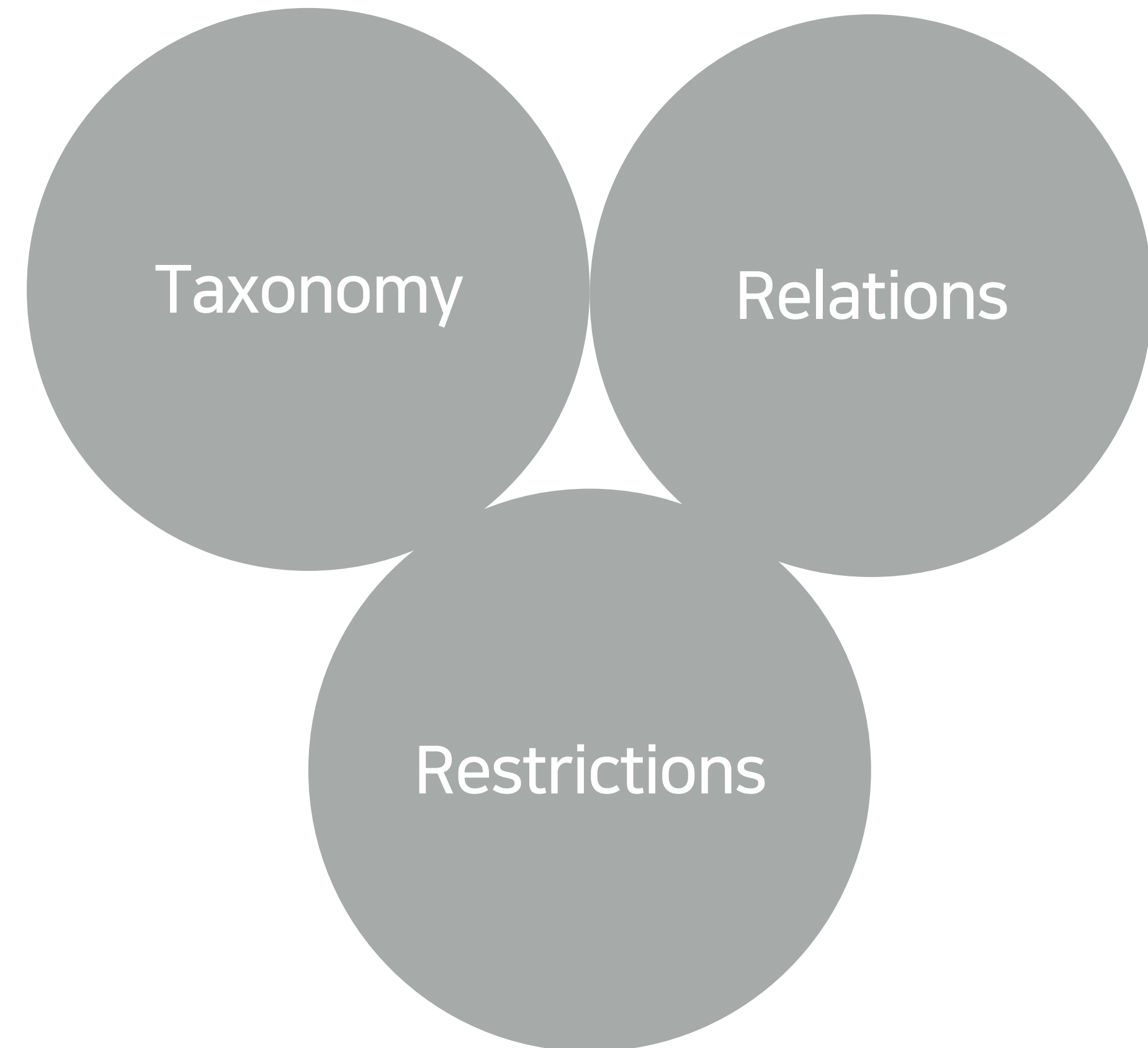
- 같은 Entity를 찾아서 연결
- DB Linkage / Entity Alignment / Entity Resolution



# 3.1 Ontology

## Ontology

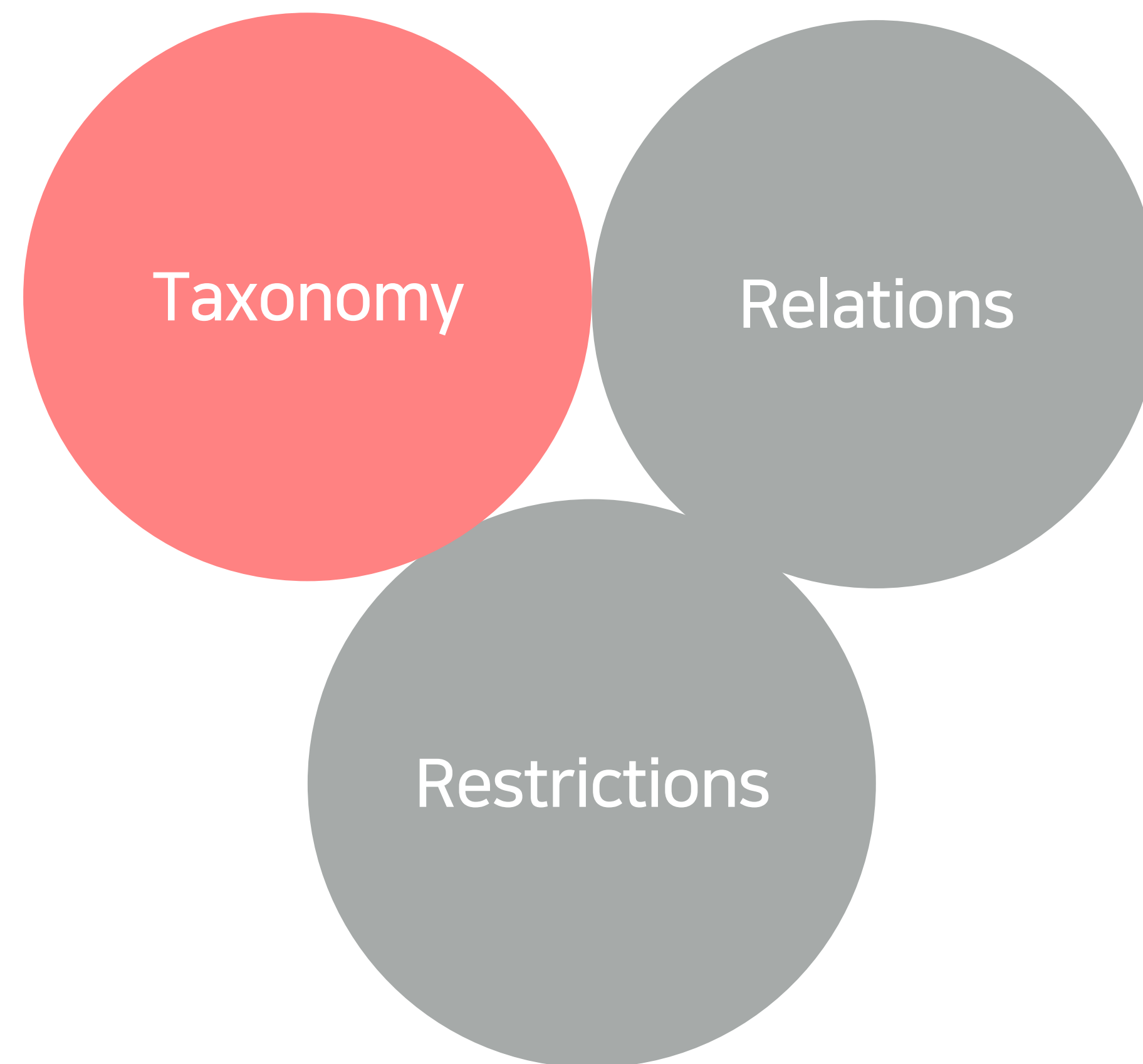
- Taxonomy
- Relations
- Restrictions



# 3.1 Taxonomy

## Ontology

- Taxonomy
- Relations
- Restrictions

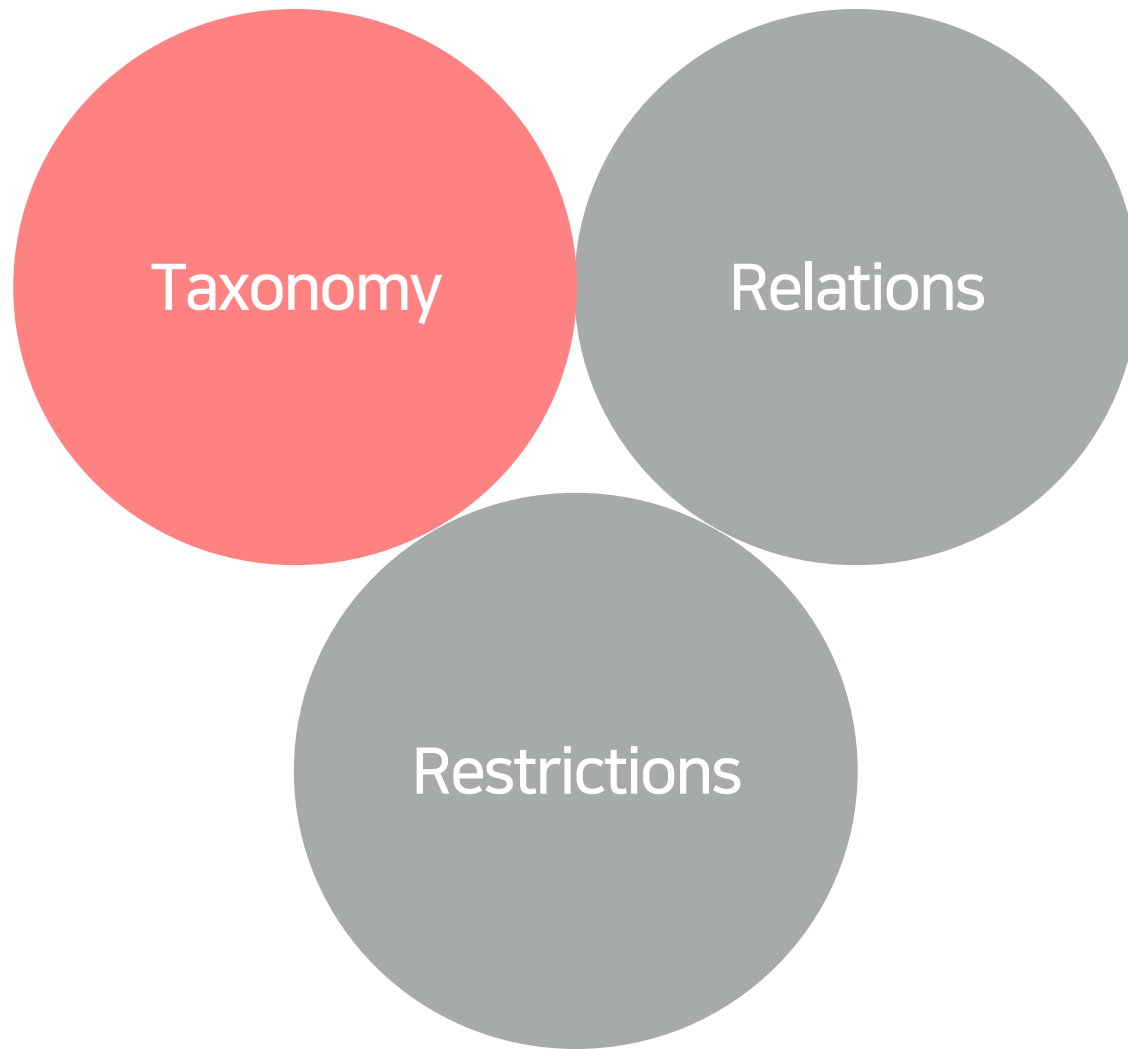


# 3.1 Taxonomy

- ▼ Thing -
  - ▶ Action +
  - ▶ BioChemEntity +
  - ▶ CreativeWork +
  - ▶ Event +
  - ▶ Intangible +
  - ▶ MedicalEntity +
  - ▶ Organization +
  - ▶ Person +
  - ▶ Place +
  - ▶ Product +
  - Taxon

```

planet of the Solar System (Q17362350) •2 ↑
├── outer planet (Q30014) •25 ↑
│   ├── Saturn (Q193)
│   ├── Jupiter (Q319)
│   ├── Uranus (Q324)
│   └── Neptune (Q332)
└── inner planet (Q3504248) •8 ↑
    ├── Earth (Q2)
    ├── Mars (Q111)
    ├── Mercury (Q308)
    └── Venus (Q313)
  
```





# 3.1 Taxonomy

제33회 한글 및 한국어 정보처리 학술대회 논문집 (2021년)

## 방향 비순환 그래프의 중심성을 이용한 위키데이터 기반 분류체계 구축

전희선<sup>o</sup>, 김현호, 강인호

네이버

{heeseon.cheon, kim.hh, once.ihkang}@navercorp.com

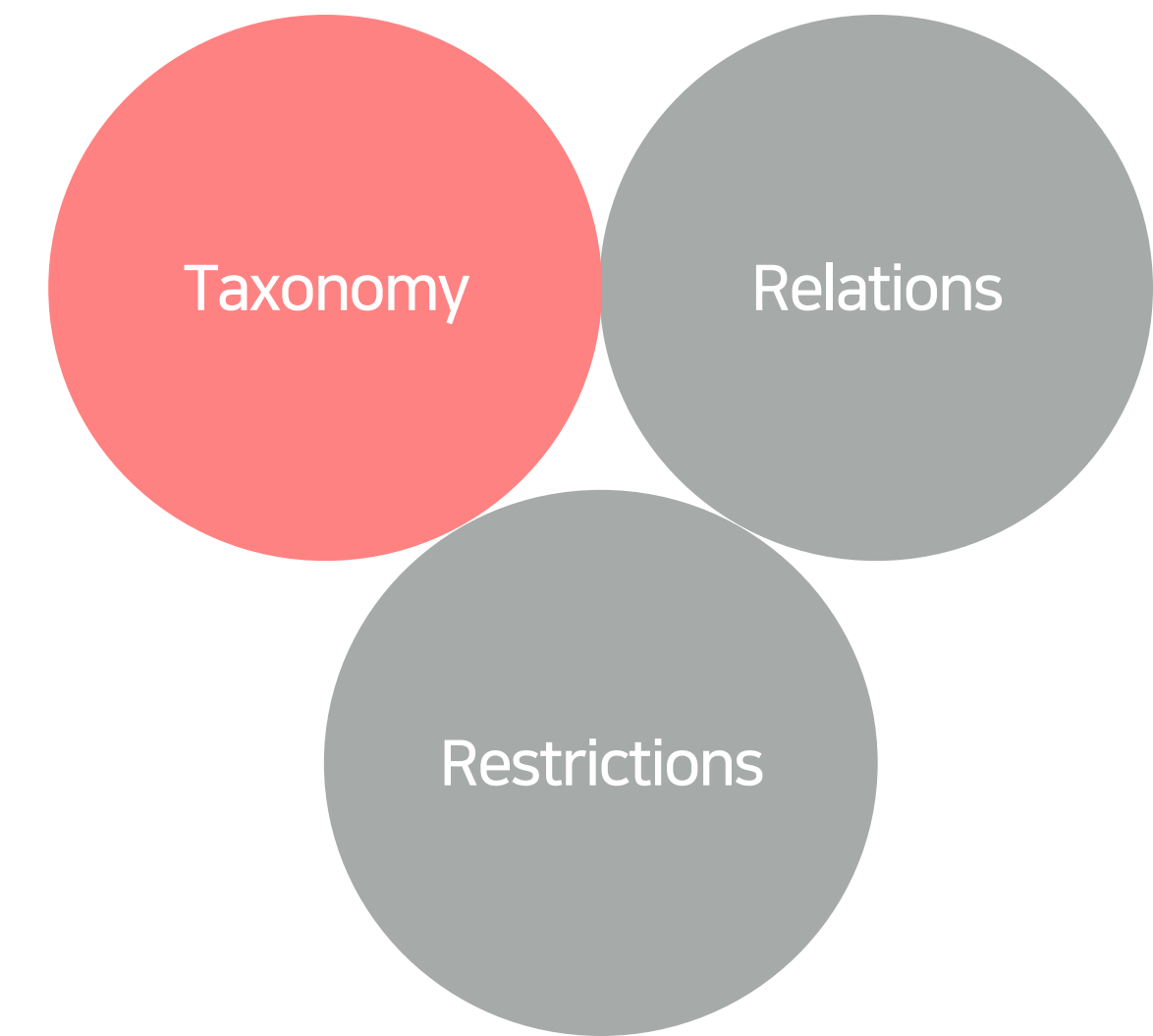
## Taxonomy Induction from Wikidata using Directed Acyclic Graph's Centrality

Hee-Seon Cheon<sup>o</sup>, Hyun-Ho Kim, Inho Kang

NAVER Corp.

### 요약

한국어 통합 지식베이스를 생성하기 위해 필수적인 분류체계(taxonomy)를 구축하는 방식을 제안한다. 위키데이터를 기반으로 분류 후보군을 추출하고, 상하위 관계를 통해 방향 비순환 그래프(Directed Acyclic Graph)를 구성한 뒤, 국부적 도달 중심성(local reaching centrality) 등의 정보를 활용하여 정제함으로써 246 개의 분류와 314 개의 상하위 관계를 갖는 분류체계를 생성한다. 워드넷(WordNet), 디비피디아(DBpedia) 등 기존 링크드 오픈 데이터의 분류체계 대비 깊이 있는 계층 구조를 나타내며, 다중 상위 분류를 지닐 수 있는 비트리(non-tree) 구조를 지닌다. 또한, 위키데이터 속성에 기반하여 위키데이터 정보가 있는 인스턴스(instance)에 자동으로 분류를 부여할 수 있으며, 해당 방식으로 실험한 결과 99.83%의 분류 할당 커버리지(coverage) 및 99.81%의 분류 예측 정확도(accuracy)를 나타냈다.



# 3.2 Taxonomy – 기존 분류체계 구축 방식

## 1. 수동 구축

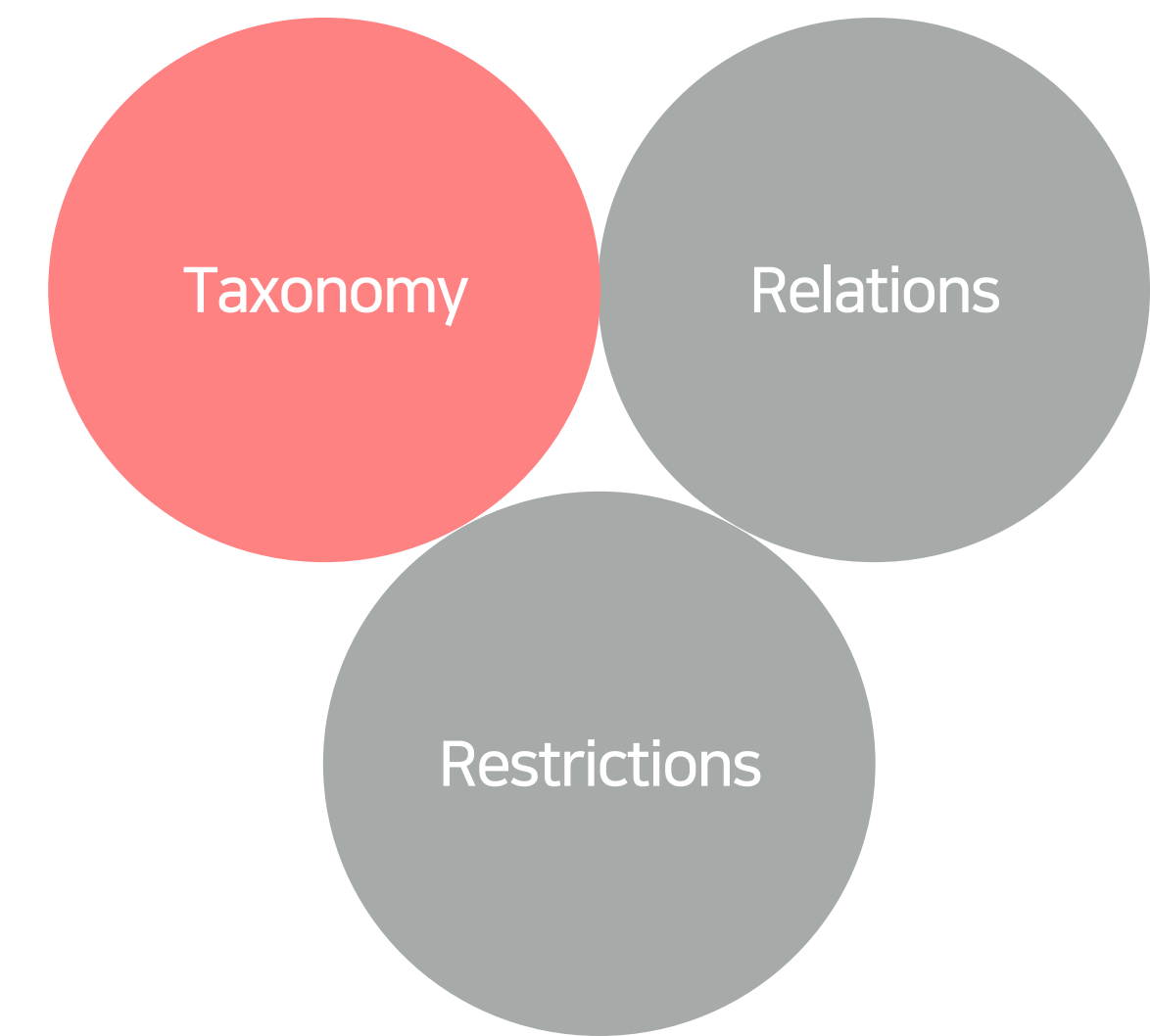
- 전문가들에 의해 수동으로 분류체계 생성
- ex : WordNet, SUMO 등

## 2. 위키피디아 분류 네트워크 (Wikipedia Category Network, WCN) 이용

- 위키피디아 내 존재하는 분류체계 이용
- “subclass of / not subclass of” 로 나누는 등의 방식으로 새로운 분류체계 구축
- ex : WikiTaxonomy, WiBi(Wikipedia Bitaxonomy)

## 3. 말뭉치(text corpora) 이용

- 상하위(subclass of) 관계를 나타내는 일정 패턴 정의 후, 말뭉치에서 해당 패턴 추출
- ex : 허스트 패턴(Hearst pattern)



# 3.2 Taxonomy – 기존 분류체계 구축 방식

## 1. 수동 구축

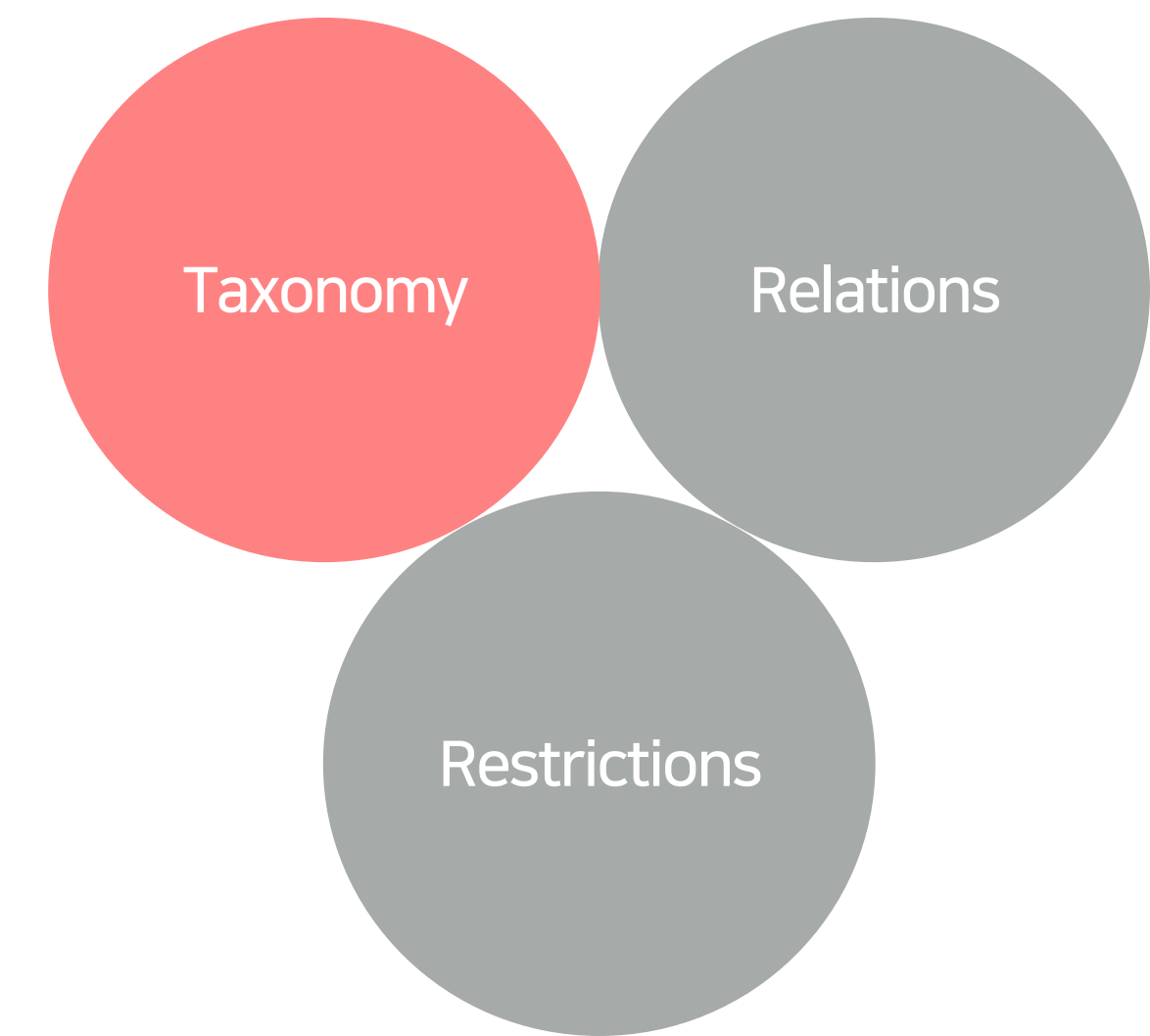
- 전문가들에 의해 수동으로 분류체계 생성
- ex : WordNet, SUMO 등

## 2. 위키피디아 분류 네트워크 (Wikipedia Category Network, WCN) 이용

- 위키피디아 내 존재하는 분류체계 이용
  - “subclass of / not subclass of” 로 나누는 등의 방식으로 새로운 분류체계 구축
  - ex : WikiTaxonomy, WiBi(Wikipedia Bitaxonomy)
- “위키데이터(wikidata)” 기반으로 진행

## 3. 말뭉치(text corpora) 이용

- 상하위(subclass of) 관계를 나타내는 일정 패턴 정의 후, 말뭉치에서 해당 패턴 추출
- ex : 허스트 패턴(Hearst pattern)



# 3.3 후보군 추출

레이블  
(label)

Item Discussion Read View history Search Wikidata

**human** (Q5) 식별자 (item identifier)

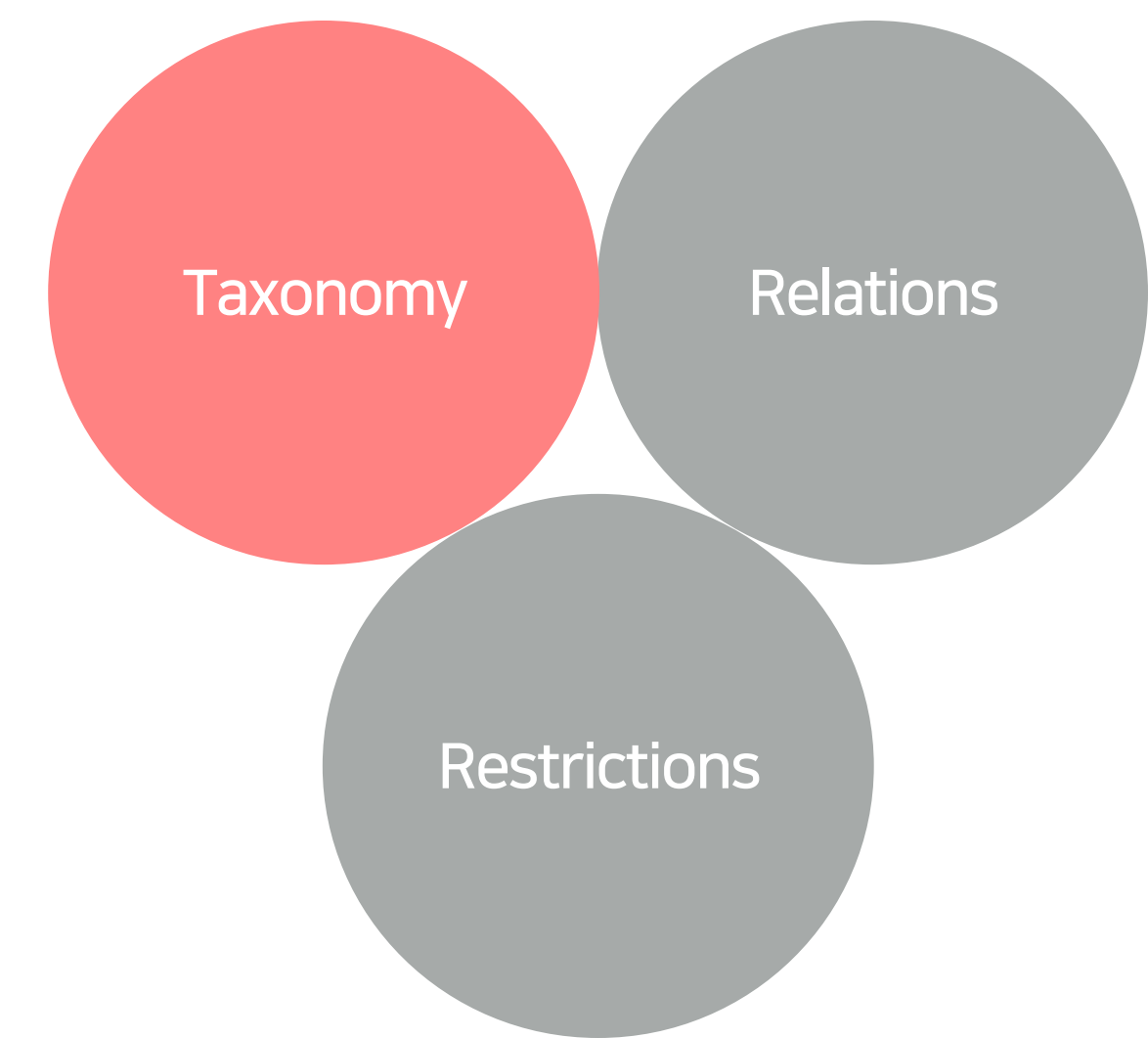
common name of Homo sapiens, unique extant species of the genus Homo  
human being | humankind | people | homosapiens | person | mankind | peoplekind | personkind | persons | humans | Human

▶ In more languages

## Statements

instance of	<ul style="list-style-type: none"> <li>organisms known by a particular common name                     <ul style="list-style-type: none"> <li>of                             <ul style="list-style-type: none"> <li>Homo sapiens</li> <li>Homo sapiens sapiens</li> <li>Homo</li> </ul> </li> </ul> </li> </ul>	값(value)
subclass of	<ul style="list-style-type: none"> <li>person</li> </ul>	
	<ul style="list-style-type: none"> <li>natural person</li> </ul>	
	<ul style="list-style-type: none"> <li>omnivore</li> </ul>	

속성  
(property)



# 3.3 후보군 추출

## Category:Television producers (Q6967490)

Wikimedia category

▼ In more languages  
Configure

Language	Label	Description
English	Category:Television producers	Wikimedia category
Korean	분류:텔레비전 프로듀서	위키미디어 분류

All entered languages

### Statements

instance of Wikimedia category  
▶ 1 reference

---

category contains human  
occupation television producer  
▼ 0 references

-한국어 label을 지닌 wikidata item 중,

label이 "분류:~"로 시작하는 item : 25.4% (307,609 / 1,213,405)

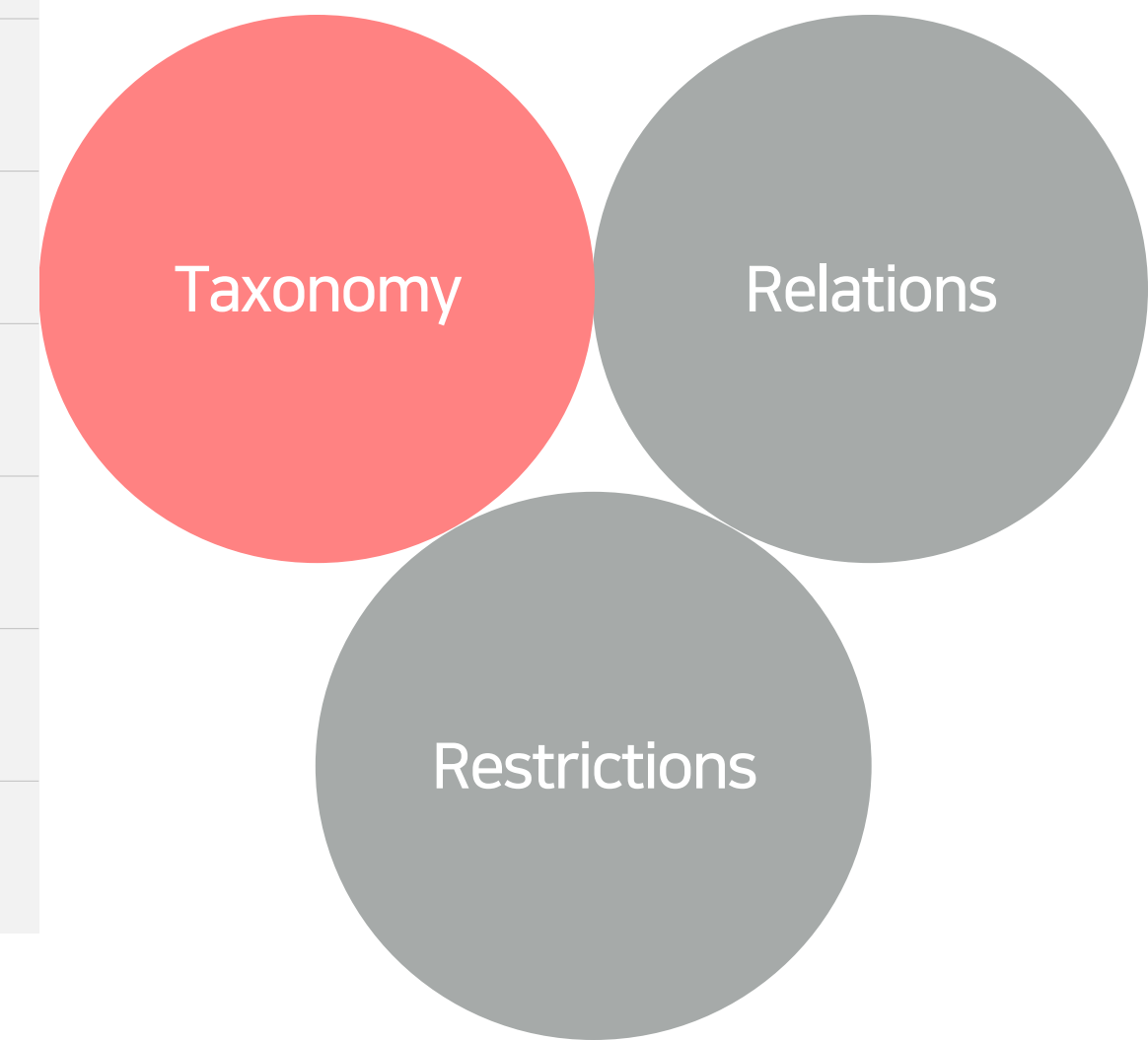
-해당 item의 property 중 "category contains (P4224)"의 value 이용

: 분류가 어떤 item에 속하는지 나타냄

-총 967개 분류 후보군 추출

(label이 "분류:", "틀:", "위키"로 시작하거나, 비어있는 경우는 제거)

분류명	등장 빈도 수
사람 (human, Q5)	101073 (78.46%)
영화 (film, Q11424)	7695 (5.97%)
분류군 (taxon, Q16521)	4555 (3.54%)
앨범 (album, Q482994)	1816 (1.41%)
구조물 (architectural structure, Q811979)	1598 (1.24%)
노래 (song, Q7366)	1204 (0.94%)
...	...

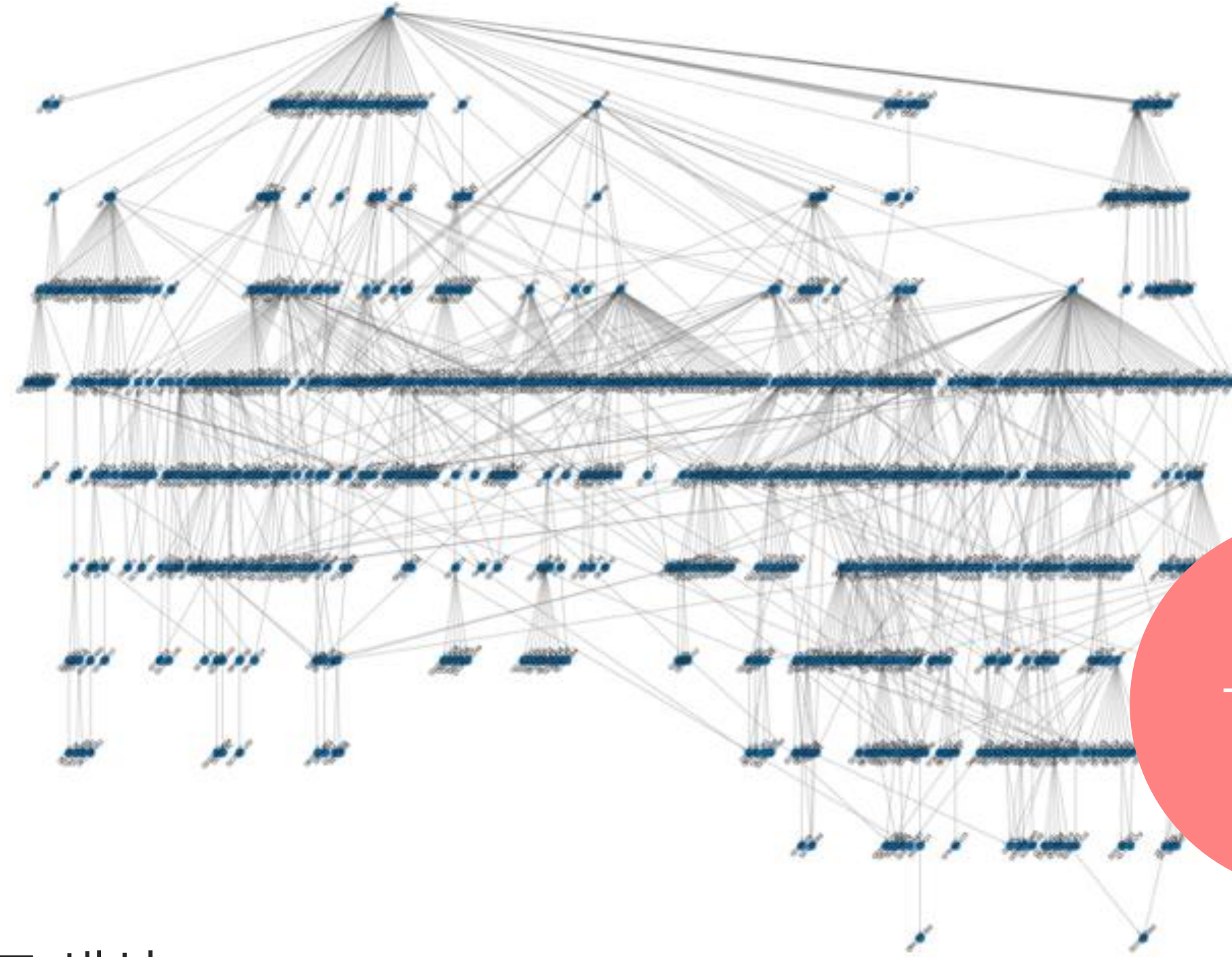
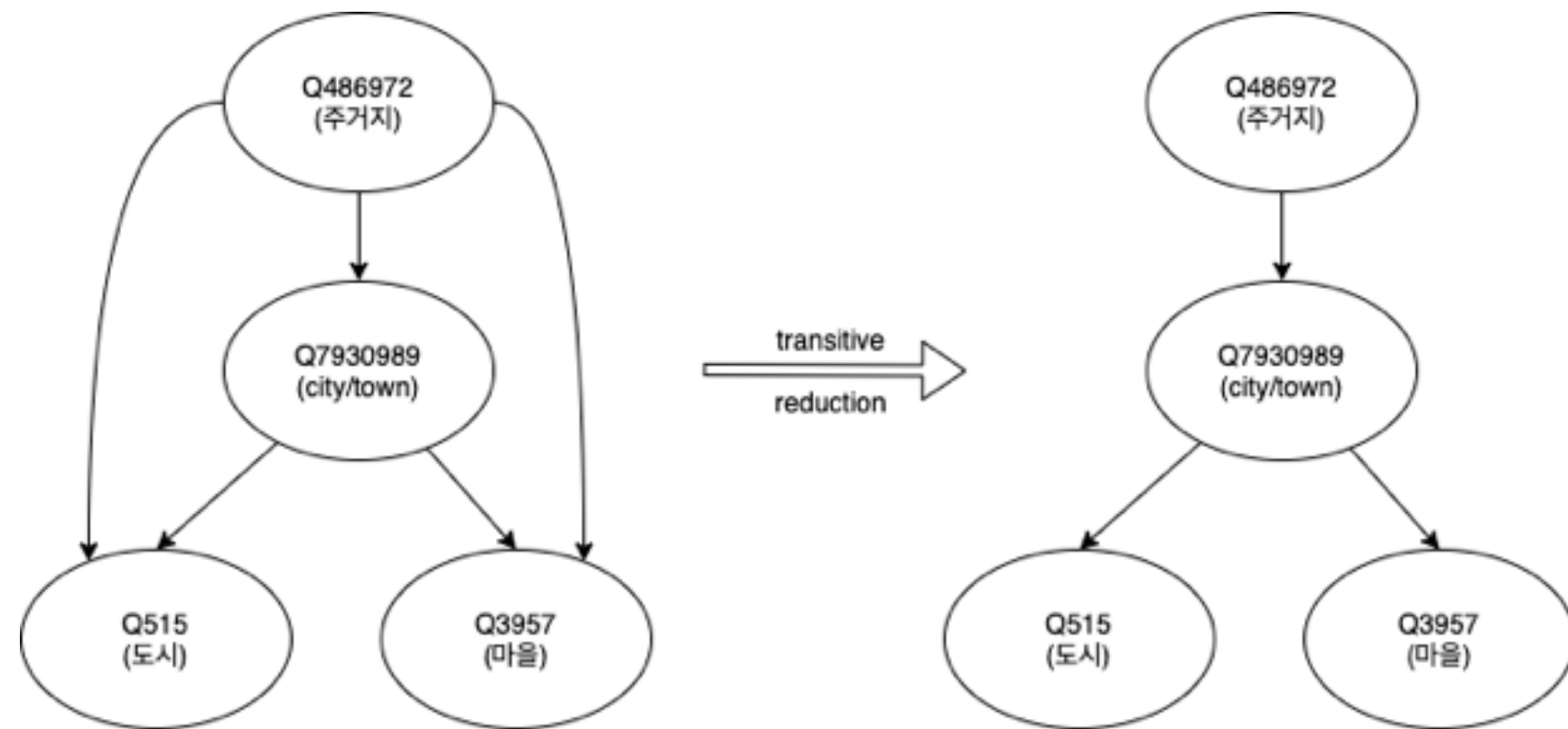


# 3.4 분류체계 생성 - 1) 방향 비순환 그래프

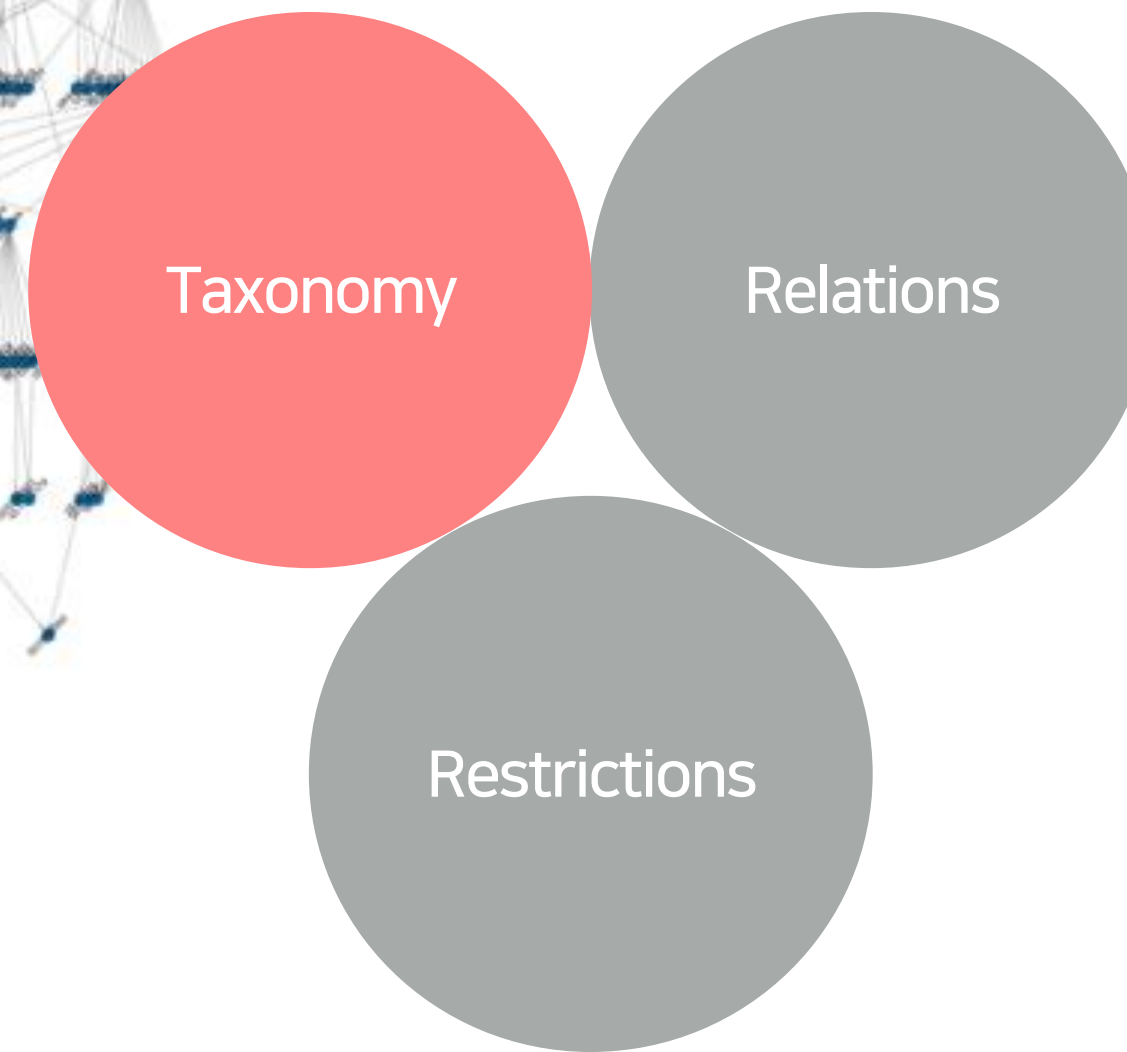
-위키데이터의 "subclass of (P279)" 속성을 통해 분류 후보군 간 상하위 관계 정의

-상위 분류에서 하위 분류로의 방향성을 나타내는  
**방향 비순환 그래프 (Directed Acyclic Graph, DAG) 생성**

-transitive reduction 수행 통해 그래프 정제



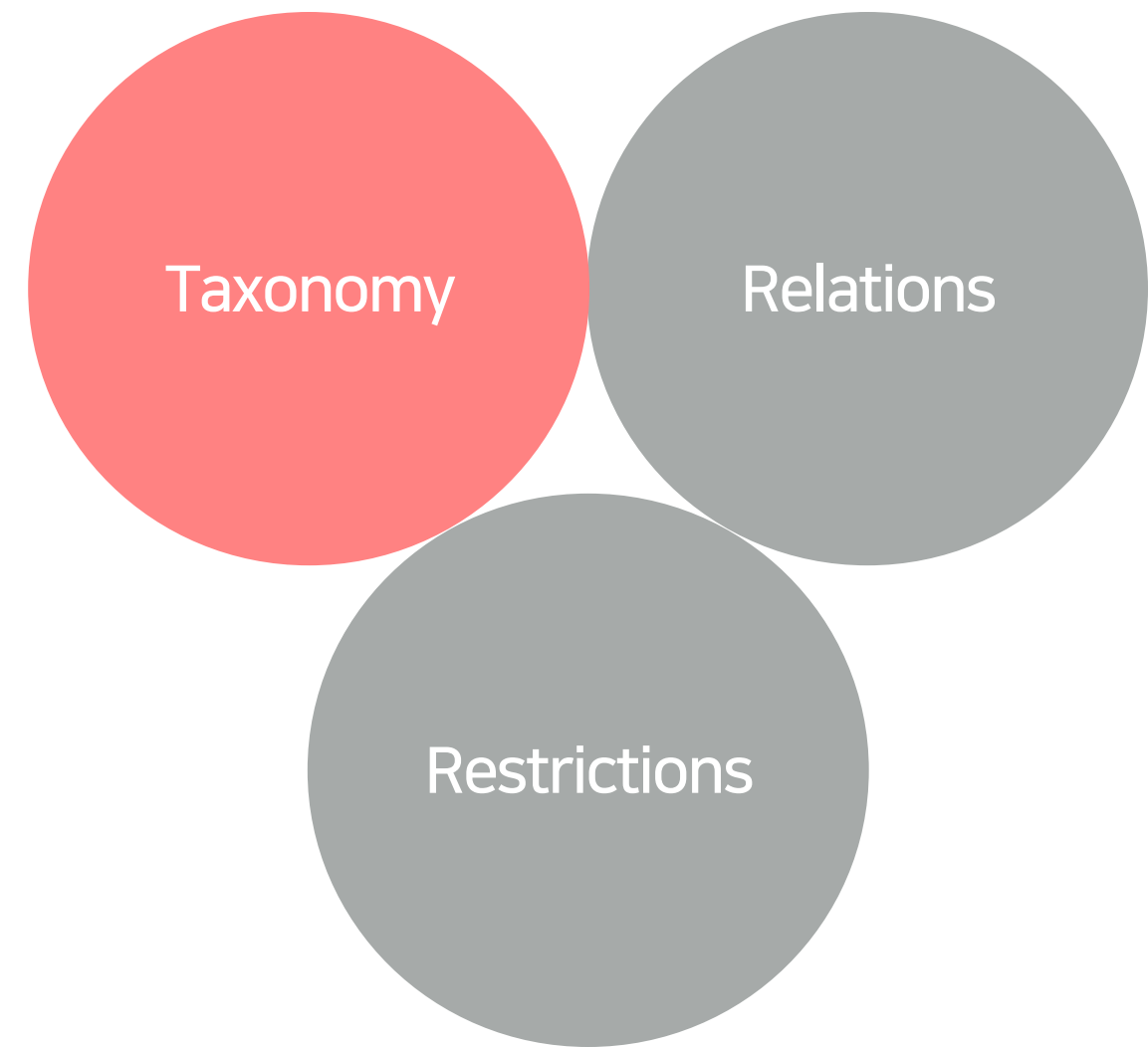
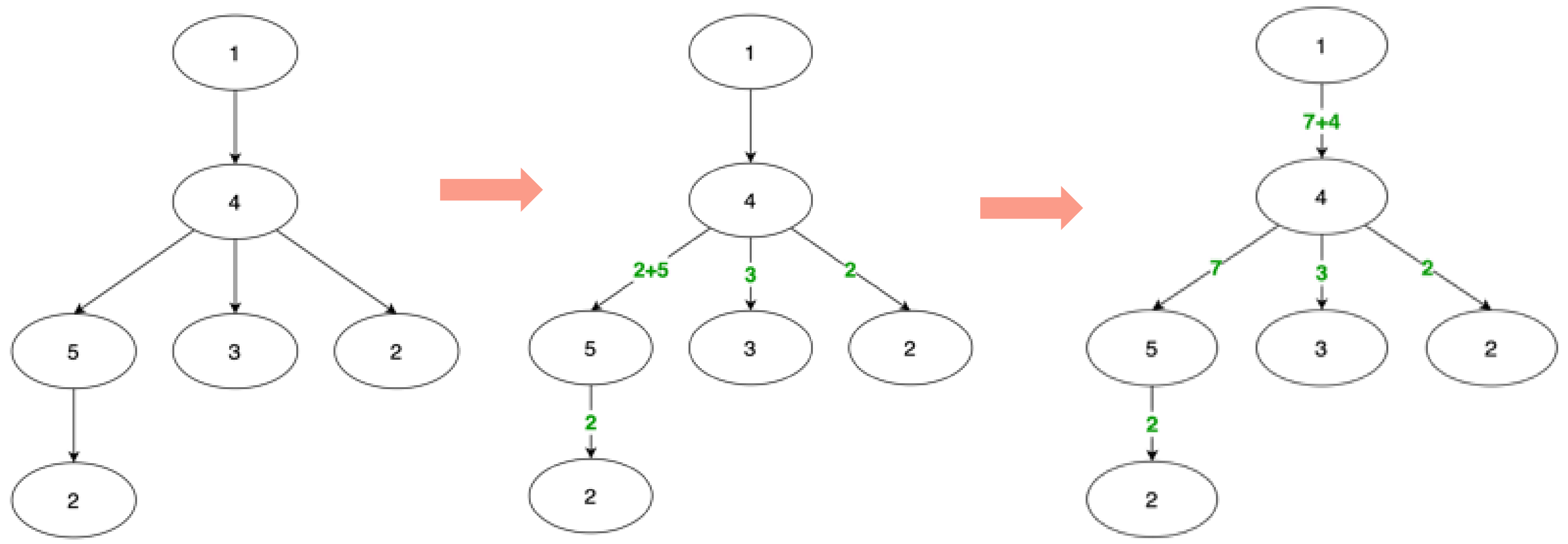
-827개 node(분류) + 1,064개 edge(상하위 관계) 지닌 그래프 생성



# 3.4 분류체계 생성 - 2) 그래프 내 가중치 할당

-node weight : 분류 후보군 추출 시 등장 빈도 수 할당

-edge weight : leaf node 부터 node weight 누적 시키며 부여 (OntoLearn Reloaded\* 논문 방식 변형)



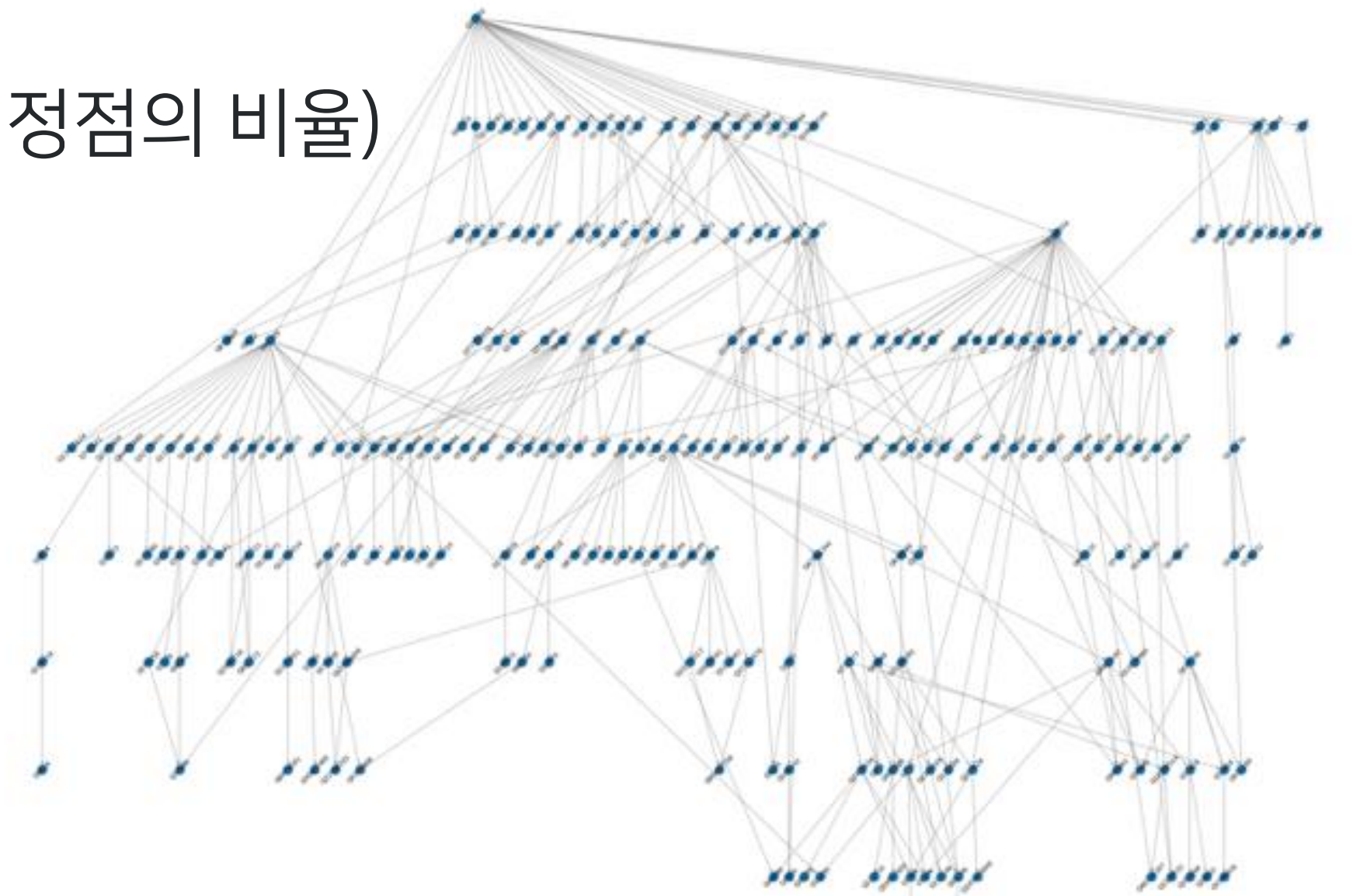
# 3.4 분류체계 생성 - 3) 가지치기(pruning)

-pruning 기준 : local reaching centrality (한 정점에서 도달할 수 있는 다른 정점의 비율)

$$C(i) = \frac{1}{N-1} \sum_{j:0 < d^{out}(i,j) < \infty} \frac{\sum_{k=1}^{d^{out}(i,j)} w_i^k(j)}{d^{out}(i,j)}$$

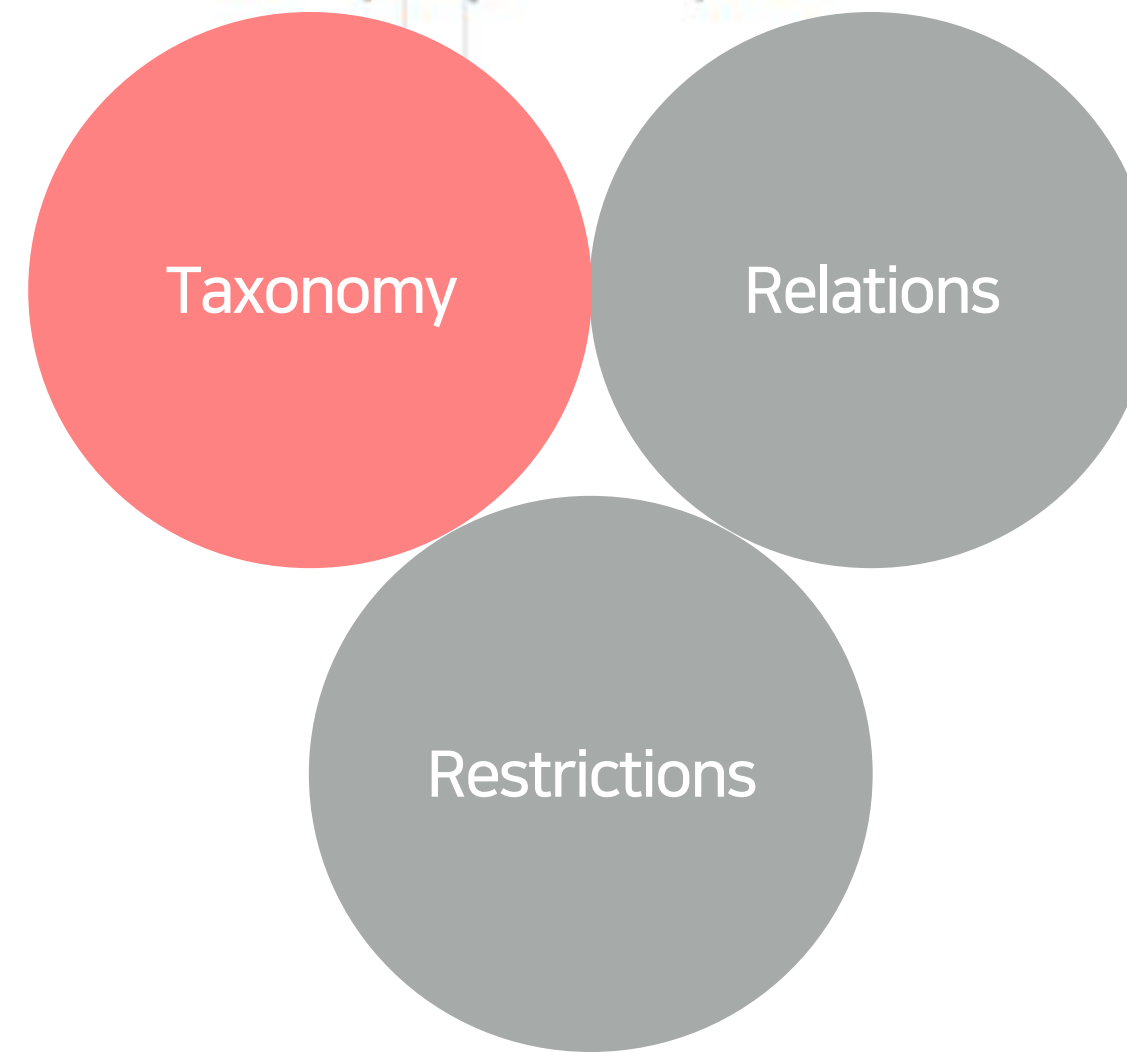
node i에 대한 local reaching centrality

- $N$  : 그래프 크기 (node 개수)
- $d^{out}(i,j)$  : node i에서부터 node j로 향하는 경로 거리
- $w_i^k(j)$  :  $d^{out}(i,j)$ 에 해당하는 경로 중 k번째 edge에 대한 weight



-해당 centrality가 0인 경우 (70.3%)에 제거

-246개 node(분류) + 314개 edge(subclass 관계) 지닌 분류체계 생성





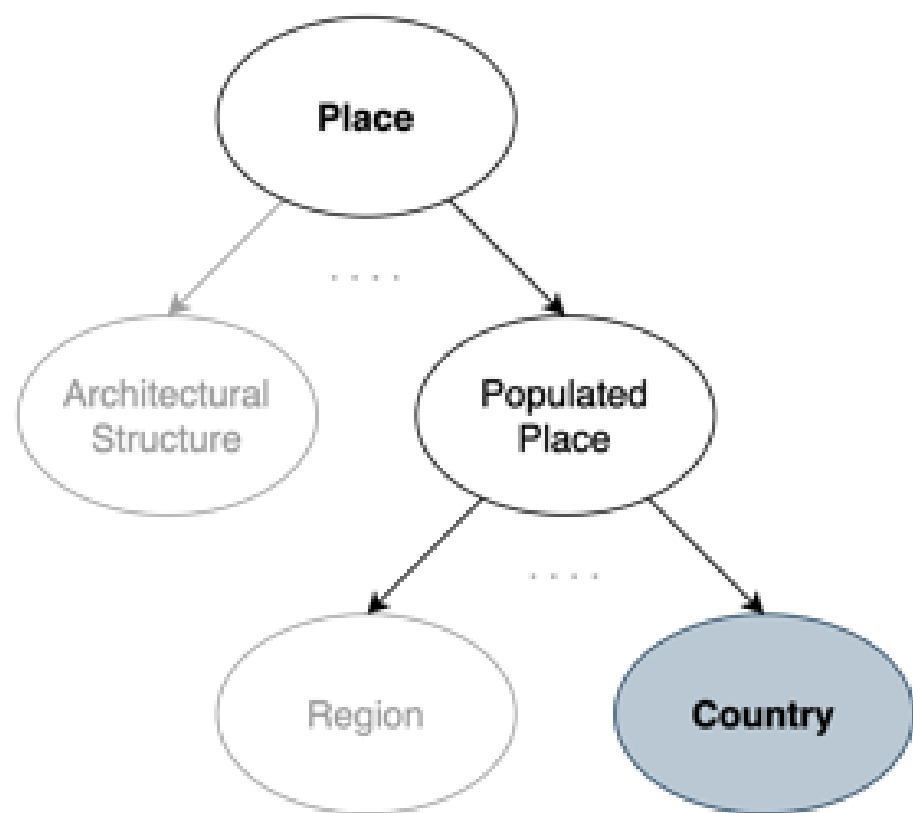
# 3.5 분류체계의 특징 - 1) 다중 상위 분류

-의미론적으로 여러 상위 분류를 지녀야 하는 경우 존재

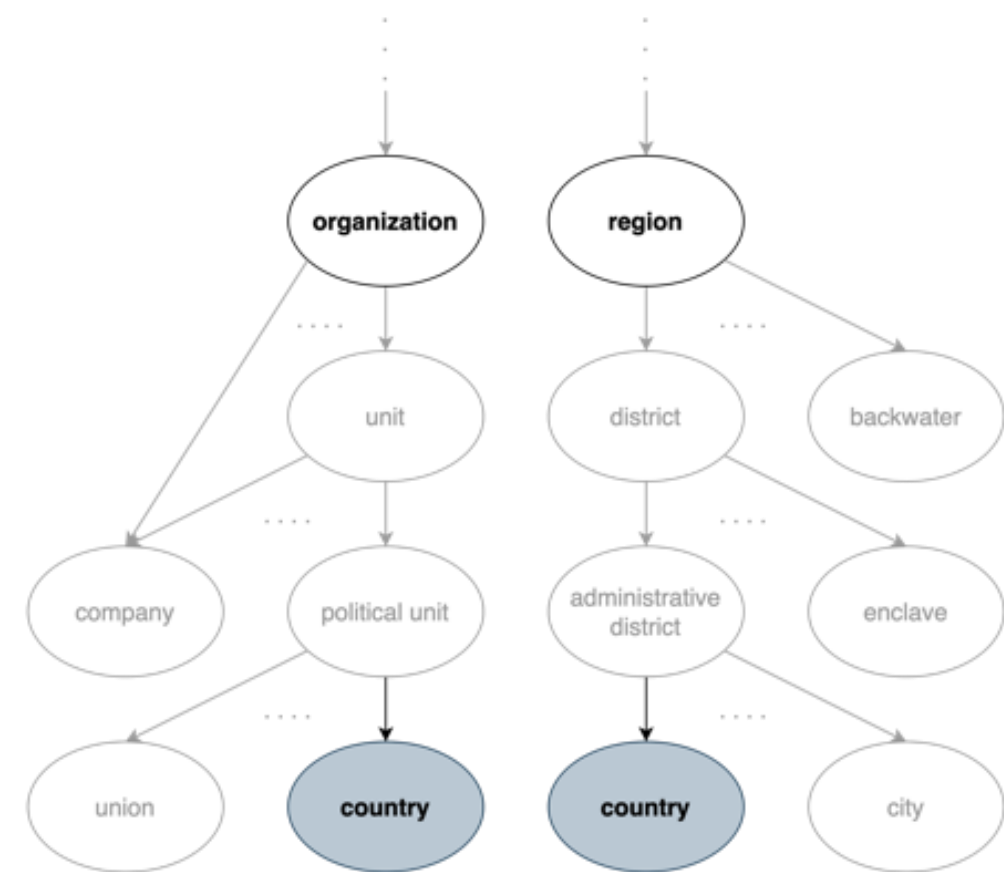
ex) 인공위성(artificial satellite) → 기기(device), 천체(astronomical object)

국가(state) → 행정구역(administrative territorial entity), 단체(organization)

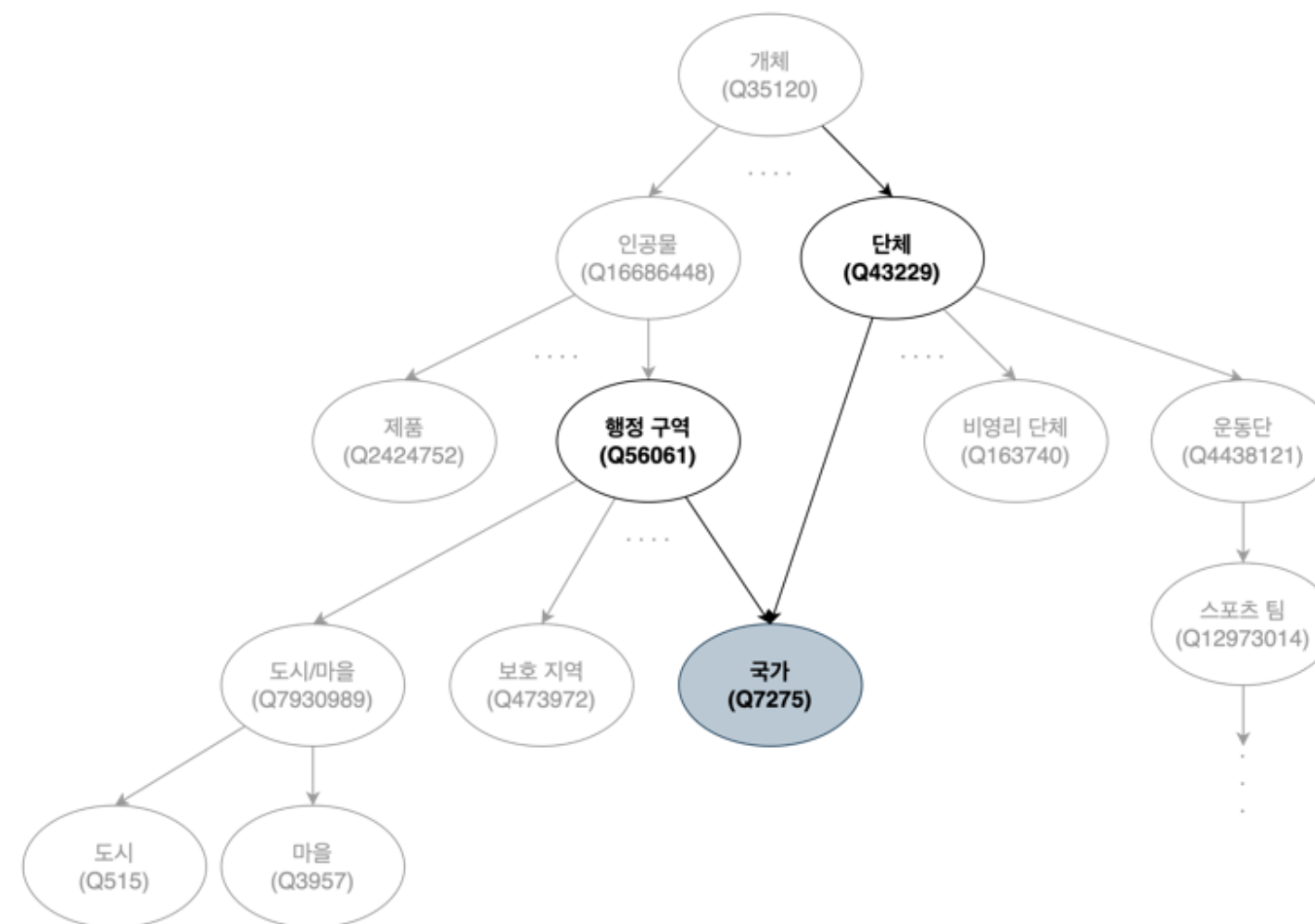
-기존 분류체계(DBpedia, WordNet)에서는 1개의 상위 분류만을 지님



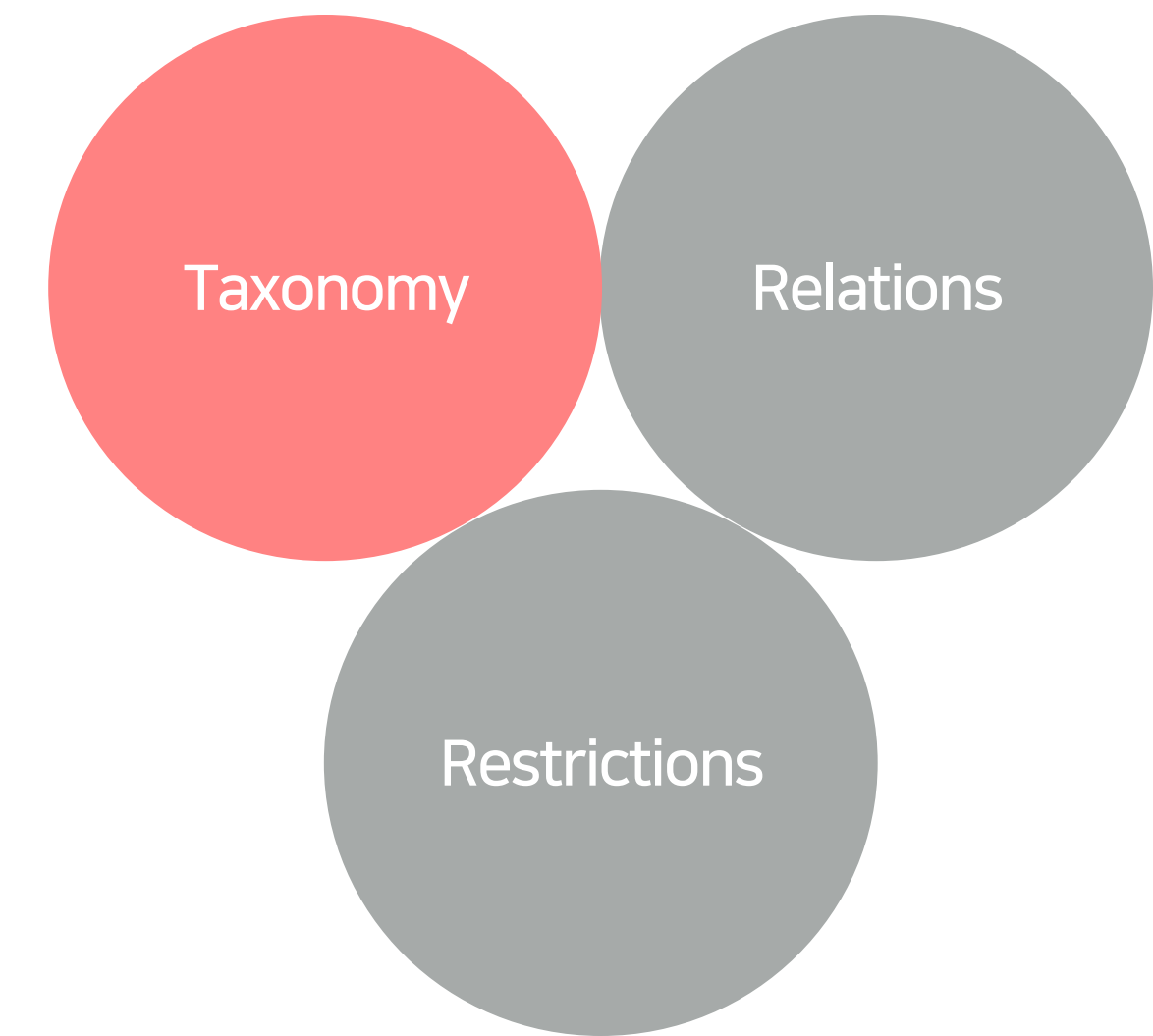
<DBpedia>



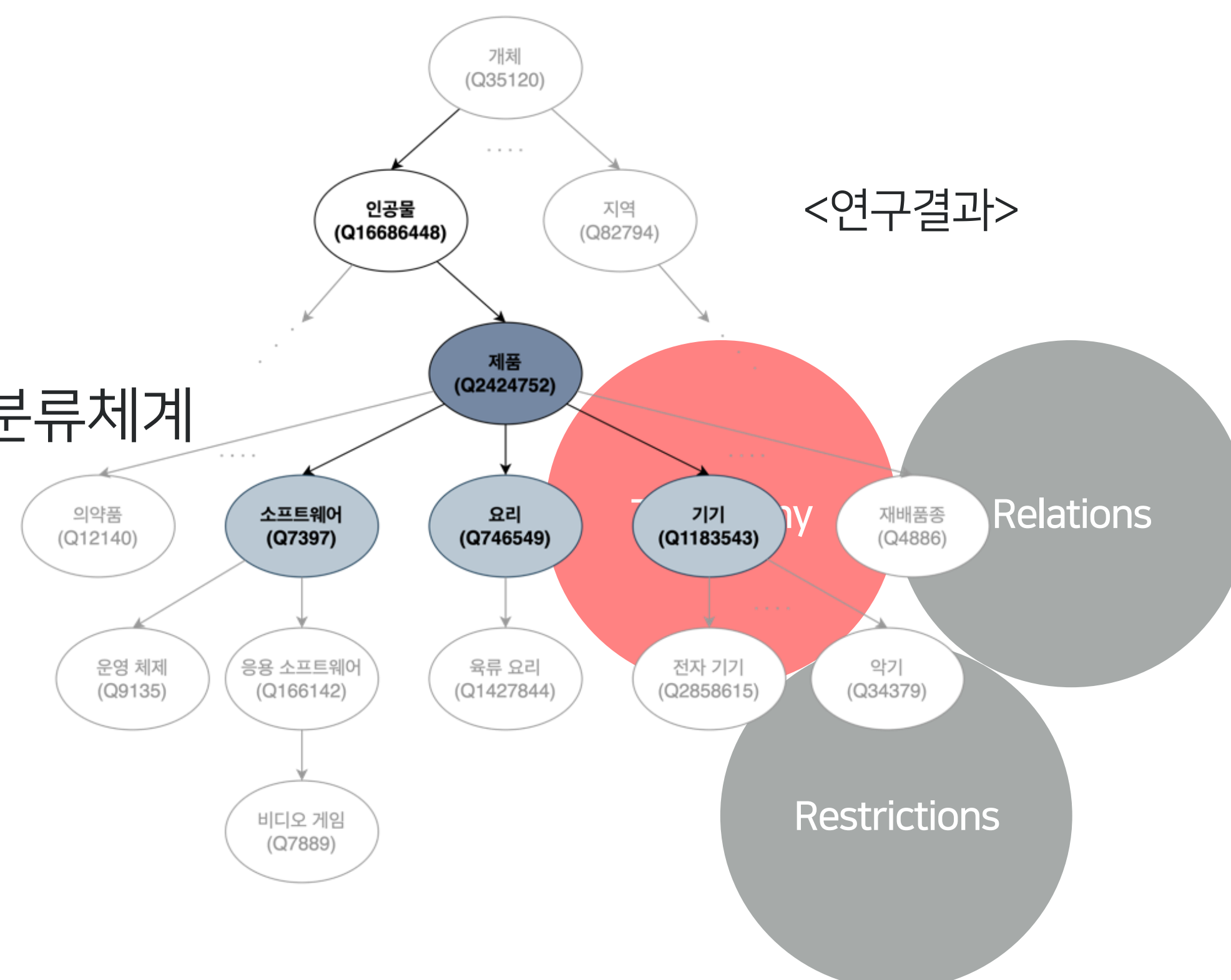
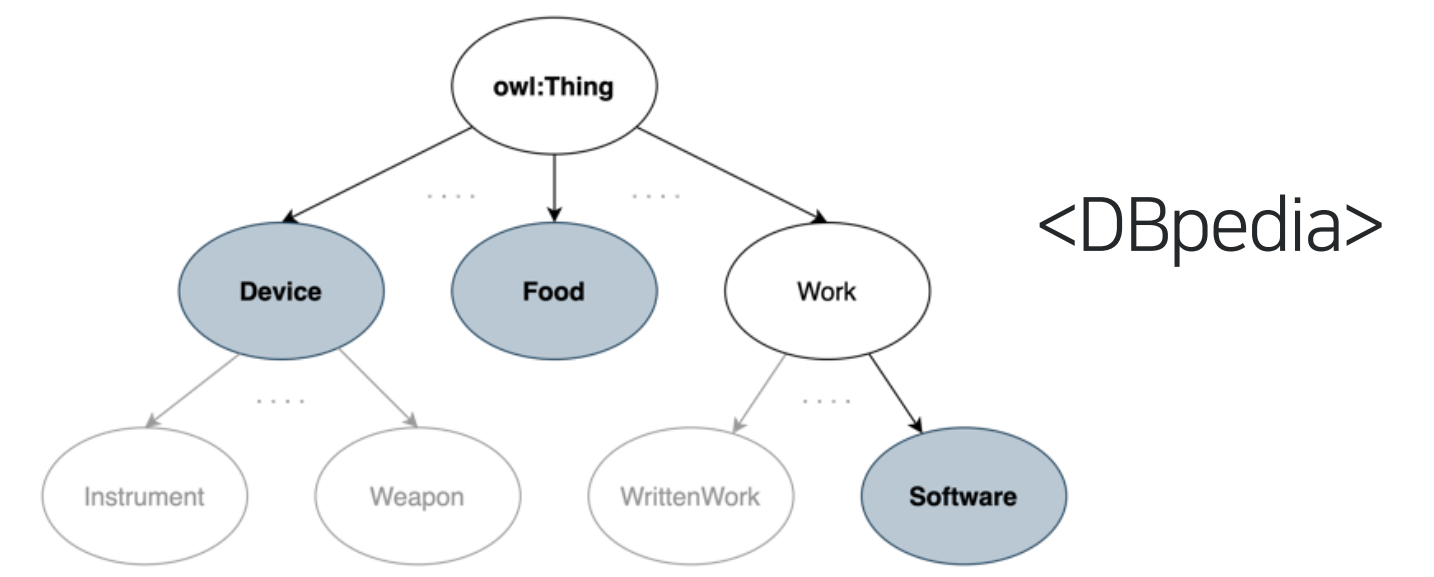
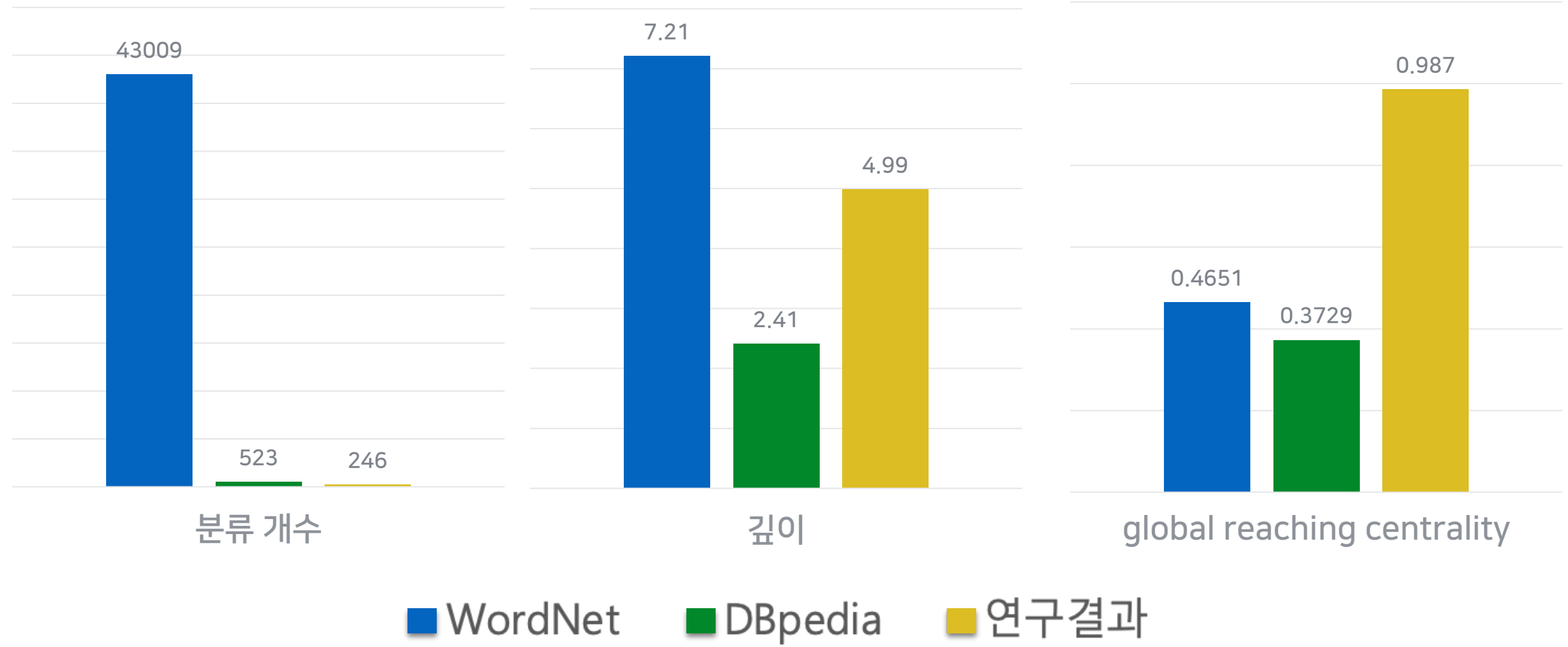
<WordNet>



<연구결과>



# 3.5 분류체계의 특징 - 2) 깊이 있는 계층 구조



- 분류 개수는 가장 적은 core한 분류체계이지만, 더욱 깊으며 중심성이 강한 분류체계
- 깊이 : 각 분류별 최상위 분류로부터의 거리에 대한 평균
- global reaching centrality : 각 node별 local reaching centrality에 대한 평균

# 3.5 분류체계의 특징 - 3) 분류 자동 할당

-위키데이터의 "instance of (P31)" 및 "subclass of (P279)" 속성 기반으로 위키데이터 각 item마다 자동으로 분류 할당 가능

-정답 데이터 : DBpedia 데이터 중, 위키데이터 식별자와의 연결 정보가 있는 entity 3,365,936개

-분류 할당 결과

- 커버리지 : 99.83% (3,360,335 / 3,365,936)

- 정확도 : 99.81% (3,342,695 / 3,365,936)

- DBpedia 보다 세부적인 분류로 할당

ex) 겨울왕국 (Q246283)

→ DBpedia 분류 : "영화 (Q11424)"

→ 연구결과 : "애니메이션 영화 (Q202866)" 로 분류

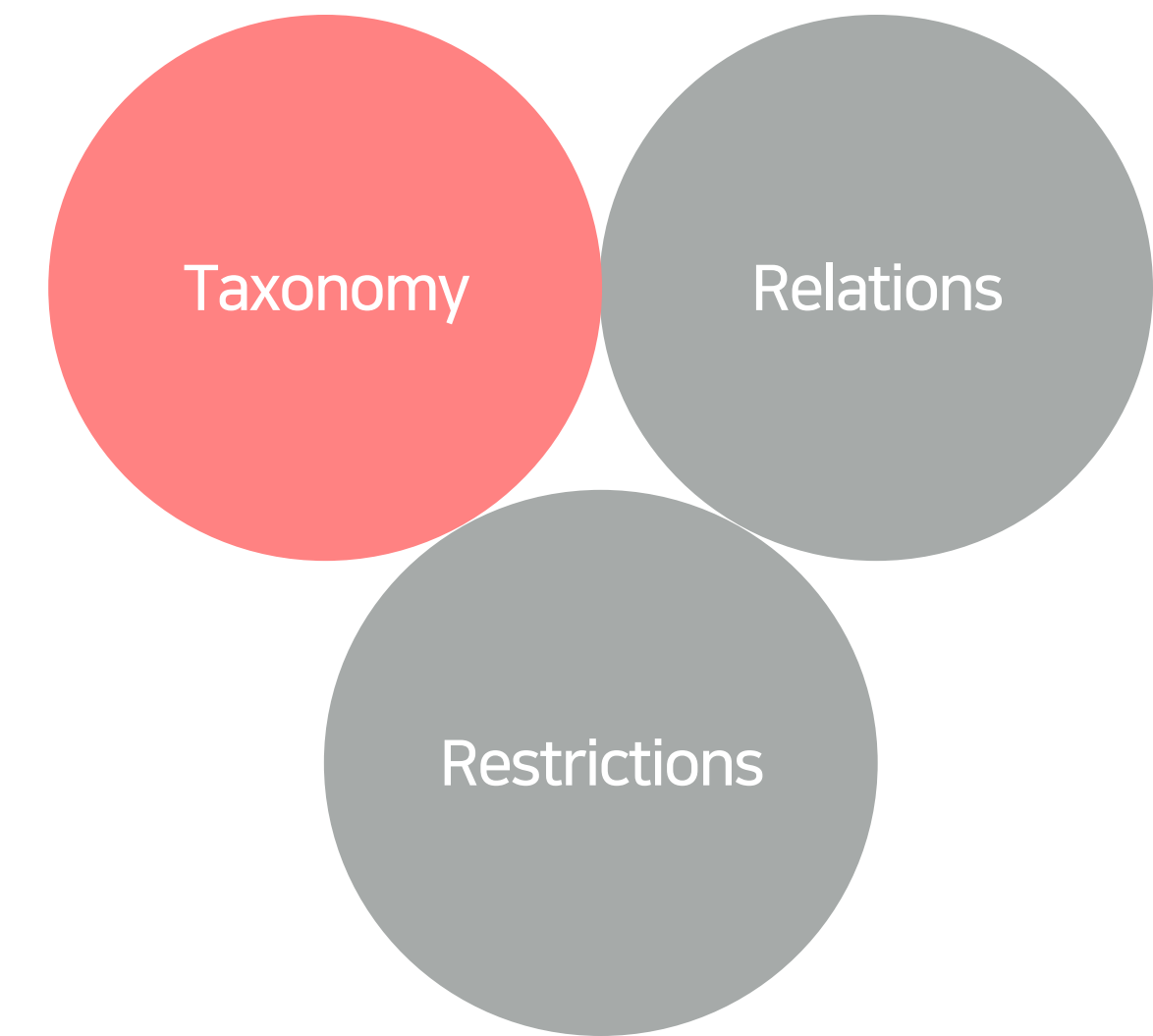
- 위키데이터의 속성이 잘못 할당되어 있을 경우 오류 발생

ex) 디지털 마케팅 (Q1323528)

→ DBpedia 분류 : "학문 (Q11862829)"

→ 연구결과 : "직업 (Q192581)" 로 잘못 분류

(instance of 속성값 "프리랜서 (Q215279)" 에 의함)

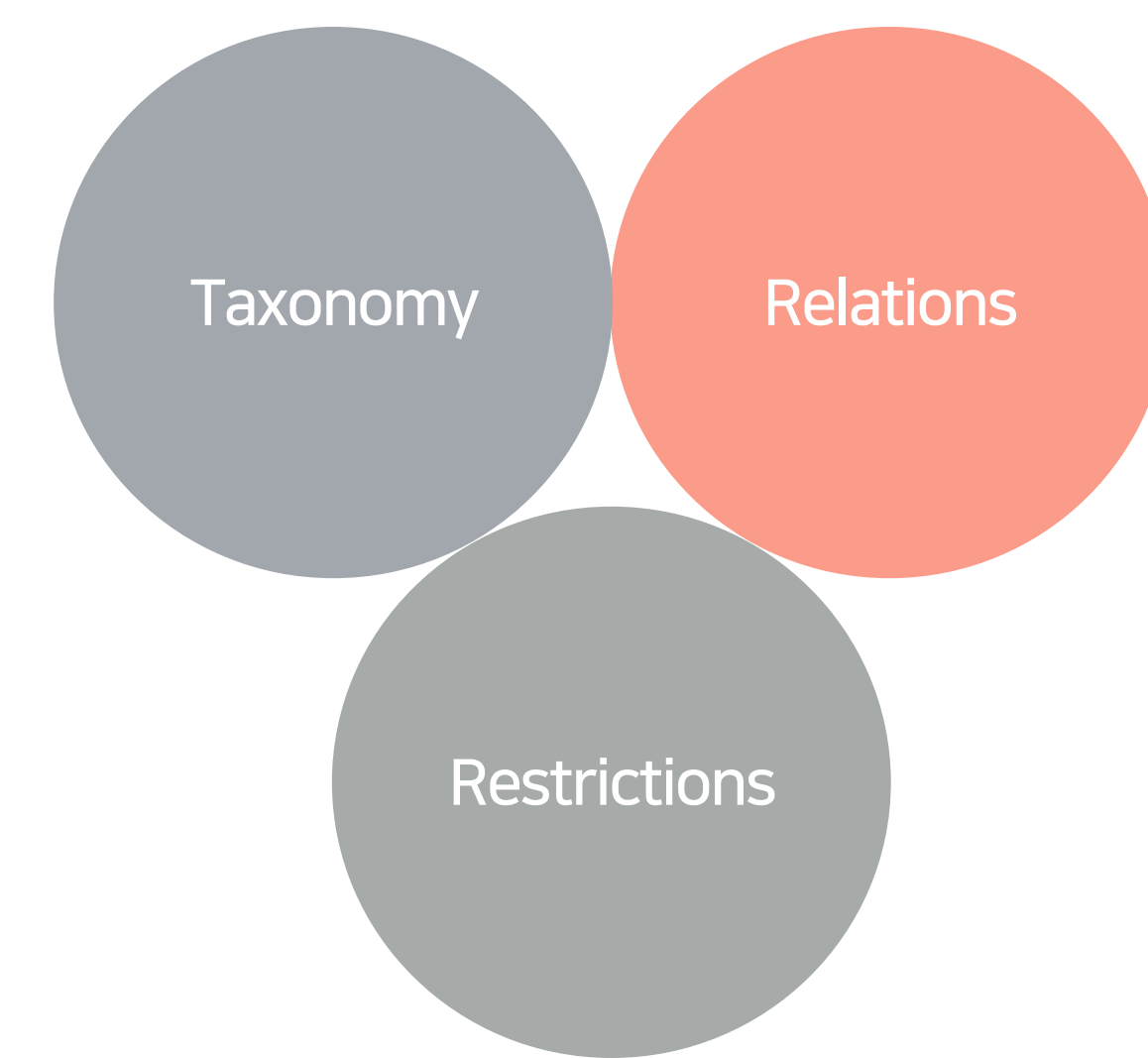


# 3.6 Relations

properties for this type	
sex or gender	▼ 0 references
date of birth	▼ 0 references
date of death	▼ 0 references
given name	▼ 0 references
family name	▼ 0 references
country of citizenship	▼ 0 references
occupation	▼ 0 references

## human (Q5)

common name of Homo sapiens, unique extant species of the genus Homo  
human being | humankind | people | homosapiens | person | mankind | peo



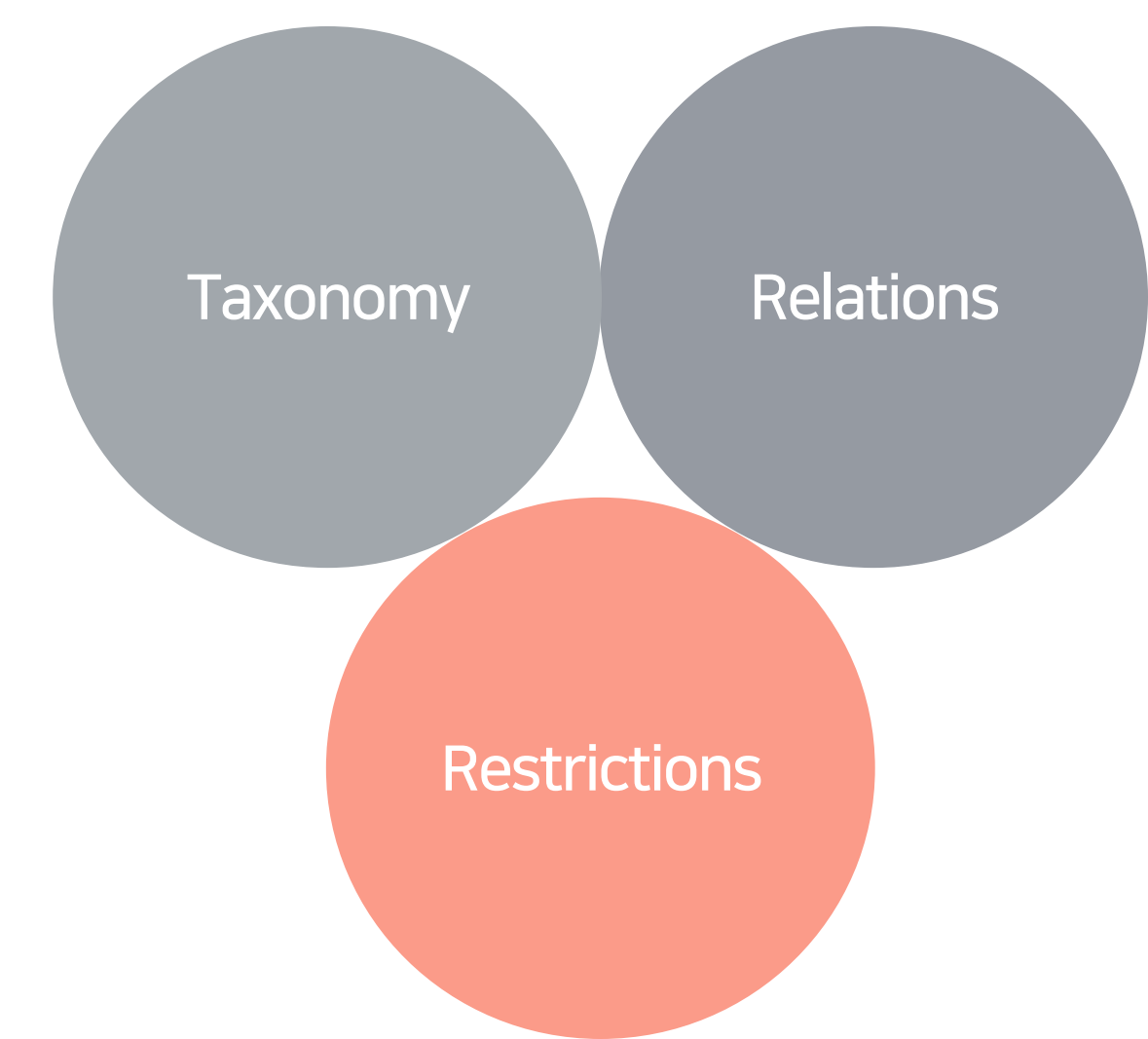
# 3.7 Restrictions

## Constraints

property constraint	value-type constraint	
	class	<ul style="list-style-type: none"> <li>geographic location</li> <li>fictional location</li> <li>geographic region</li> <li>mythical place</li> <li>building</li> <li>architectural ensemble</li> <li>residence</li> <li>geographic entity</li> <li>house</li> <li>nursing home</li> </ul>
	relation	instance of
	▼ 0 references	
	type constraint	
	class	<ul style="list-style-type: none"> <li>human</li> <li>fictional character</li> <li>domesticated animal</li> <li>human biblical figure</li> <li>animal living in captivity</li> </ul>

## residence (P551)

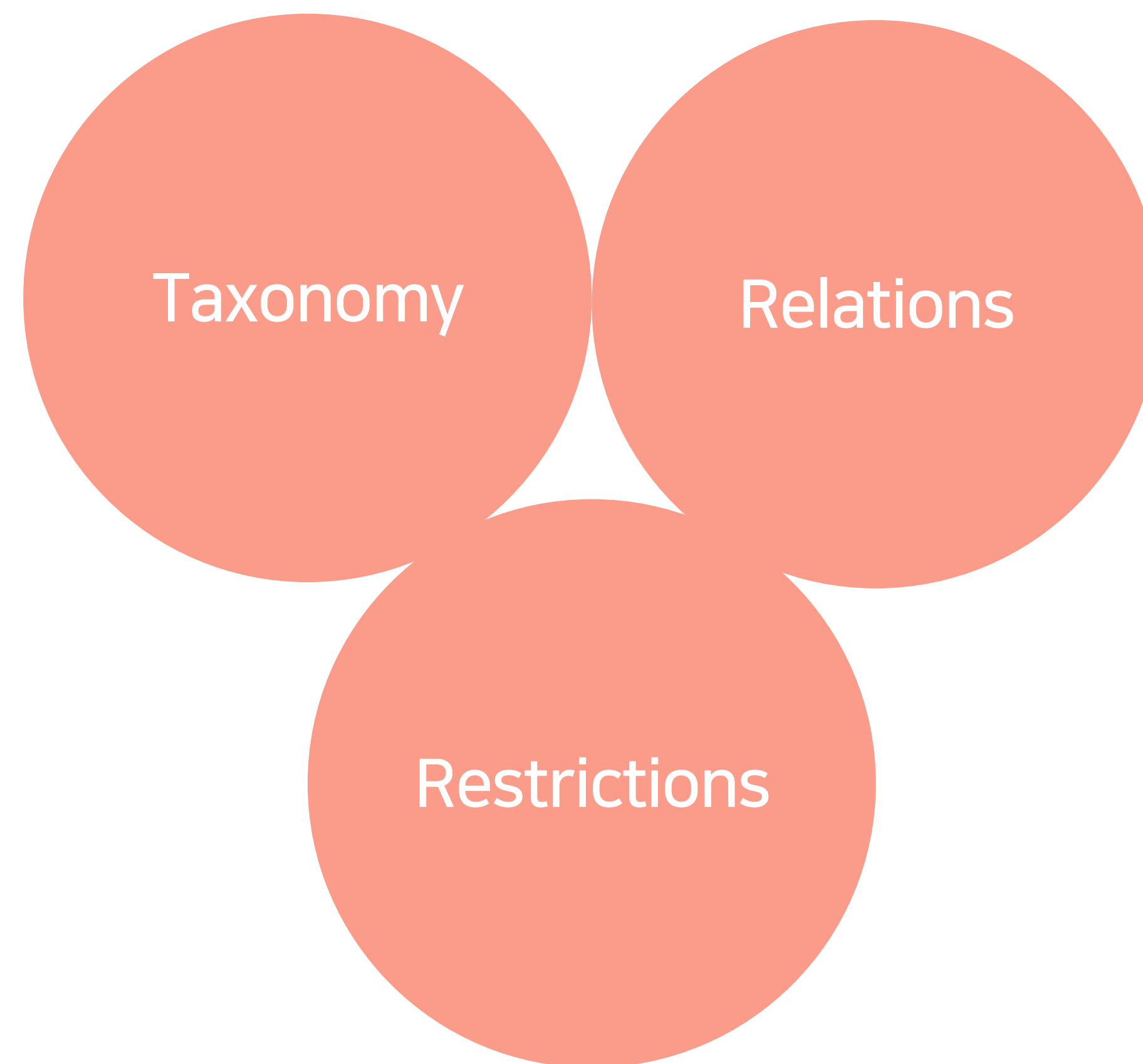
the place where the person is or has been, resident lives in | hometown | home town | place of residence |



# 3.1 Ontology

## Ontology

- Taxonomy
- Relations
- Restrictions

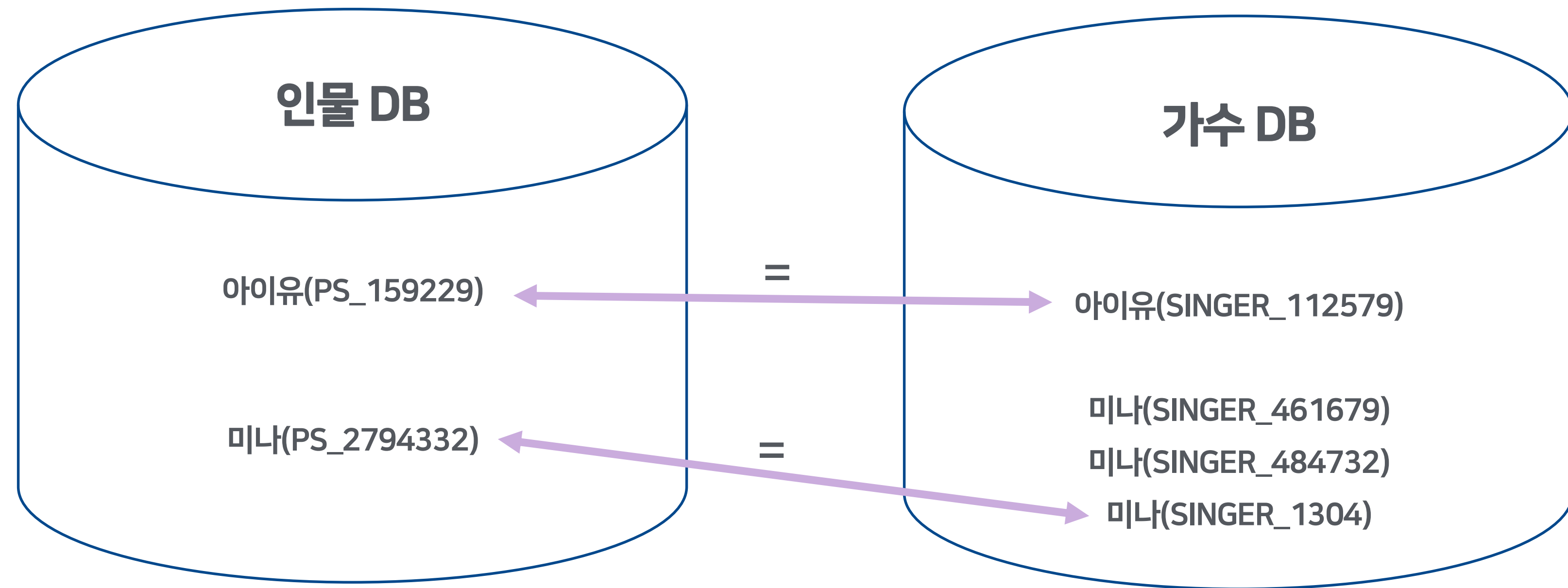


# 4. Record Linkage Entity Alignment

# 4.1 소개

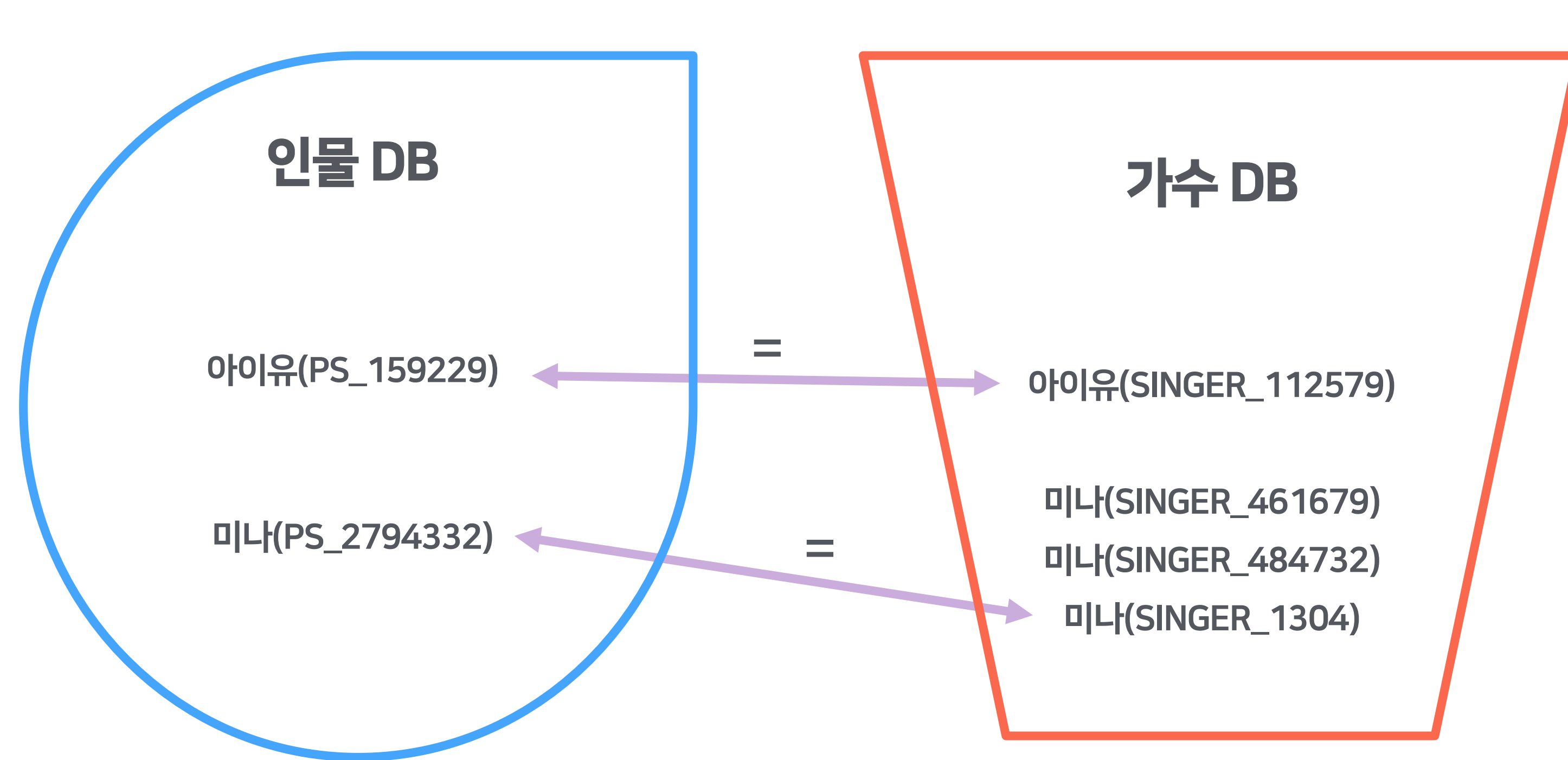
## 2개 이상의 DB에서 동일한 record(entity)를 찾아내는 task

- Record Linkage
- DB Linkage
- Data Matching
- Entity Resolution
- Entity Alignment

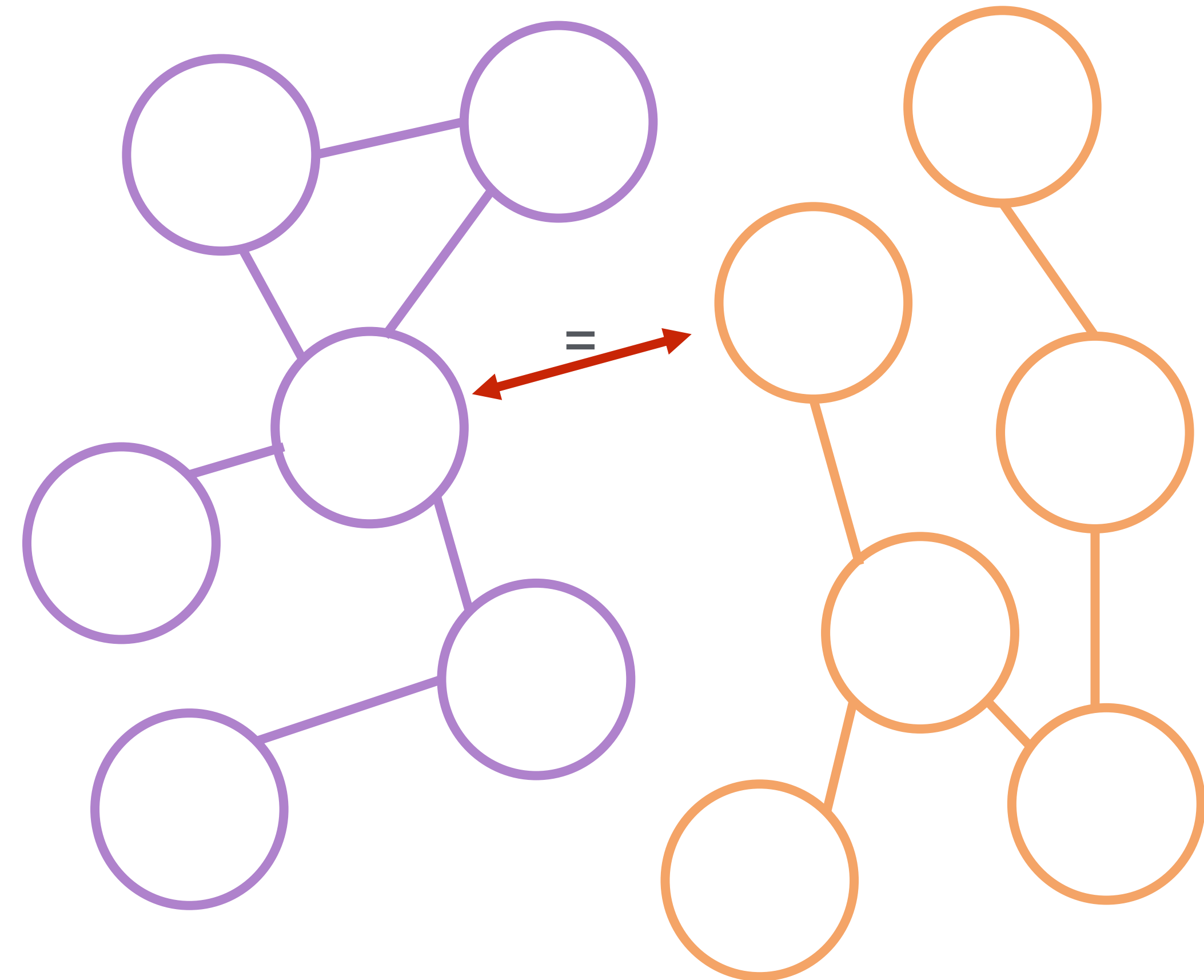




# 4.1 소개

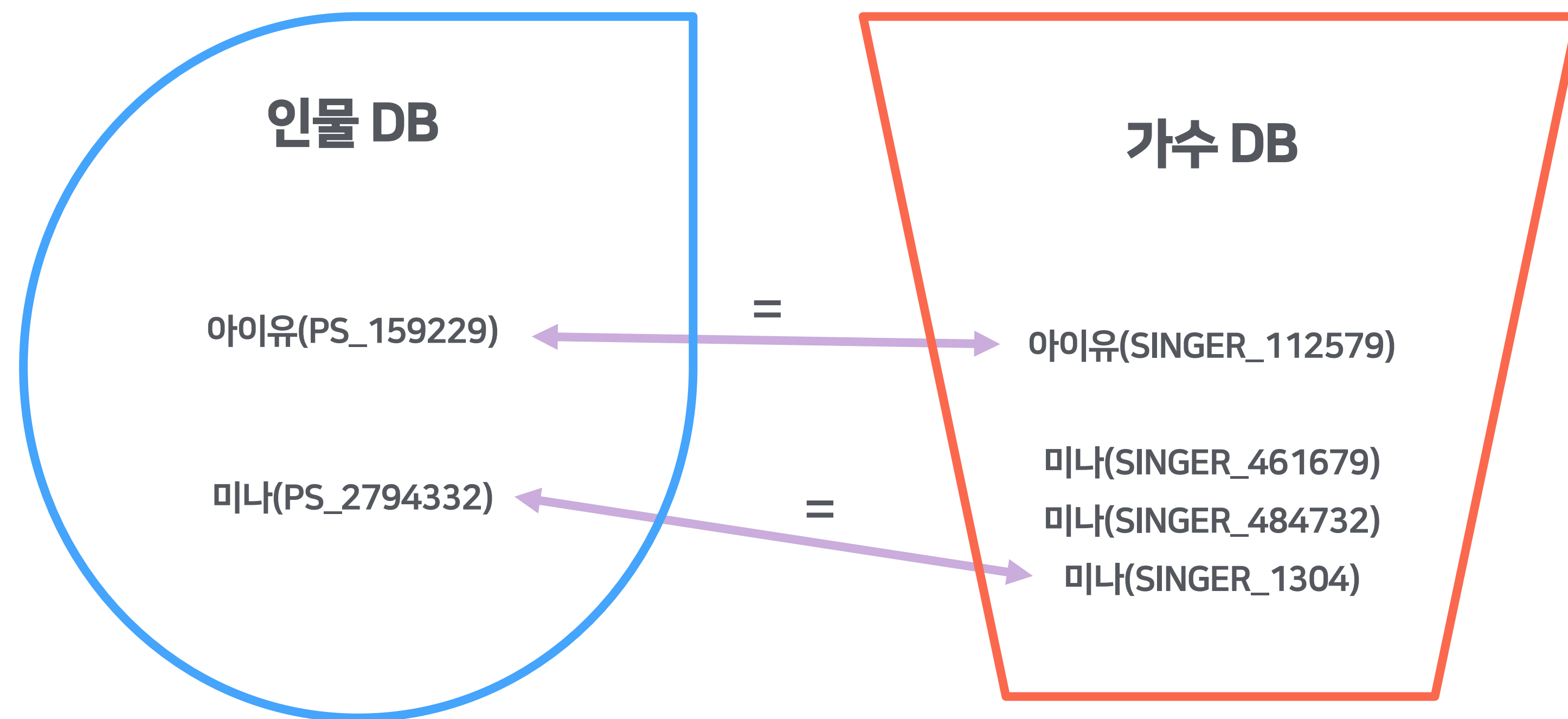


- Record Linkage
- DB Linkage
- Data Matching

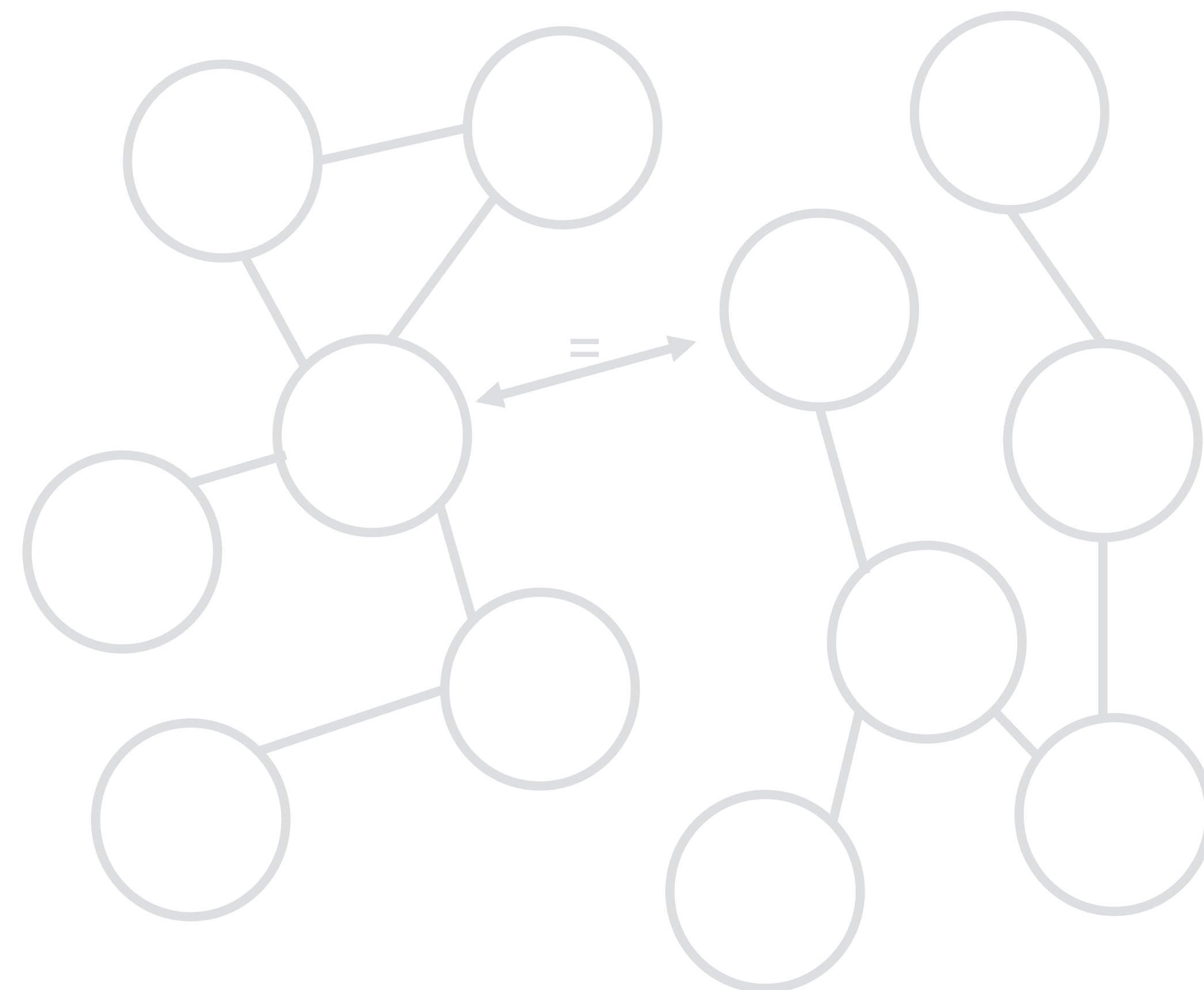


- Entity Resolution
- Entity Alignment

# 4.2 Record Linkage



- Record Linkage
- DB Linkage
- Data Matching



- Entity Resolution
- Entity Alignment

# 4.2 Record Linkage

제33회 한글 및 한국어 정보처리 학술대회 논문집 (2021년)

## 레코드 연결을 위한 속성인지 메타블로킹

이주현<sup>o</sup>, 김현호, 강인호

네이버

dev\_joohyun.lee@navercorp.com, kim.hh@navercorp.com, once.ihkang@navercorp.com

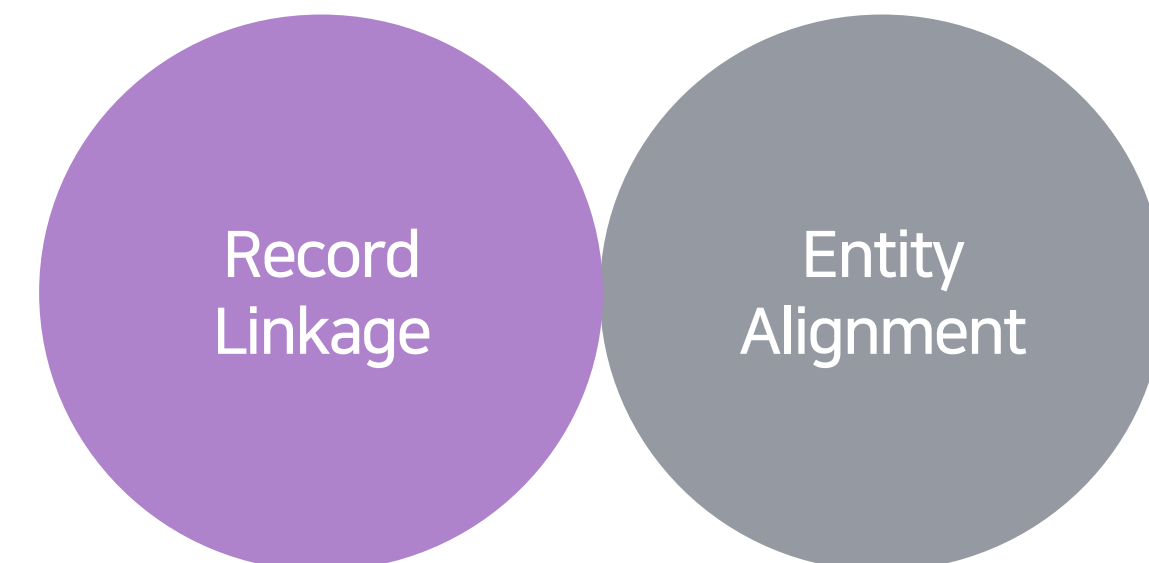
## Property-aware Meta Blocking for Record Linkage

Joo-Hyun Lee<sup>o</sup>, Hyun-Ho Kim, In Ho Kang

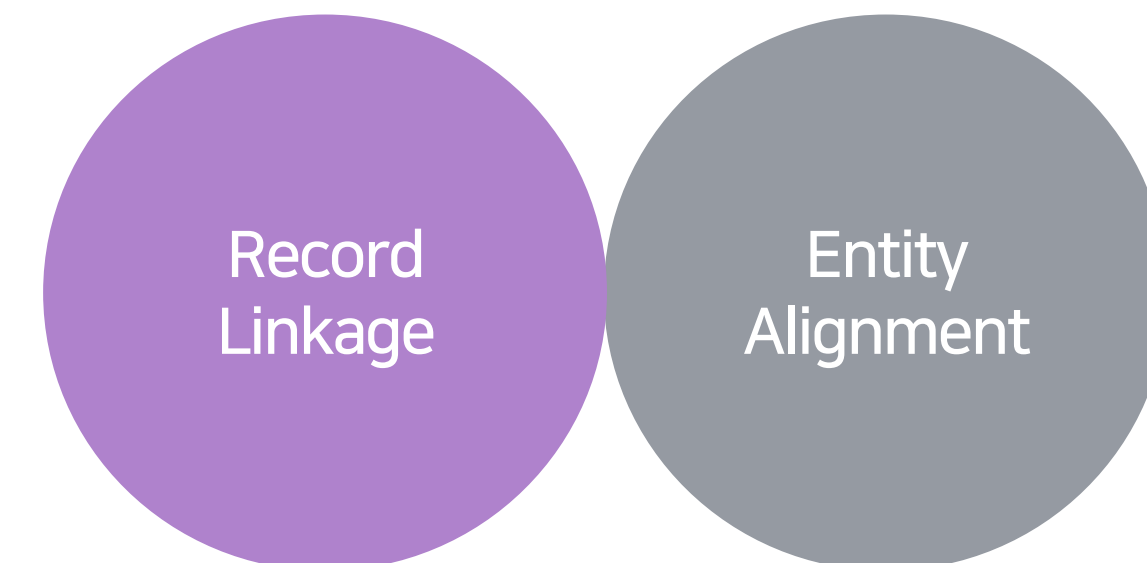
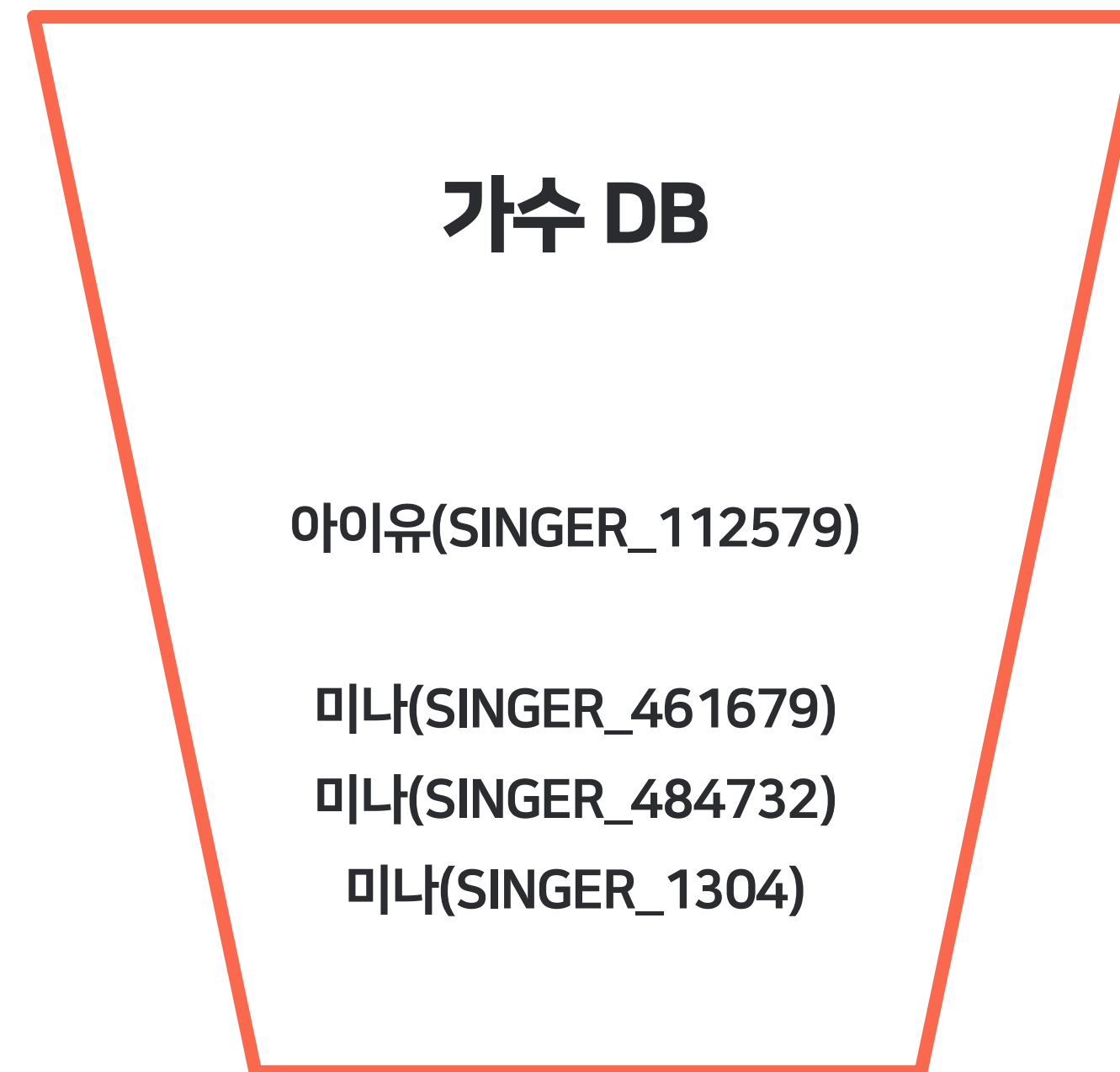
Naver Corp.

### 요약

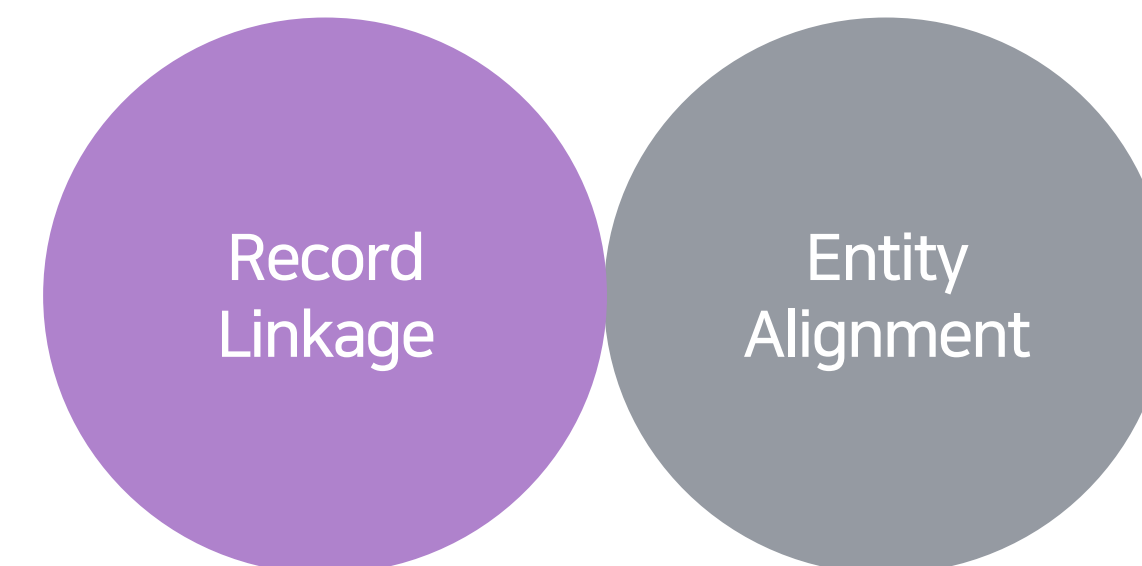
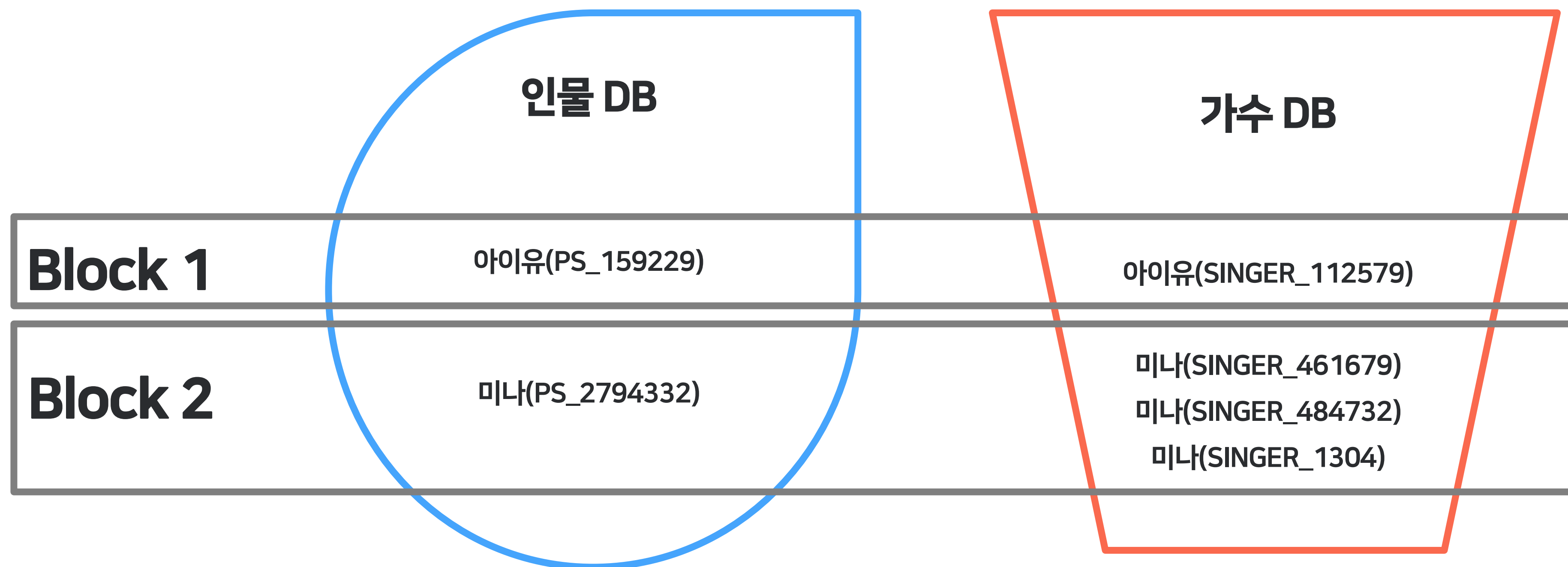
레코드 연결의 대표적인 문제 중 하나는 레코드 간 비교 비용이 크다는 것이다. 이러한 문제를 해결하기 위해서는 레코드 연결에 필수적으로 블로킹 단계가 포함되어야 한다. 블로킹이란 같은 레코드일 가능성이 높은 대상들을 그룹화하여 비교연산을 수행할 대상을 선정하는 단계를 말한다. 블로킹의 목적은 최대한 결과의 recall을 희생시키지 않으면서 비교 연산 횟수 최소화하는 것이다. 메타 블로킹은 가중치 그래프를 블로킹에 적용함으로써 전통적인 블로킹 방식의 한계를 극복하고 더 좋은 성능을 나타내는 모델이다. 본 논문에서는 메타블로킹에서 주목하지 않았던 블록 생성방식을 데이터베이스 속성에 따라 블록을 생성하는 방식으로 개선하고 그에 맞는 가중치 계산식을 제안하였다. 또한 키 기반 블로킹, 메타블로킹, 속성인지 메타블로킹으로 생성된 블로킹 결과에 대한 성능을 측정 및 비교하였다.



# 4.2 Record Linkage



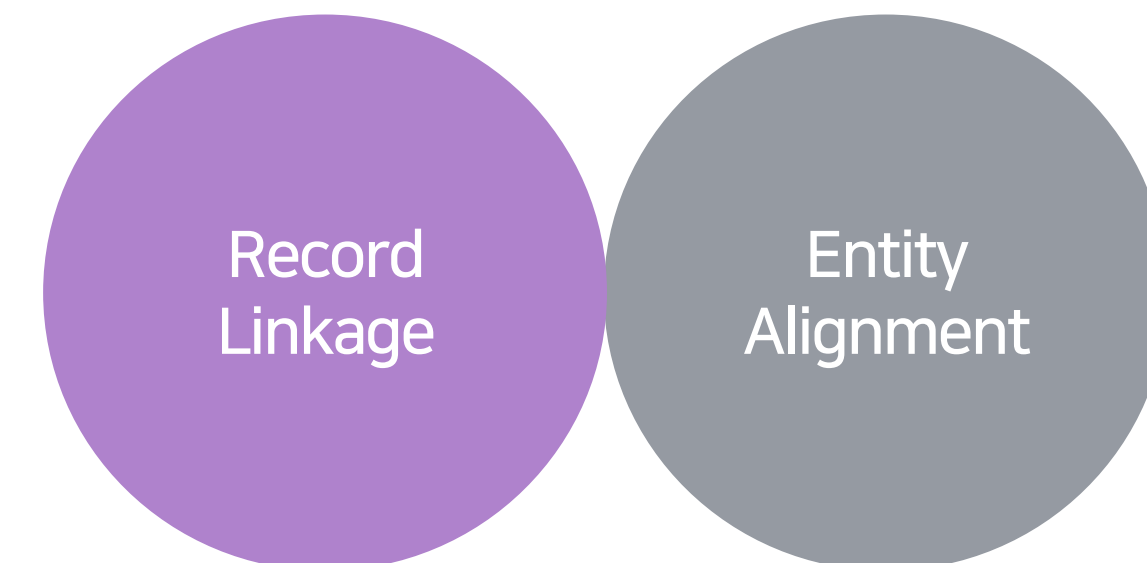
# 4.2 Record Linkage



# 4.2 Record Linkage

## Blocking

- **지수적으로 증가(quadratic increase)하는 complexity 문제 극복을 위한 단계**  
→ DB1: N records , DB2 : M records →  $N * M$ 의 비교 필요
- **같은 레코드일 가능성이 높은 대상들을 그룹화하여 비교를 수행할 대상을 선정하는 단계**
- **동일한 블록에 속한 record들만 비교 수행**  
→ recall을 희생하지 않으면서 비교 횟수를 줄이는 것이 핵심
- **두 가지 블로킹 모델 제공**
  1. Key-base blocking  
: record를 대표하는 컬럼의 값이 일치하는 대상 블록 생성
  2. Meta-Blocking  
: record의 다수 컬럼을 활용하여 블록 생성 및 가중치 그래프 변환 후 최종 블록 정보 생성



# 4.3 Meta blocking

Blocking key가 다른 경우



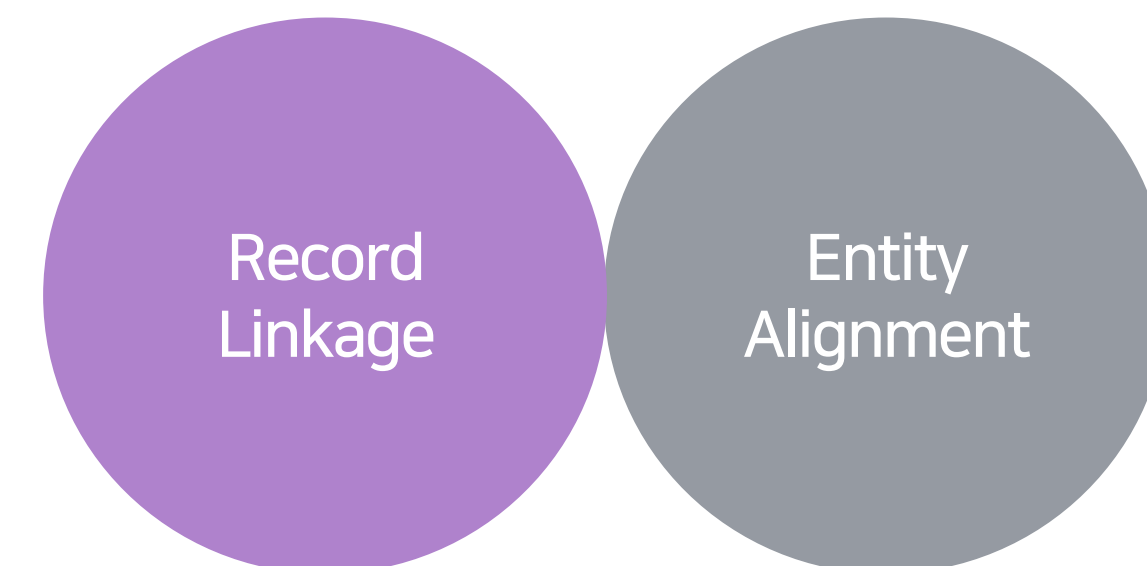
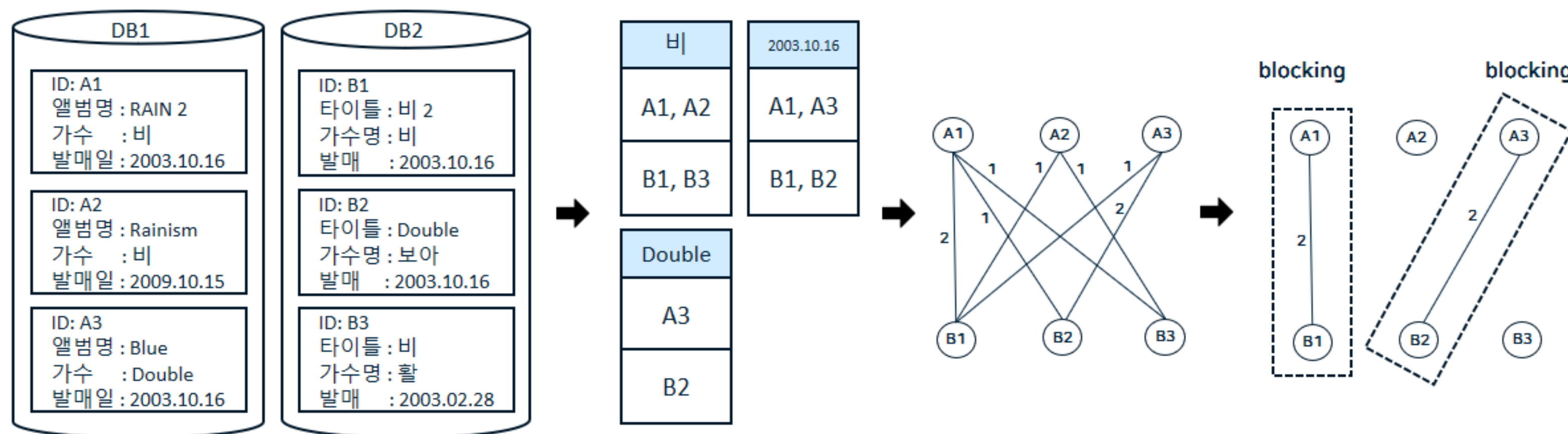
비2



아티스트	비
앨범 종류	정규
발매일	2003.10.16
장르	댄스/팝
스타일	댄스 팝, 팝
기획사	JYP 엔터테인먼트
유통사	Dreamus

## Meta Blocking 수행

block 정보를 가중치 그래프로 표현 후 edge pruning 수행



# 4.4 Property-aware Meta Blocking

## 블록 생성 방식과 가중치 계산식을 개선한 메타 블로킹

### • 블록 생성 방식 개선

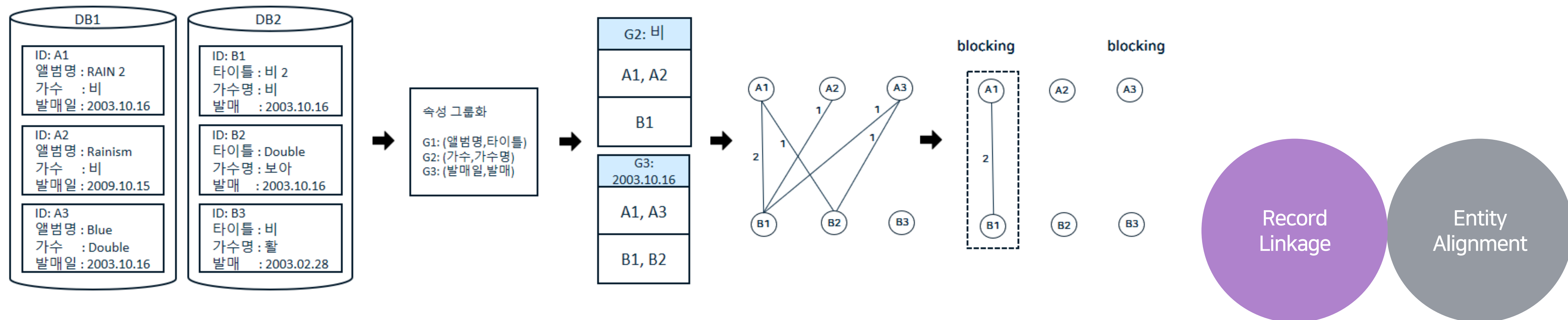
동일 컬럼 pair를 선정하여 블록 생성

### • 가중치 계산식 개선

생성되는 블록의 수 감소를 반영한 새로운 가중치 식 사용

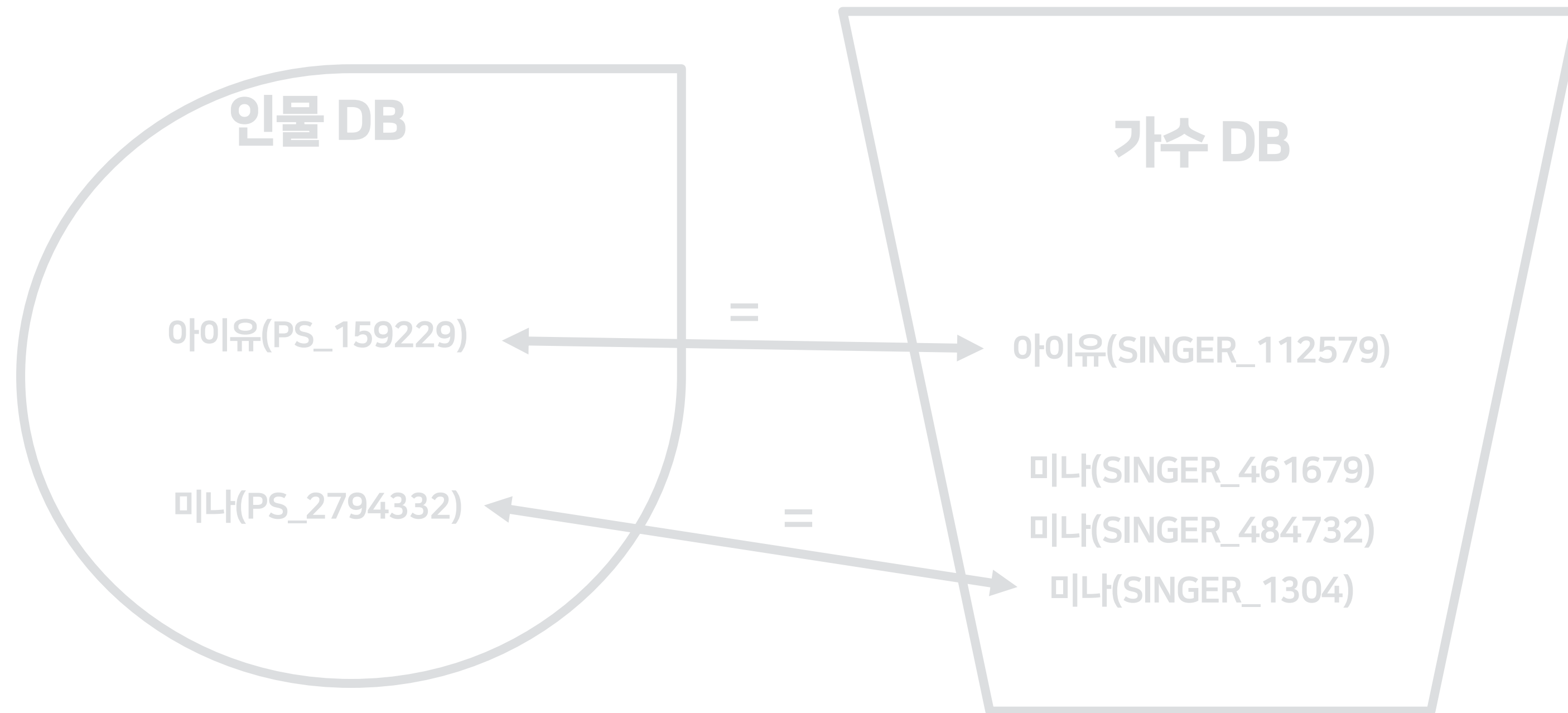
→ 공통블록의 수에 공통블록의 중요도 수치 반영

$$e_{i,j} \cdot weight = |B_{i,j}| \cdot \log \left( \frac{\sum_{b_k \in B_{i,j}} \frac{|N|}{\|b_k\|}}{|B_{i,j}|} \right)$$

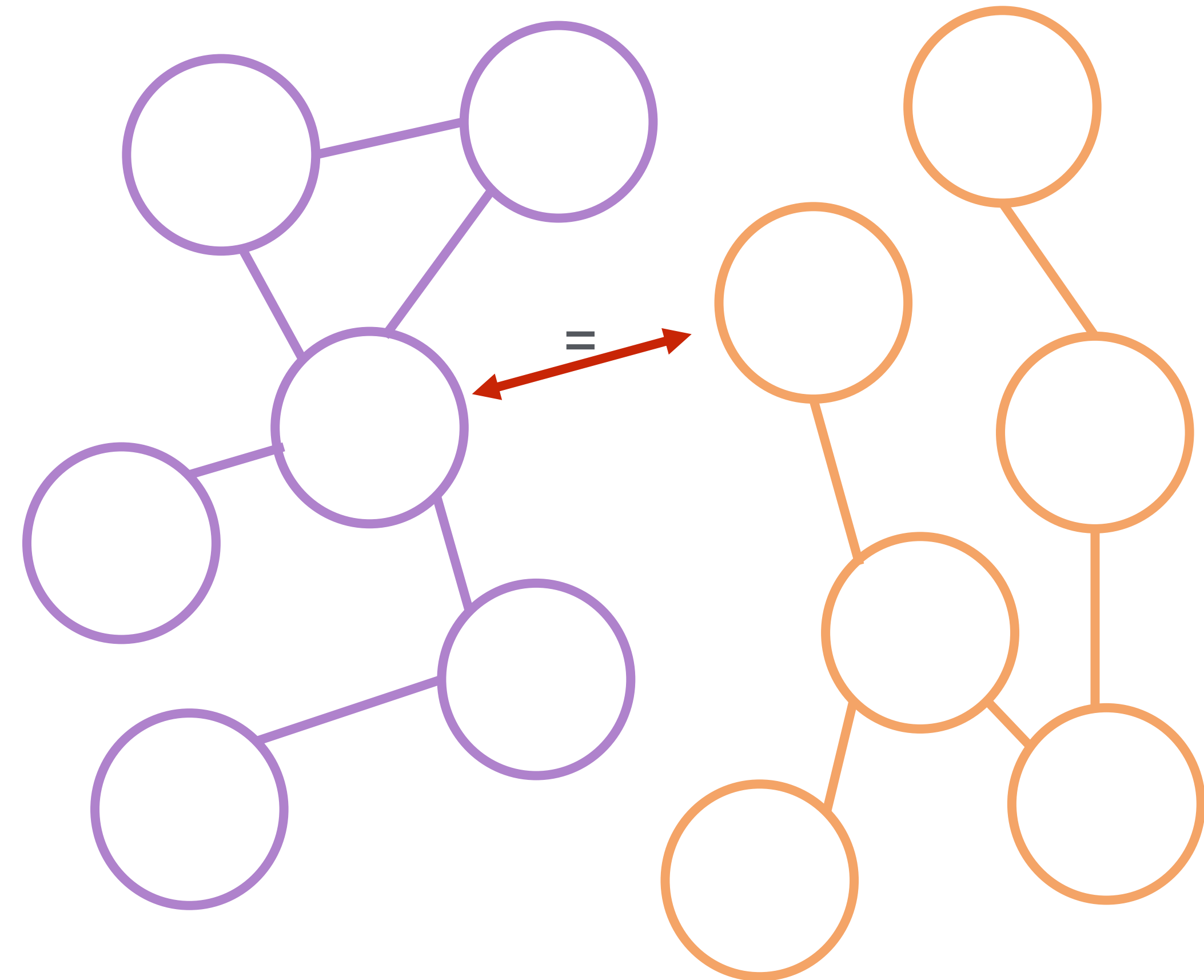




# 4.5 Entity Alignment



- Record Linkage
- DB Linkage
- Data Matching



- Entity Resolution
- Entity Alignment

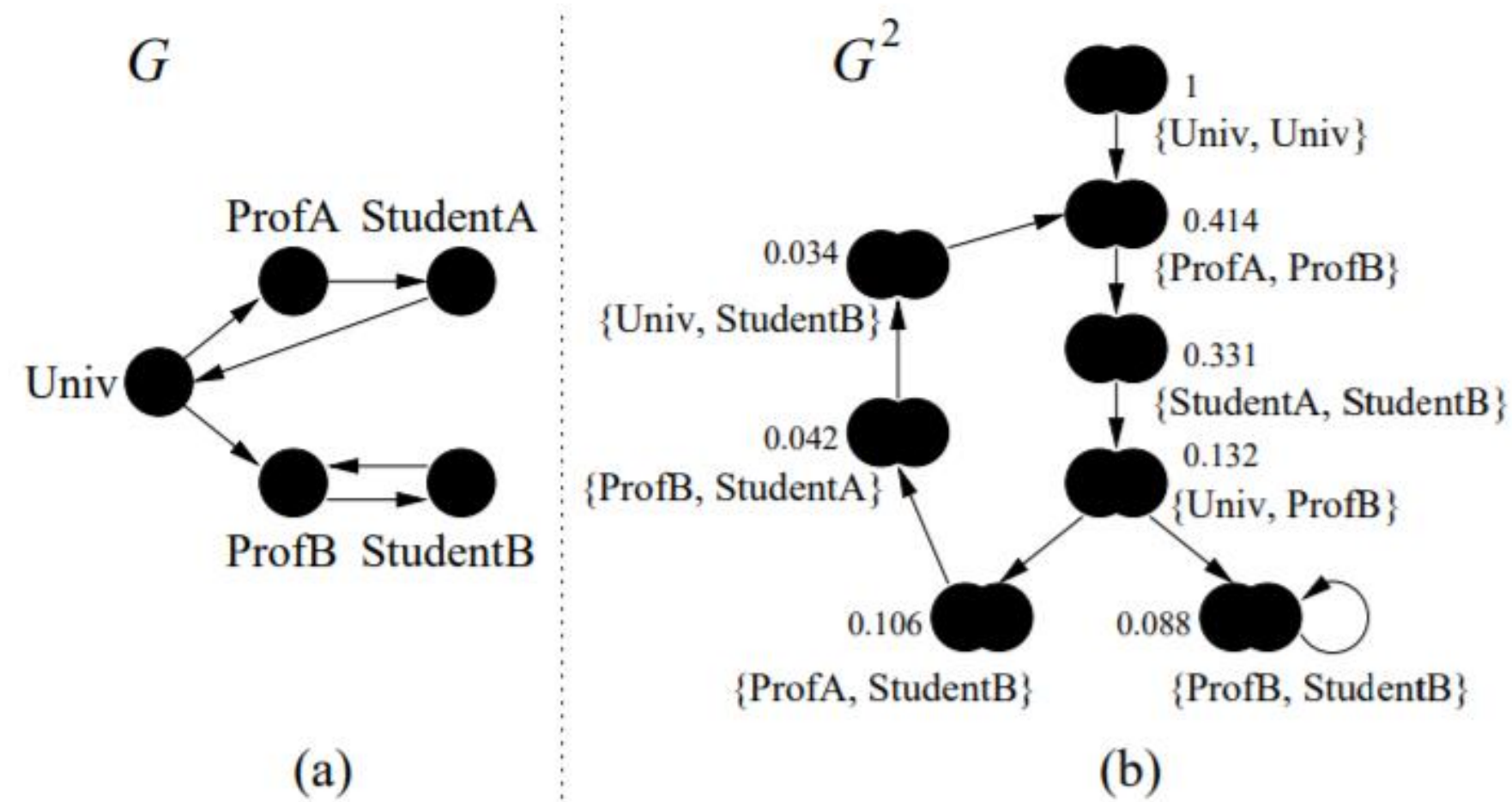
# 4.5 Entity Alignment

## SimRank: A Measure of Structural-Context Similarity\*

Glen Jeh  
glenj@db.stanford.edu

Jennifer Widom  
widom@db.stanford.edu

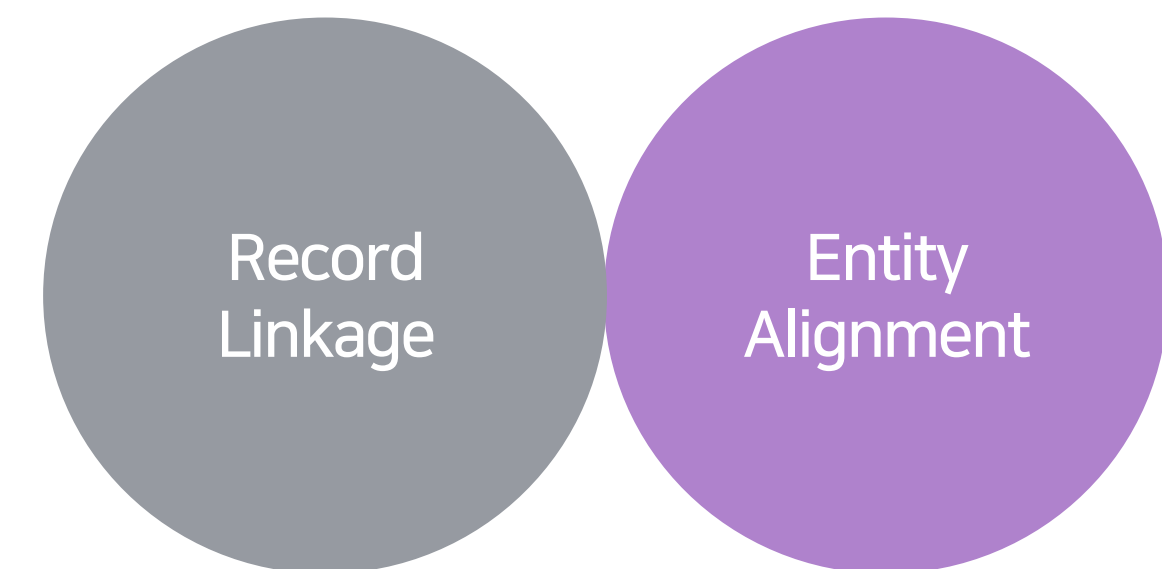
Stanford University



SimRank:

[1.	0.	0.818	0.	]
[0.	1.	0.	0.818	]
[0.818	0.	1.	0.	]
[0.	0.818	0.	1.	]]

Figure 1: A small Web graph  $G$  and simplified node-pairs graph  $G^2$ . SimRank scores using parameter  $C = 0.8$  are shown for nodes in  $G^2$ .



# 4.5 Entity Alignment

## Delta-SimRank Computing on MapReduce

Liangliang Cao  
IBM Watson Research Center  
liangliang.cao@us.ibm.com

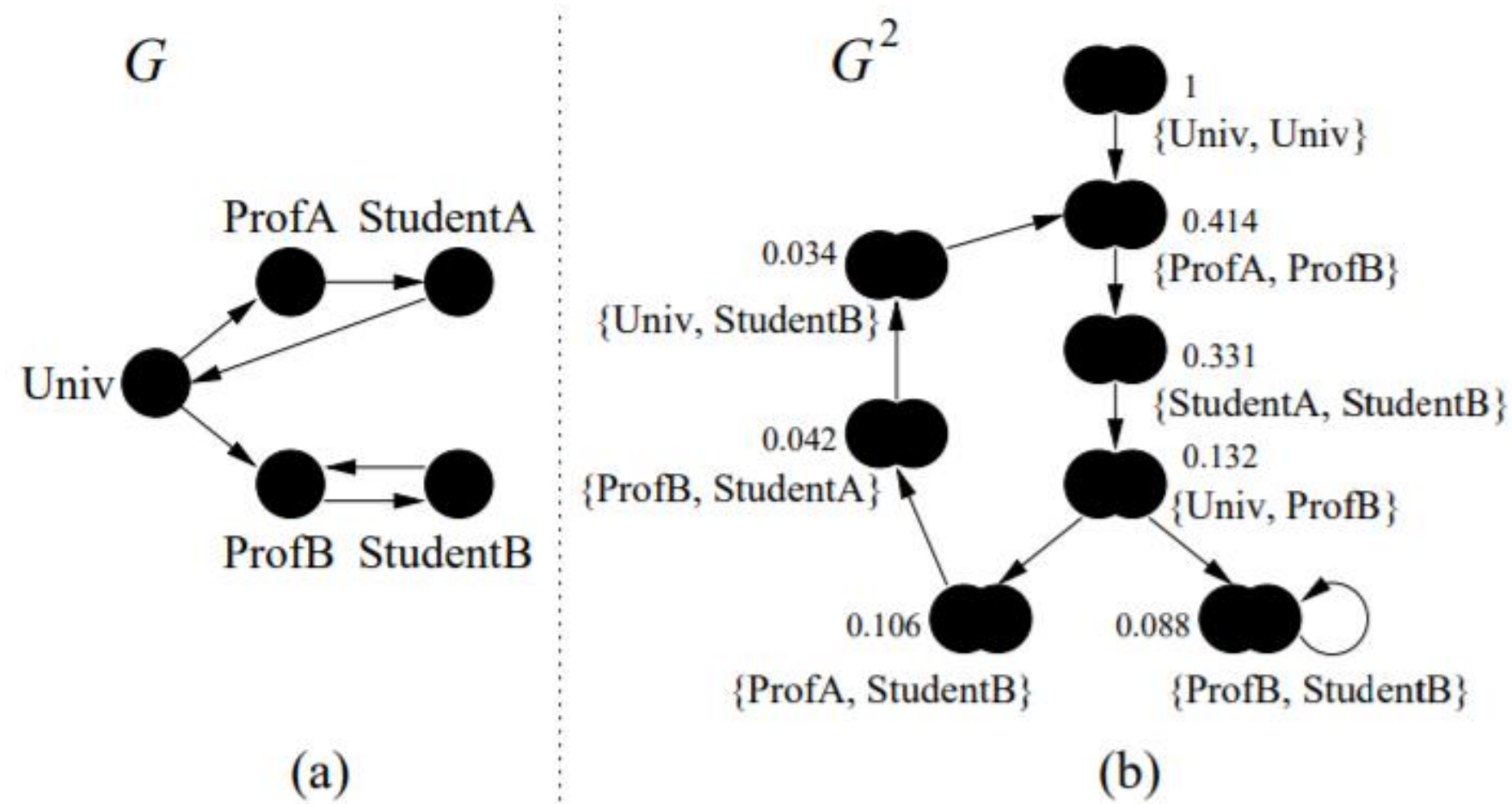
Hyun Duk Kim,  
University of Illinois at  
Urbana-Champaign  
hkim277@illinois.edu

Min-Hsuan Tsai  
University of Illinois at  
Urbana-Champaign  
mtsai2@illinois.edu

Brian Cho  
University of Illinois at  
Urbana-Champaign  
bcho2@illinois.edu

Zhen Li  
University of Illinois at  
Urbana-Champaign  
zhenli3@illinois.edu

Indranil Gupta  
University of Illinois at  
Urbana-Champaign  
indy@illinois.edu



### Algorithm 2 : Computing Delta-SimRank on MapReduce

**Input:** Graph  $G$ , initialized  $\Delta^t$

- 1: *Map function*  $((a,b), \Delta^t(a,b))$
- 2: if  $a = b$  or  $\Delta^t(a,b) \leq \epsilon$
- 3: return
- 4: find  $a, b$ 's neighbor  $I(a)$  and  $I(b)$  respectively
- 5: for each  $c \in I(a), d \in I(b)$
- 6: output  $(c,d), \frac{C}{|I(c)||I(d)|} \Delta^t(a,b)$
- 7: *Reduce function* (Key =  $(c,d)$ , Values =  $vs[]$ )
- 8: if  $c = d$
- 9: output  $\Delta^{t+1}(c,d) = 0$
- 10: else
- 11: output  $\Delta^{t+1} = \text{sum}(vs)$

**Output:** updated  $\Delta^{t+1}$

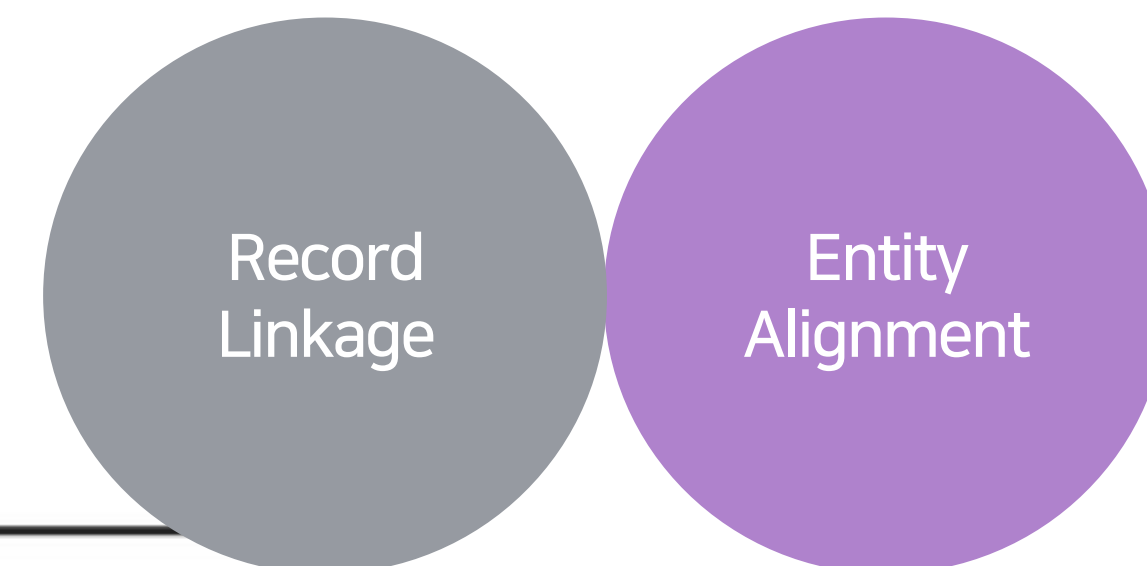


Figure 1: A small Web graph  $G$  and simplified node-pairs graph  $G^2$ . SimRank scores using parameter  $C = 0.8$  are shown for nodes in  $G^2$ .

# 4.5 Entity Alignment - Probabilistic

## PARIS: Probabilistic Alignment of Relations, Instances, and Schema

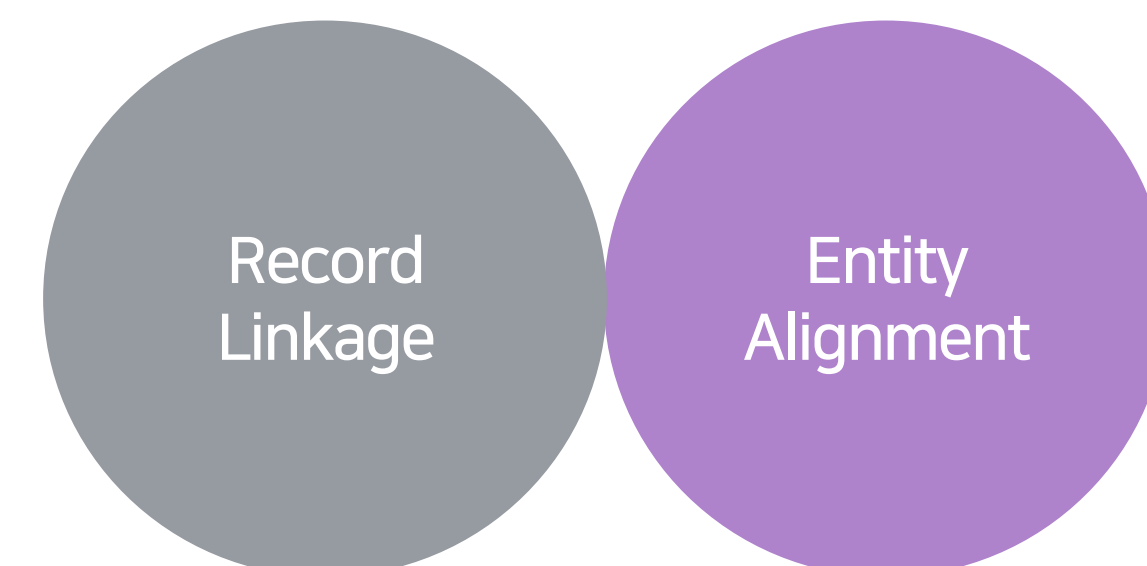
Fabian M. Suchanek  
INRIA Saclay – Île-de-France  
4 rue Jacques Monod  
91893 Orsay Cedex, France  
fabian@suchanek.name

Serge Abiteboul  
INRIA Saclay – Île-de-France  
4 rue Jacques Monod  
91893 Orsay Cedex, France  
serge.abiteboul@inria.fr

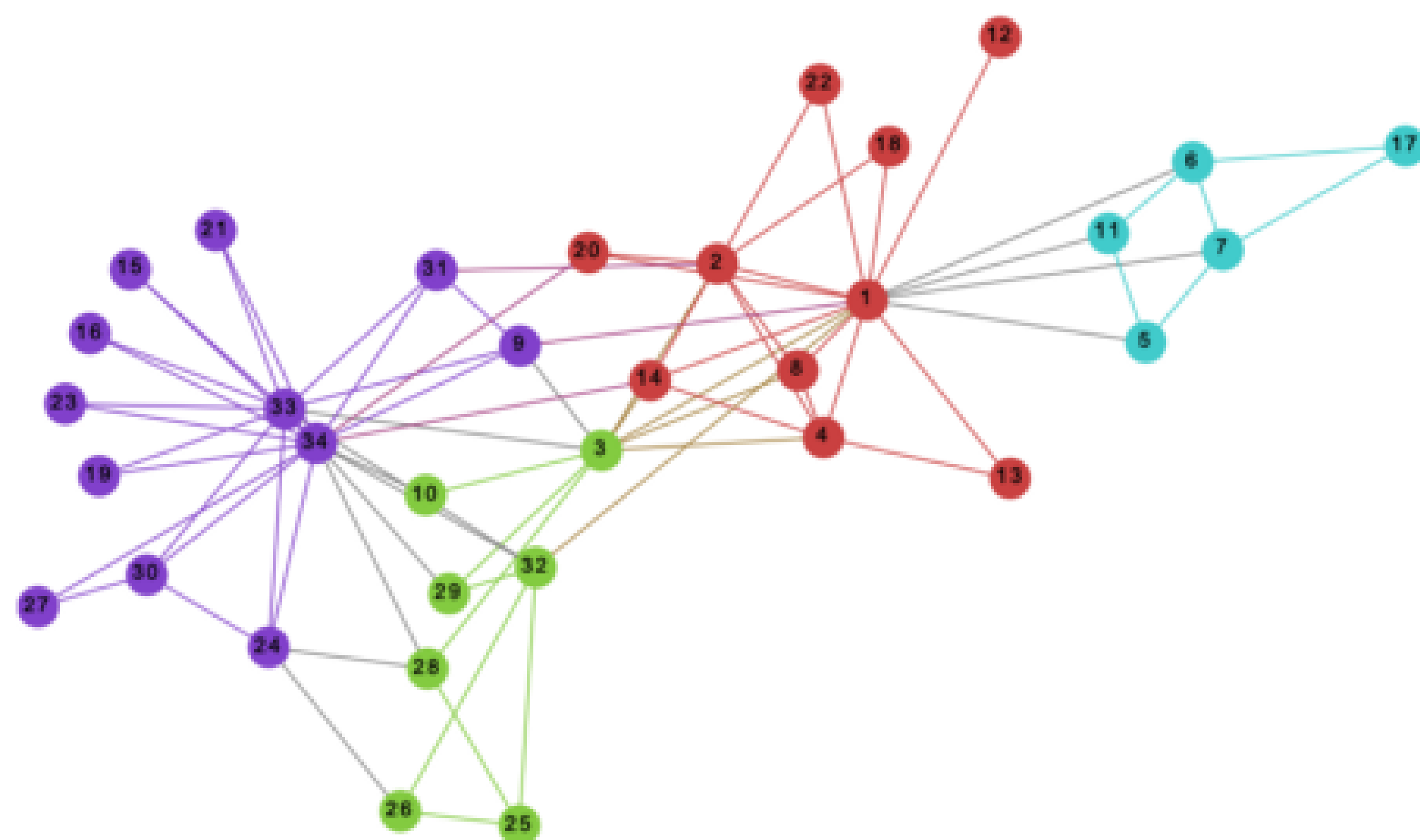
Pierre Senellart  
Institut Télécom;  
Télécom ParisTech; CNRS LTCI  
75634 Paris Cedex 13, France  
pierre.senellart@telecom-paristech.fr

- Equivalence of Instances
- Subrelations
- Subclasses

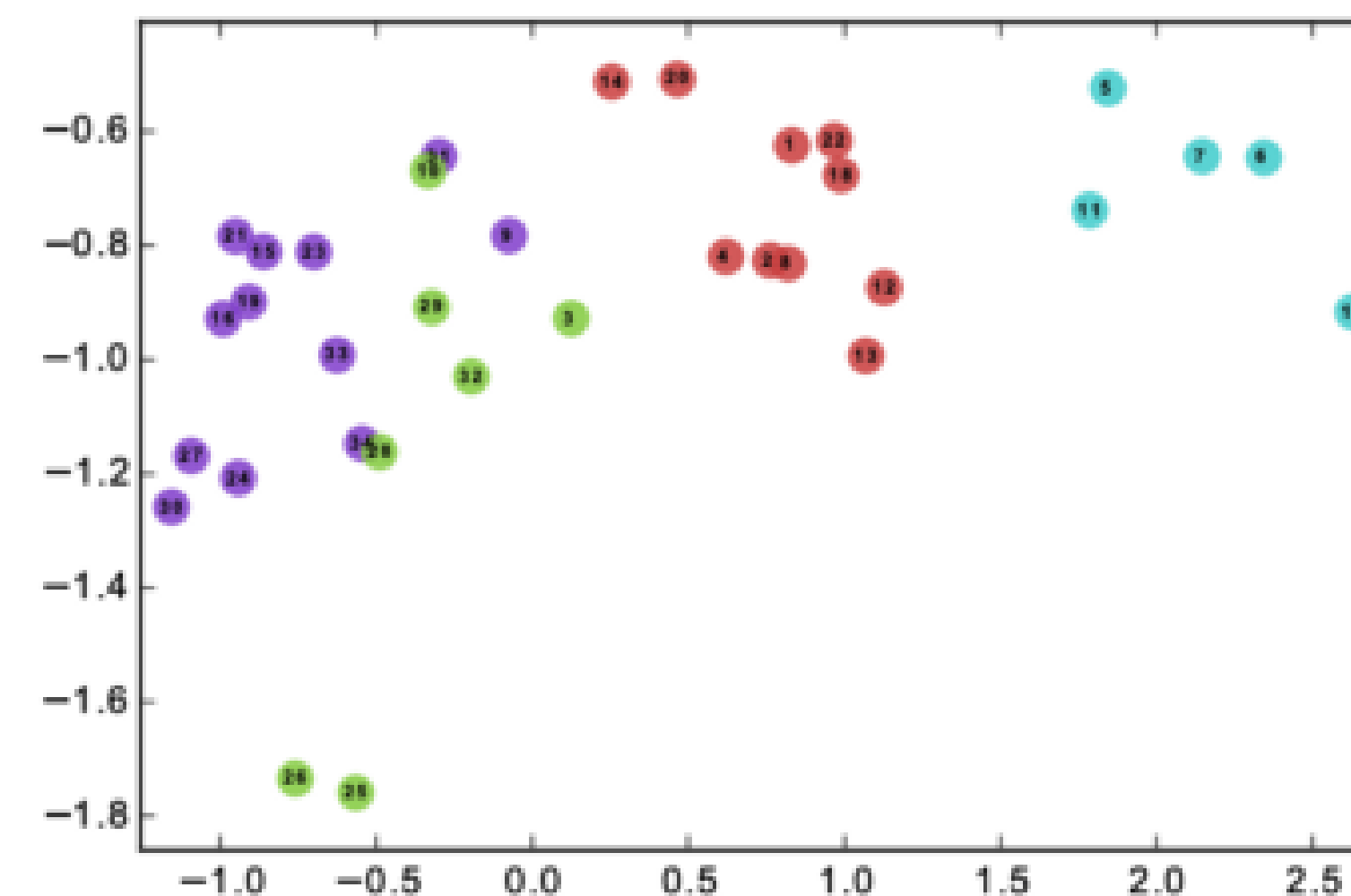
$$\frac{\sum_{r(x,y)} \left( 1 - \prod_{r'(x',y')} (1 - (\Pr(x \equiv x') \times \Pr(y \equiv y'))) \right)}{\sum_{r(x,y)} \left( 1 - \prod_{x',y'} (1 - P(x \equiv x') \times \Pr(y \equiv y')) \right)}$$



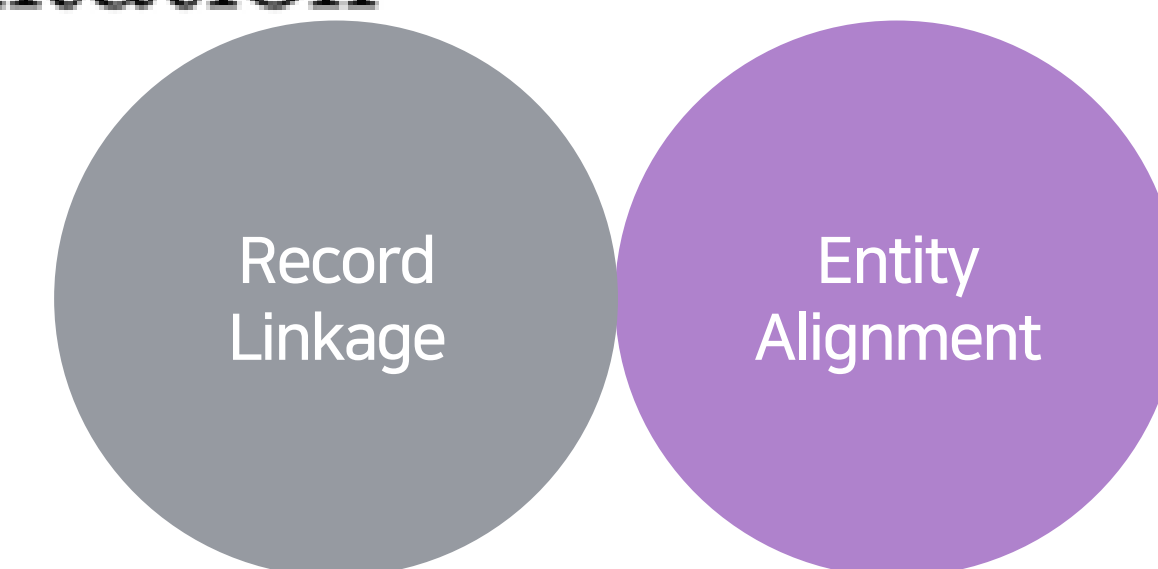
# 4.5 Entity Alignment - Embedding



(a) Input: Karate Graph



(b) Output: Representation

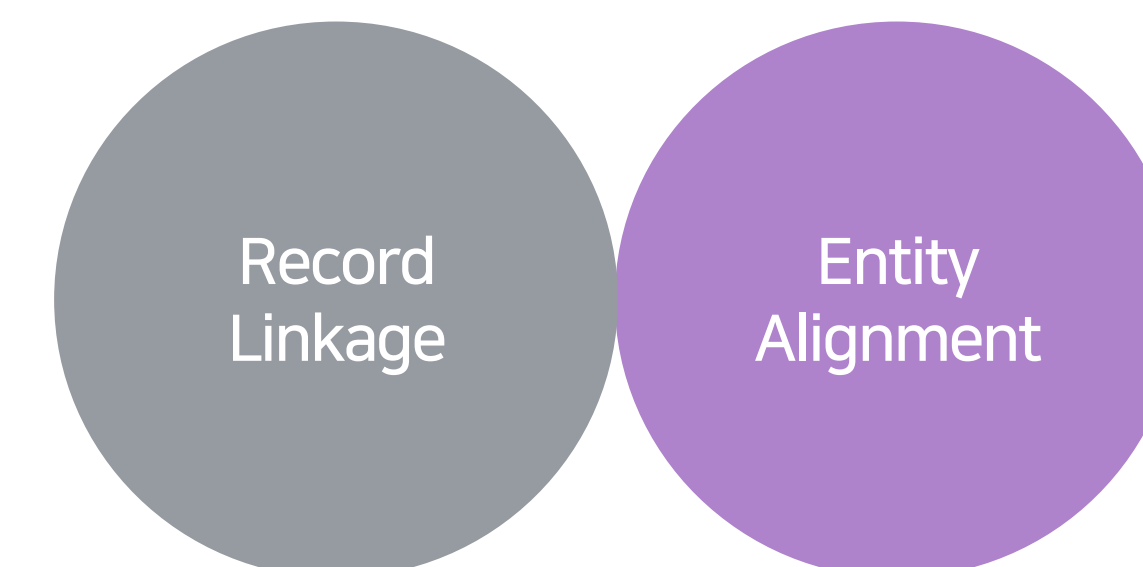


# 4.5 Entity Alignment - Embedding

Results on DBP15K(zh\_en, ja\_en, fr\_en).

	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
MTransE	0.419	0.753	0.535	0.433	0.773	0.549	0.407	0.751	0.526
BootEA	0.490	0.793	0.593	0.499	0.813	0.605	0.515	0.838	0.623
TransEdge	0.519	0.813	0.621	0.526	0.825	0.632	0.397	0.824	0.543
MMEA	0.405	0.672	0.499	0.397	0.680	0.496	0.442	0.749	0.550
GCN-Align	0.410	0.756	0.527	0.442	0.810	0.566	0.430	0.813	0.557
NAEA	0.323	0.481	0.381	0.311	0.457	0.363	0.307	0.460	0.362
KECG	0.467	0.815	0.586	0.485	0.843	0.605	0.479	0.844	0.602
TransE	0.343	0.634	0.441	0.365	0.710	0.480	0.374	0.735	0.493
TransH	0.436	0.735	0.540	0.450	0.778	0.561	0.485	0.821	0.599
TransR	0.371	0.697	0.481	0.368	0.709	0.484	0.378	0.741	0.497
RotatE	0.423	0.754	0.534	0.448	0.785	0.561	0.439	0.800	0.560
HAKA	0.288	0.588	0.391	0.319	0.607	0.421	0.319	0.638	0.428
DistMult	0.180	0.400	0.255	0.058	0.179	0.099	0.095	0.285	0.157
Complex	0.115	0.265	0.166	0.063	0.251	0.146	0.141	0.332	0.206
ConvE	0.210	0.466	0.299	0.339	0.556	0.415	0.350	0.602	0.439

< 0.65



# 4.5 Entity Alignment - pattern

## Ontology-based Entity Matching in Attributed Graphs

Hanchao Ma  
Washington State University  
hanchao.ma@wsu.edu

Morteza Alipourlangouri  
McMaster University  
alipoum@mcmaster.ca

Yinghui Wu  
Washington State University  
Pacific Northwest National  
Laboratory  
yinghui.wu@wsu.edu

Fei Chiang  
McMaster University  
fchiang@mcmaster.ca

Jiaying Pi  
Siemens Corporate  
Technology  
jiaying.pi@siemens.com

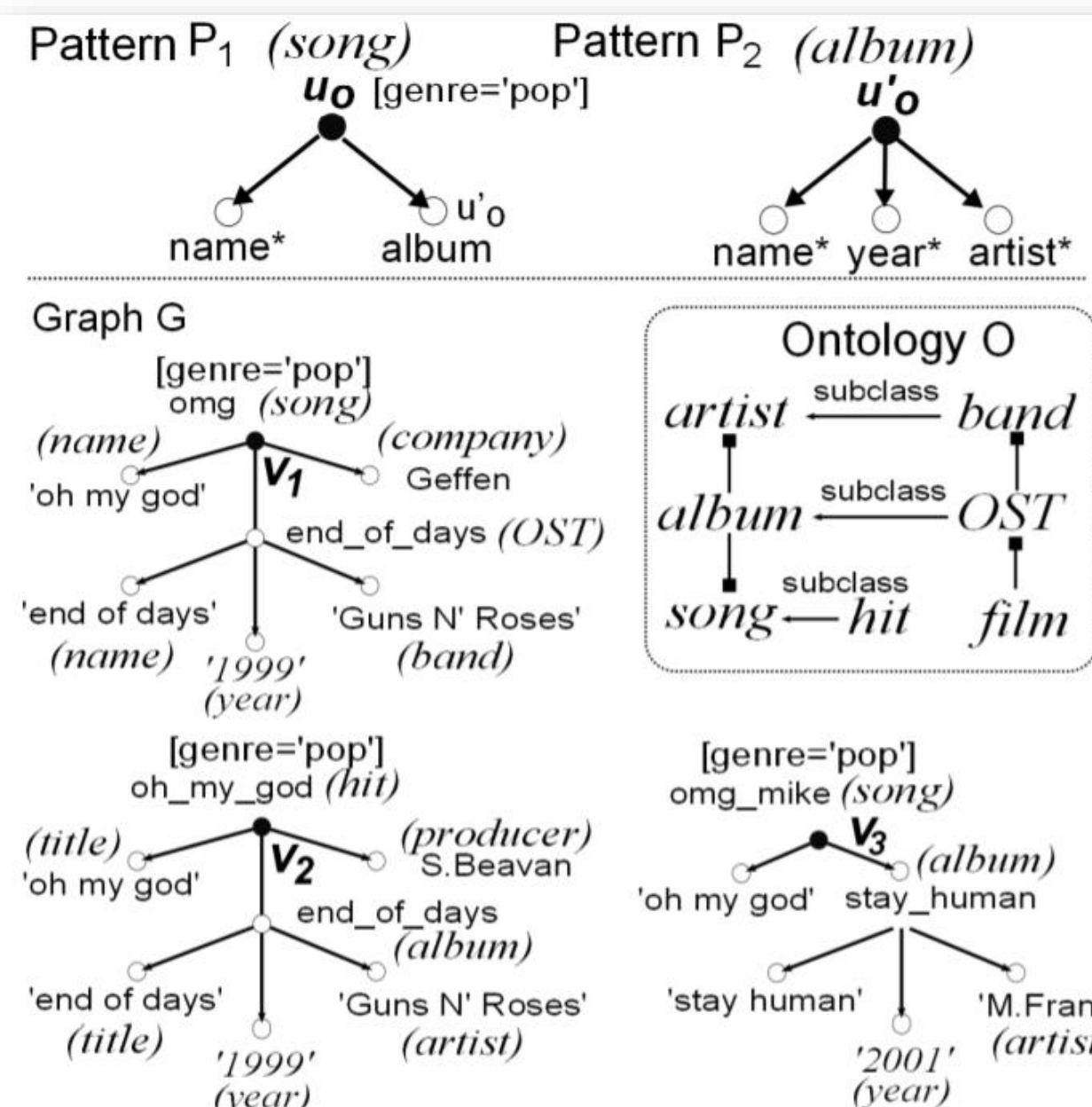


Figure 1: Entity matching with ontologies.

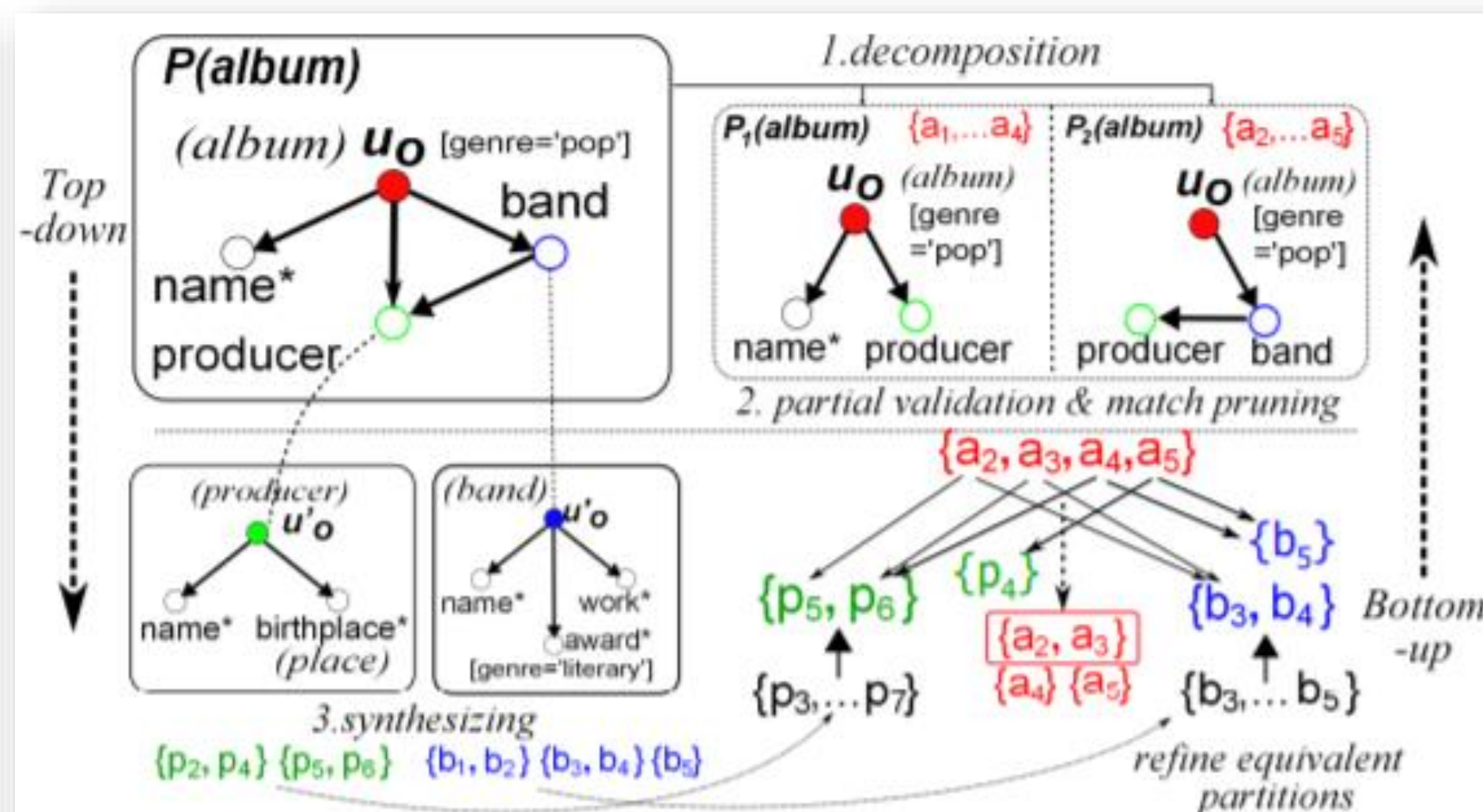


Figure 5: OGG-EM: Entity matching and pruning.

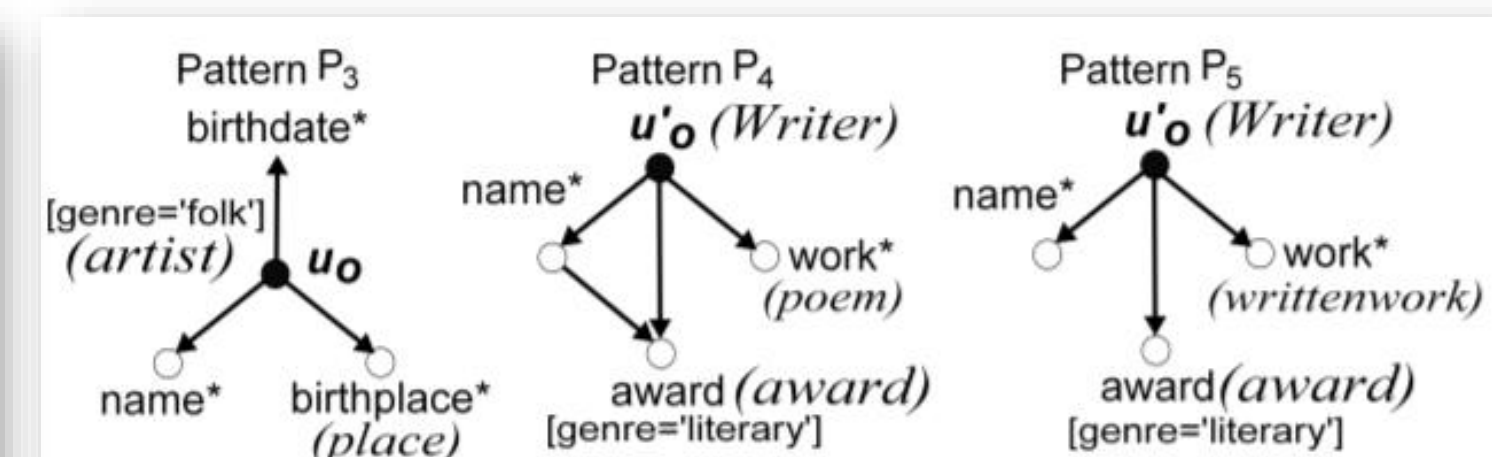
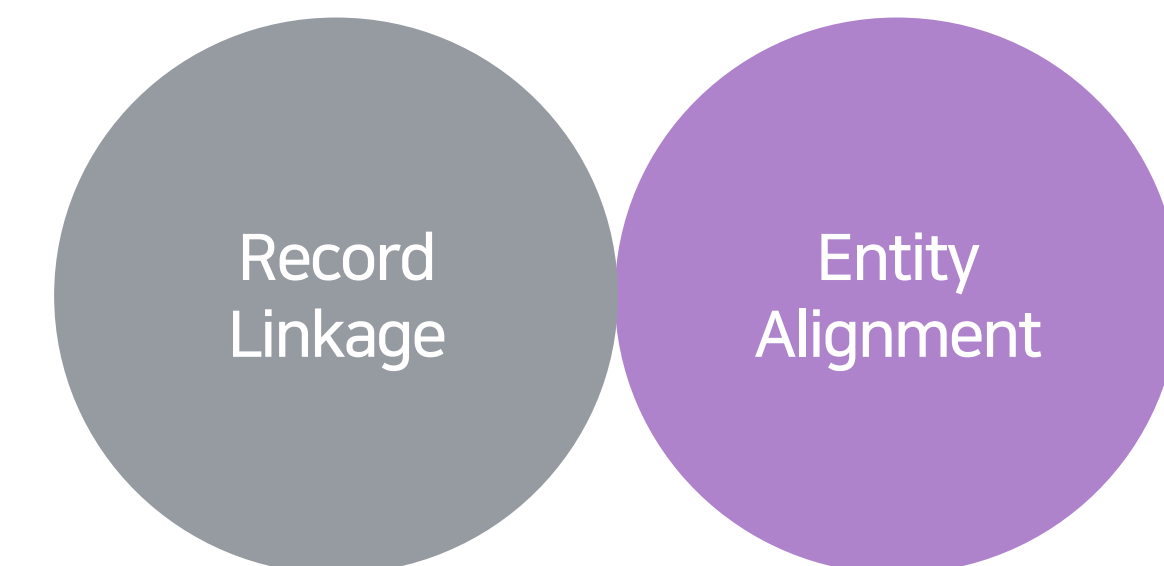


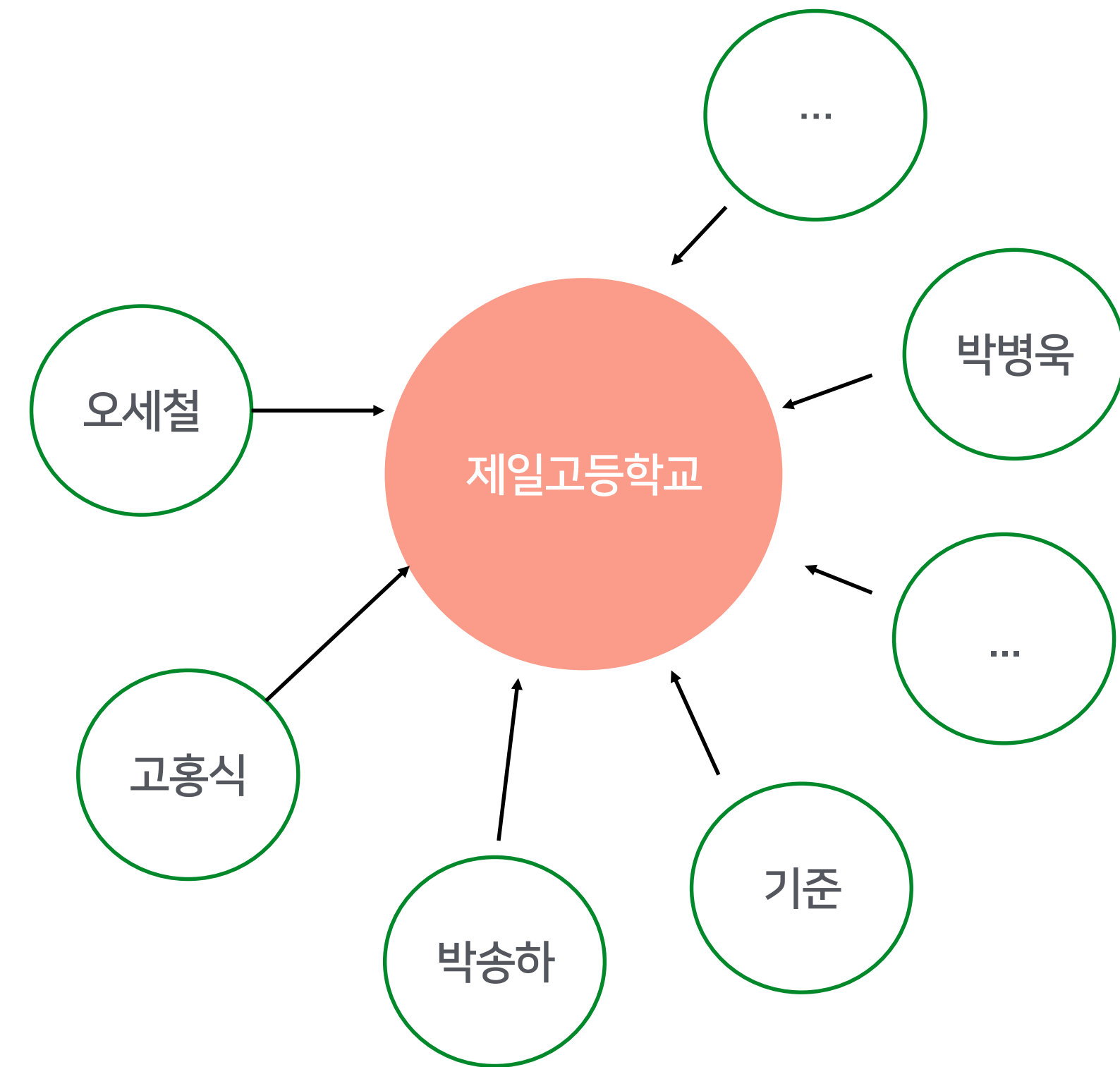
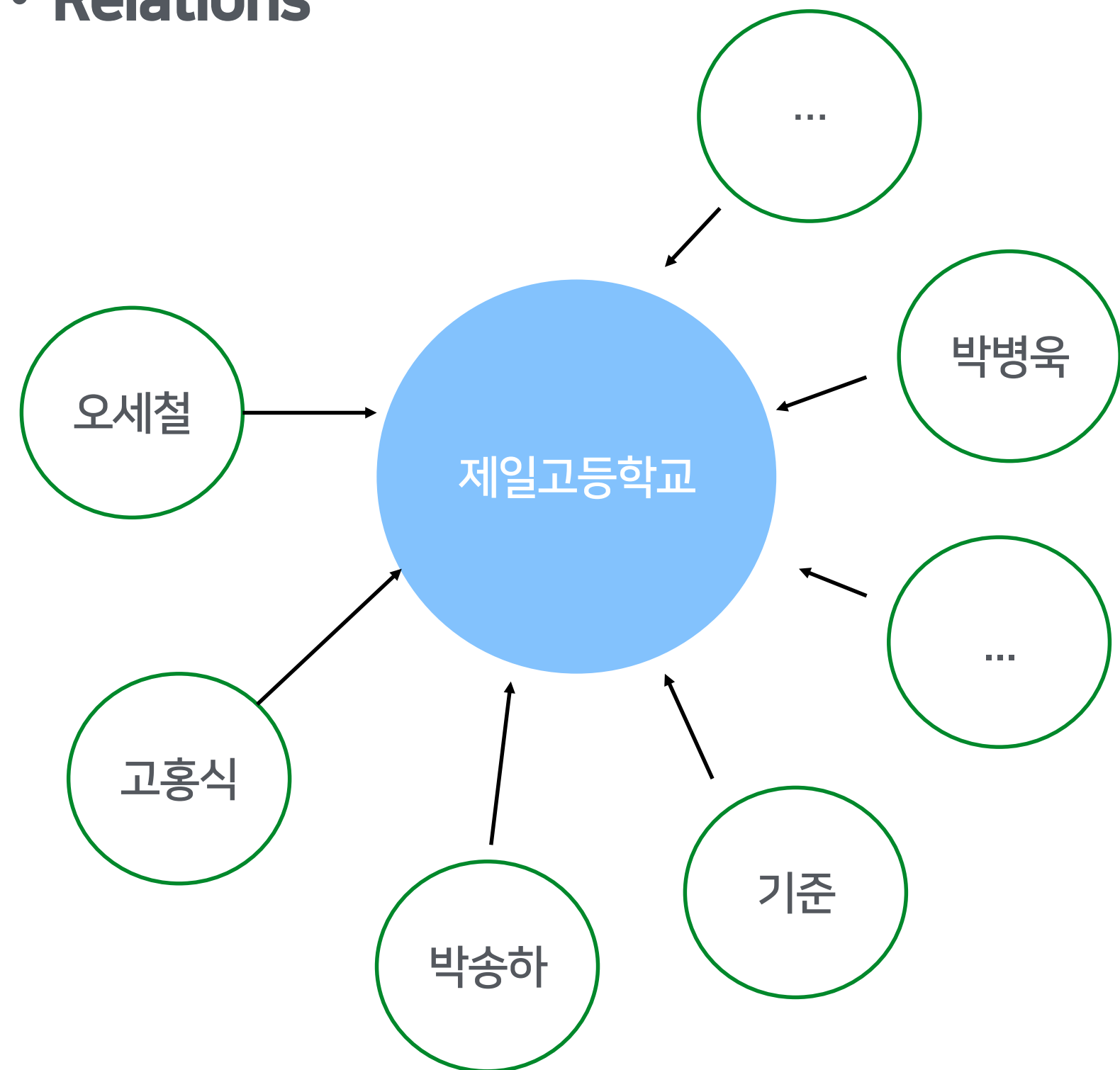
Figure 2: Ontological Graph Keys



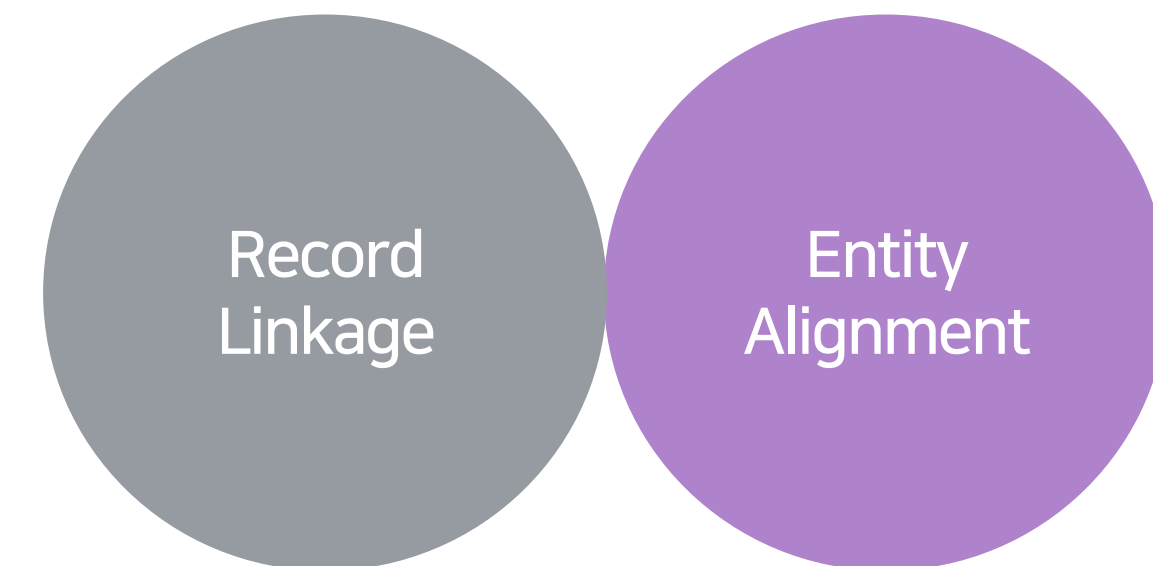
# 4.5 Entity Alignment - Ours

## 1. Ontology Alignment

- Taxonomy (class hierarchy)
- Relations ( Predicate, Property)
- Triple Pattern
- Structured Pattern
- Relations



<ztella:property/educated\_at>

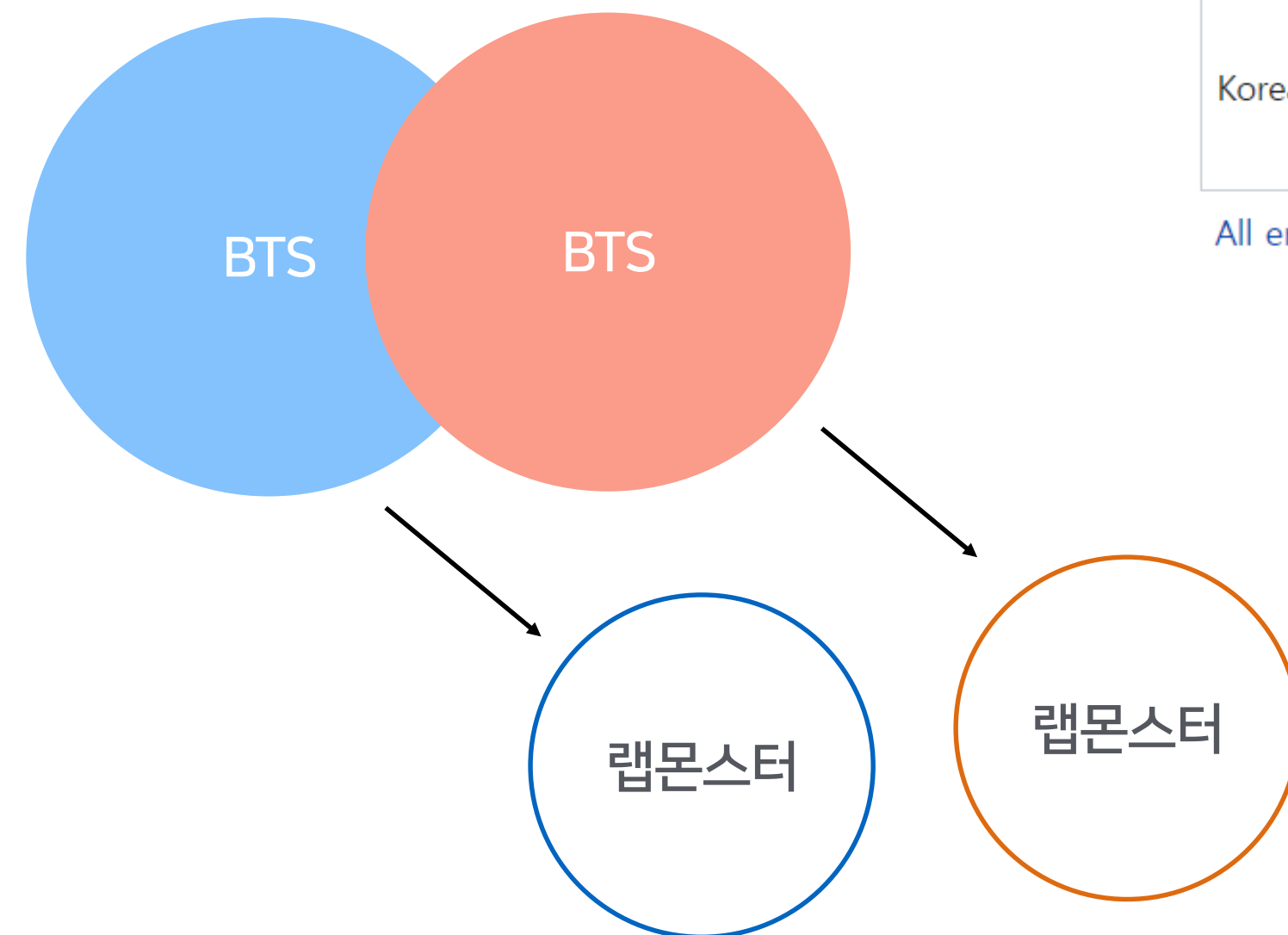




# 4.5 Entity Alignment - Ours

## 1. Ontology Alignment

- Taxonomy (class hierarchy)
- Relations ( Predicate, Property)
- Triple Pattern
- Structured Pattern
- Relations



<ztella:property/member>

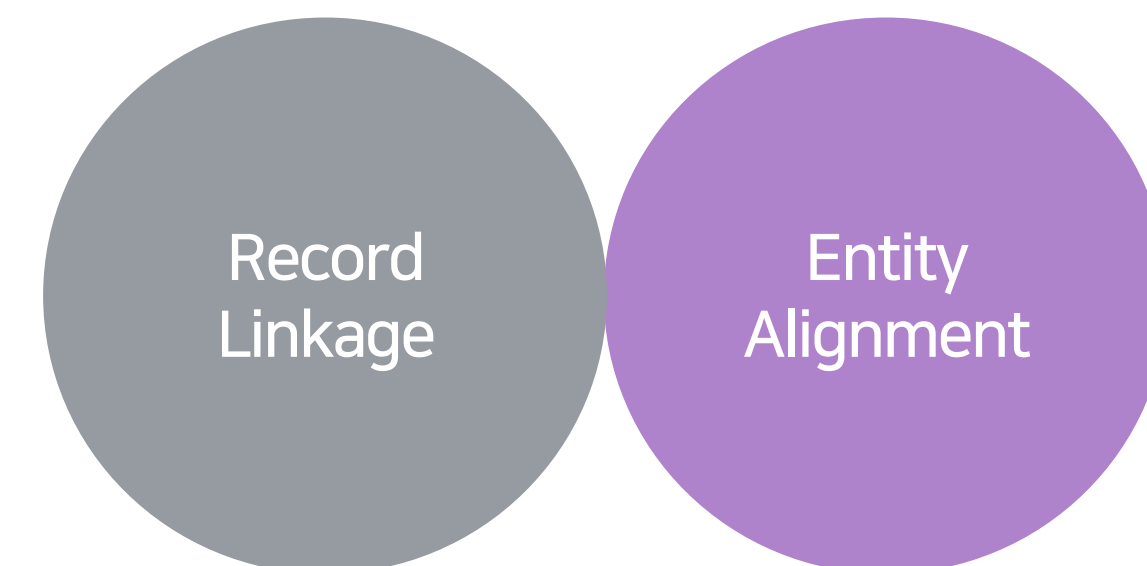
### RM (Q20514446)

South Korean rapper, songwriter, record producer, and leader of the boy band BTS  
Kim Nam-joon | Rap Monster

▼ In more languages  
Configure

Language	Label	Description	Also known as
English	RM	South Korean rapper, songwriter, record producer, and leader of the boy band BTS	Kim Nam-joon Rap Monster
Korean	김남준	대한민국 7인조 보이그룹 방탄소년단의 리더,메인래퍼	랩몬스터 알엠

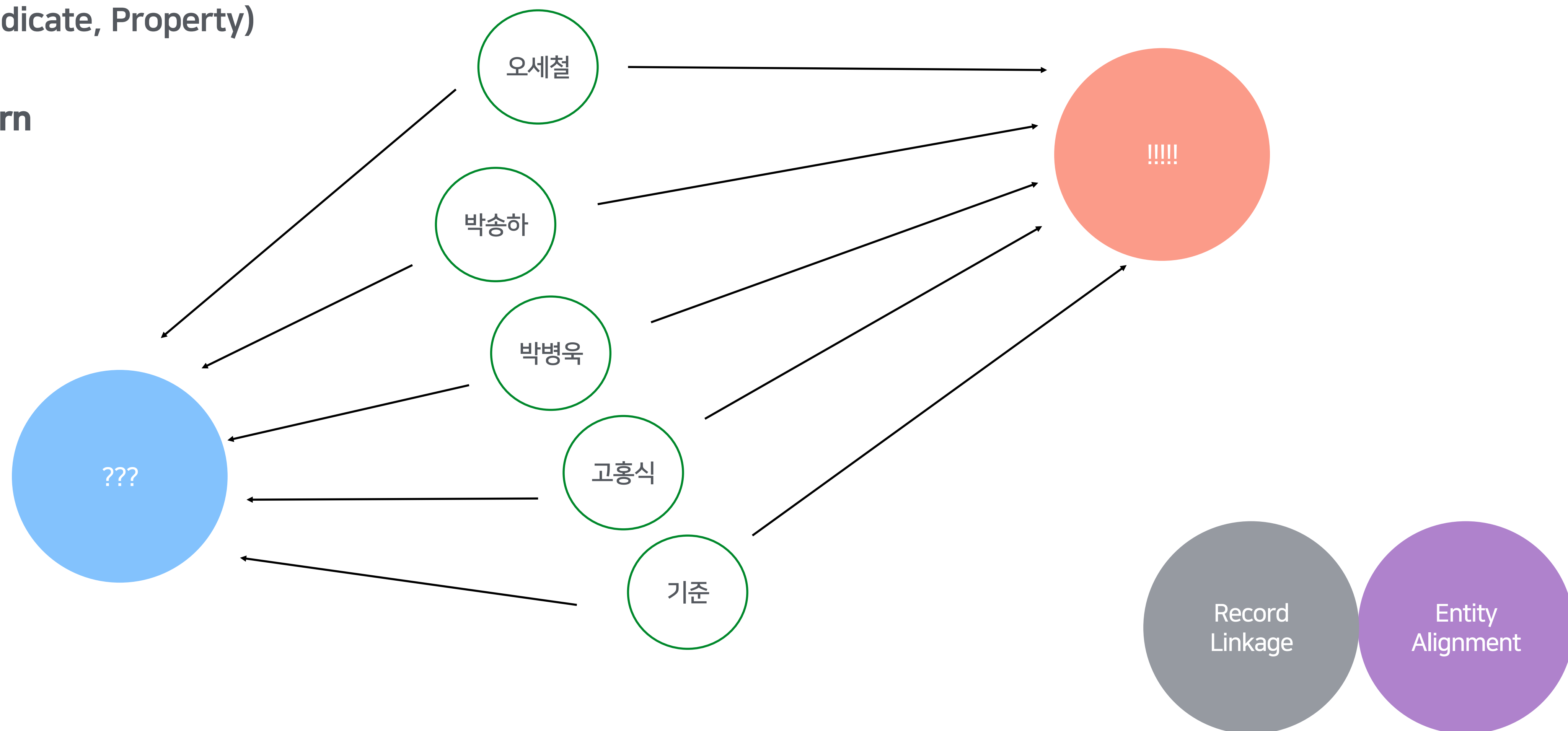
All entered languages



# 4.5 Entity Alignment - Ours

## 1. Ontology Alignment

- Taxonomy (class hierarchy)
- Relations ( Predicate, Property)
- Triple Pattern
- Structured Pattern
- Relations



# 5. Applications

# 5.1 음원 사이트 링크링

낙하 (with 아이유)  
노래 AKMU(악뮤) | 2021.07.26.

전체 곡정보 관련영상 수록앨범 관련 플레이리스트 비슷한 아티스트

03:51

곡정보

말했잖아 언젠가  
이런 날이 온다면  
난 널 혼자  
내버려두지 않을 거라고...

▶ 바로재생 MP3

유직 다른 사이트 더보기

VIBE에서 더보기 →

같은 앨범 수록곡

- 1 전쟁터 (with 이선희)
- 2 낙하 (with 아이유) TITLE
- 3 BENCH (with Zion.T)

관련 플레이리스트

세련된 모던 락 세련된 모던 락

한 템포 쉬어 가기 한 템포 쉬어 가기

월간 빌보드 K-Pop #9월 월간 빌보드 K-Pop #9월

요즘 국내 락/ 메탈 요즘 국내 락/ 메탈

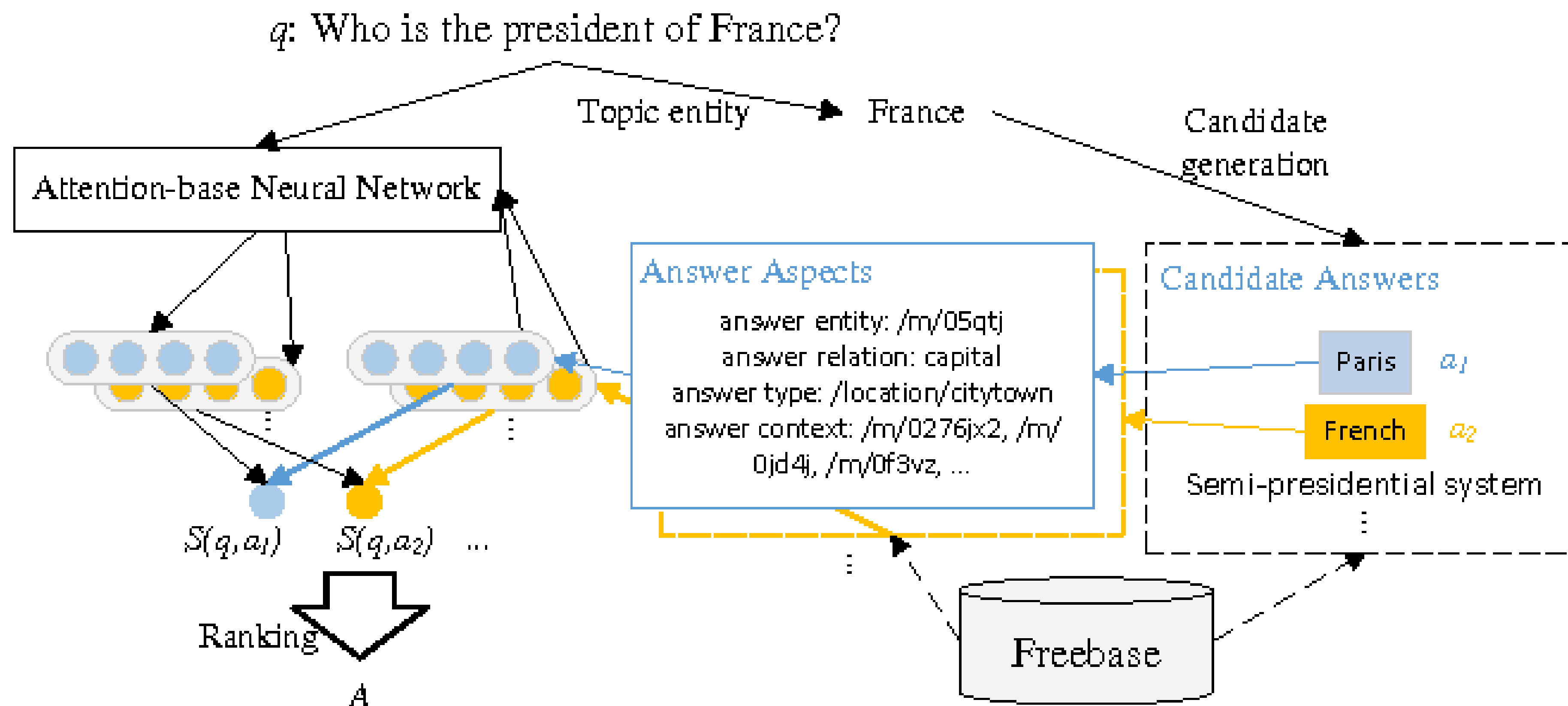
YG 대표곡 YG 대표곡

VIBE 최신곡 VIBE 최신곡

듣을 수 있는 곳

VIBE BUBS Spotify YouTube

# 5.2 Knowledge Base QA









# 5.3 Recommendations

**N** | 오징어게임

통합 이미지 VIEW 지식iN 인플루언서<sup>N</sup> 동영상 쇼핑 뉴스 어학사전 지도 ...






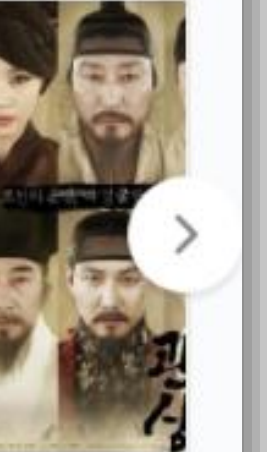
함께 볼만한 웹드라마 →

					
마이 네임	D.P.	지옥	킹덤: 아신전	이 구역의 미친 X	블루버스터

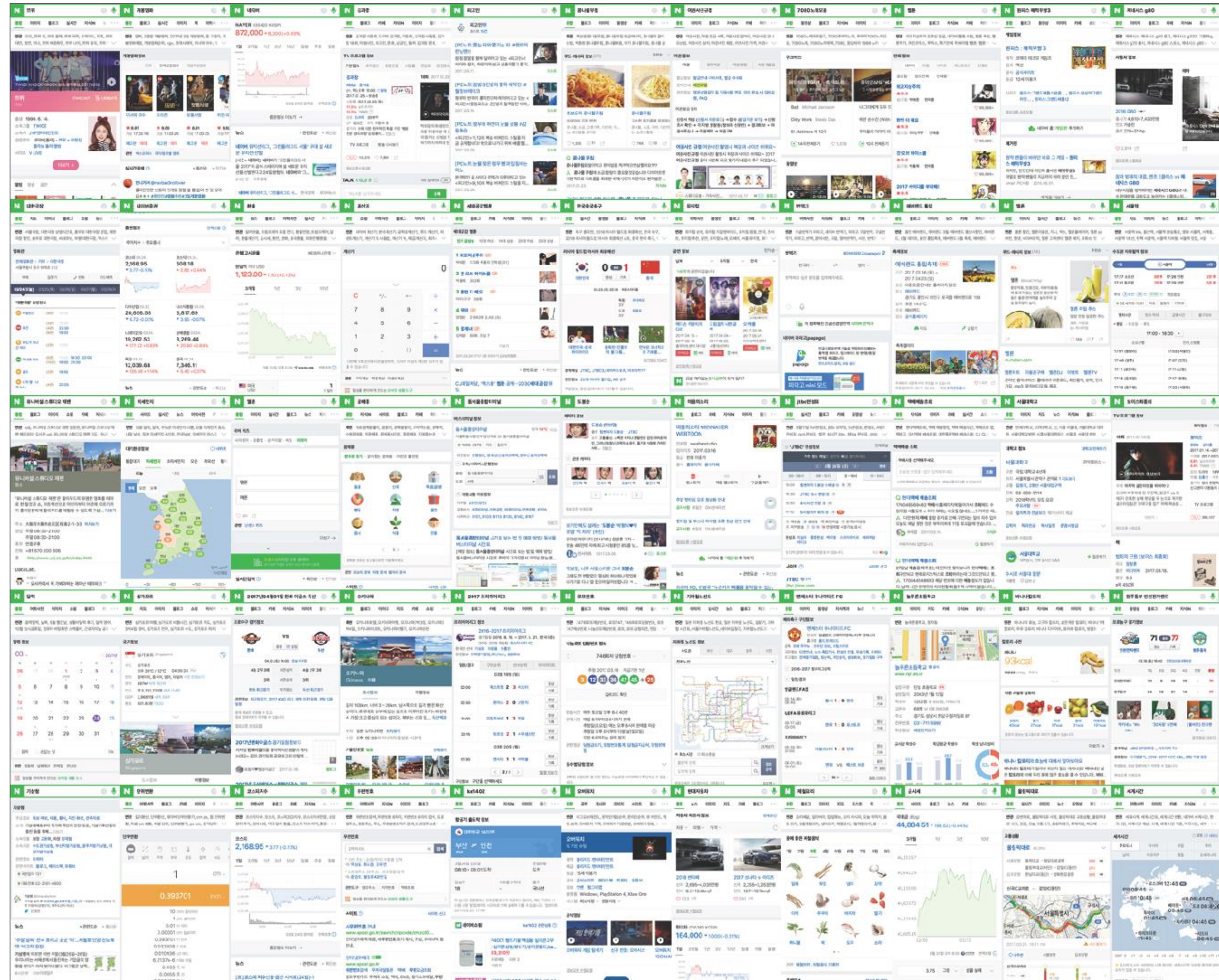
**N** | 태양은 없다

통합 이미지 VIEW 지식iN 인플루언서<sup>N</sup> 동영상 쇼핑 뉴스 어학사전 지도 ...

관련 영화 추천 →

					
비트 1997 개봉	아수라 2016 개봉	무사 2001 개봉	신세계 2013 개봉	암살 2015 개봉	관상 2013 개봉

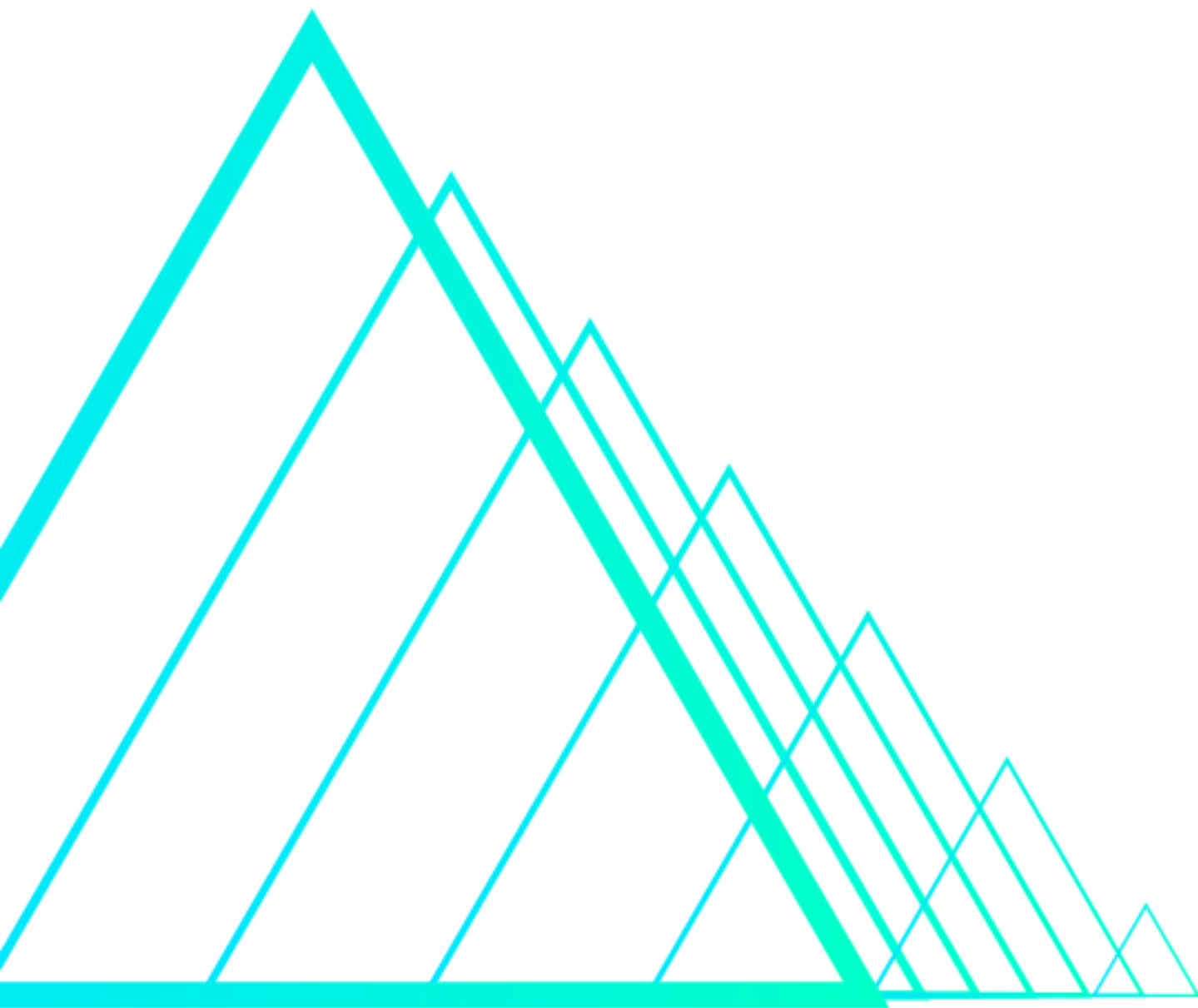
# <2019 DEVIEW 중>



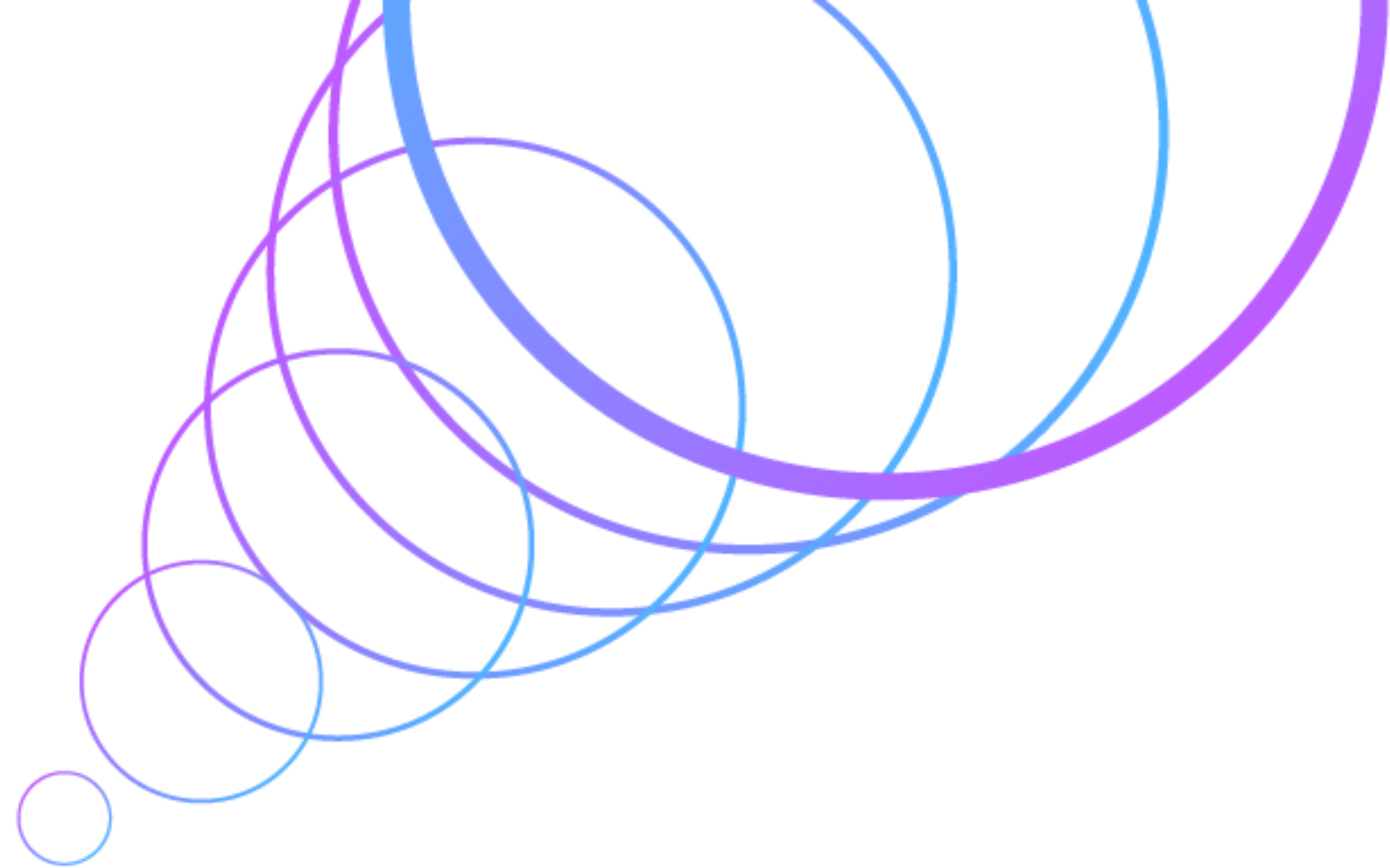
**DEVIEW 2019**

40여개 주제 250여개 컬렉션으로 서비스 중.

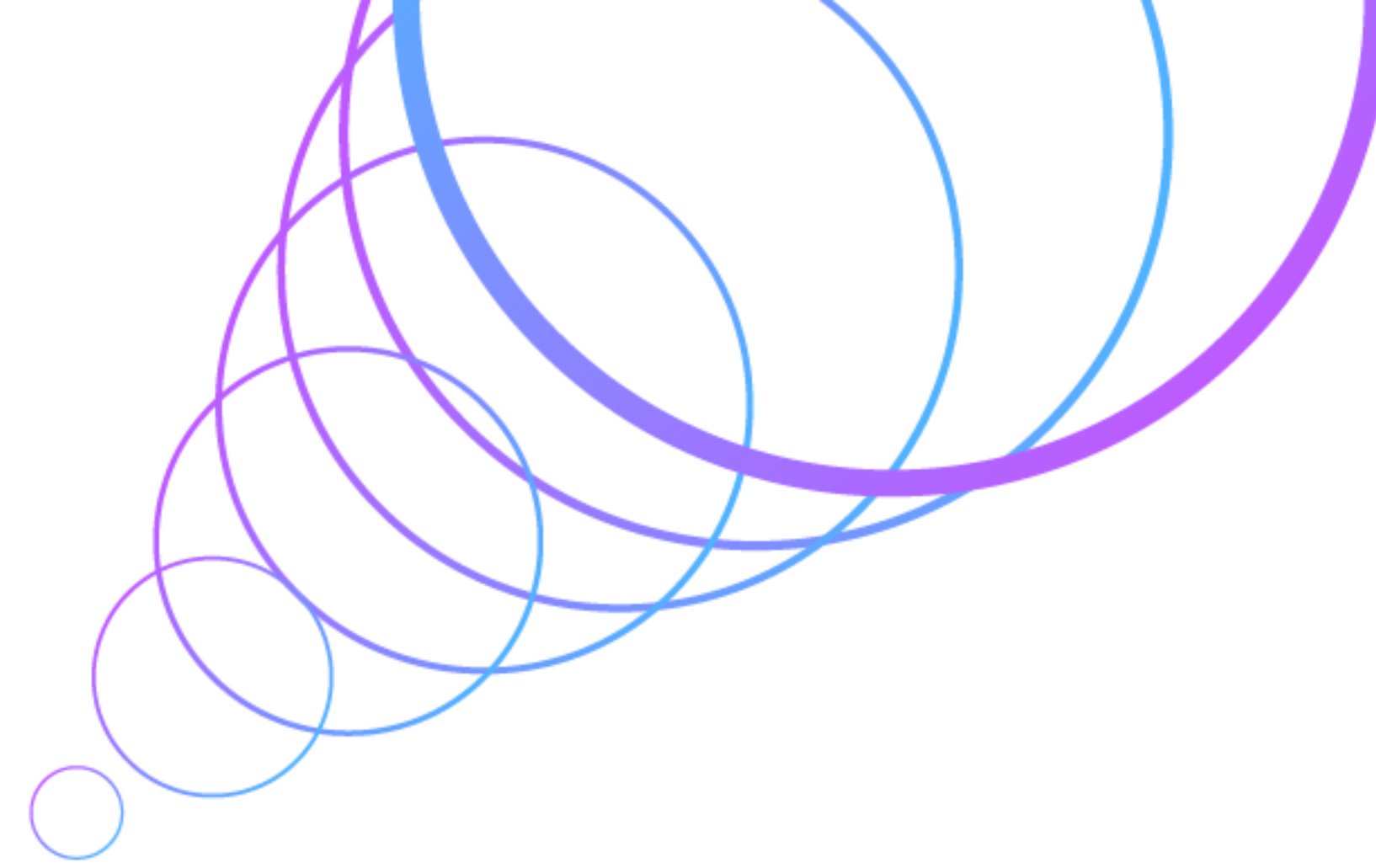
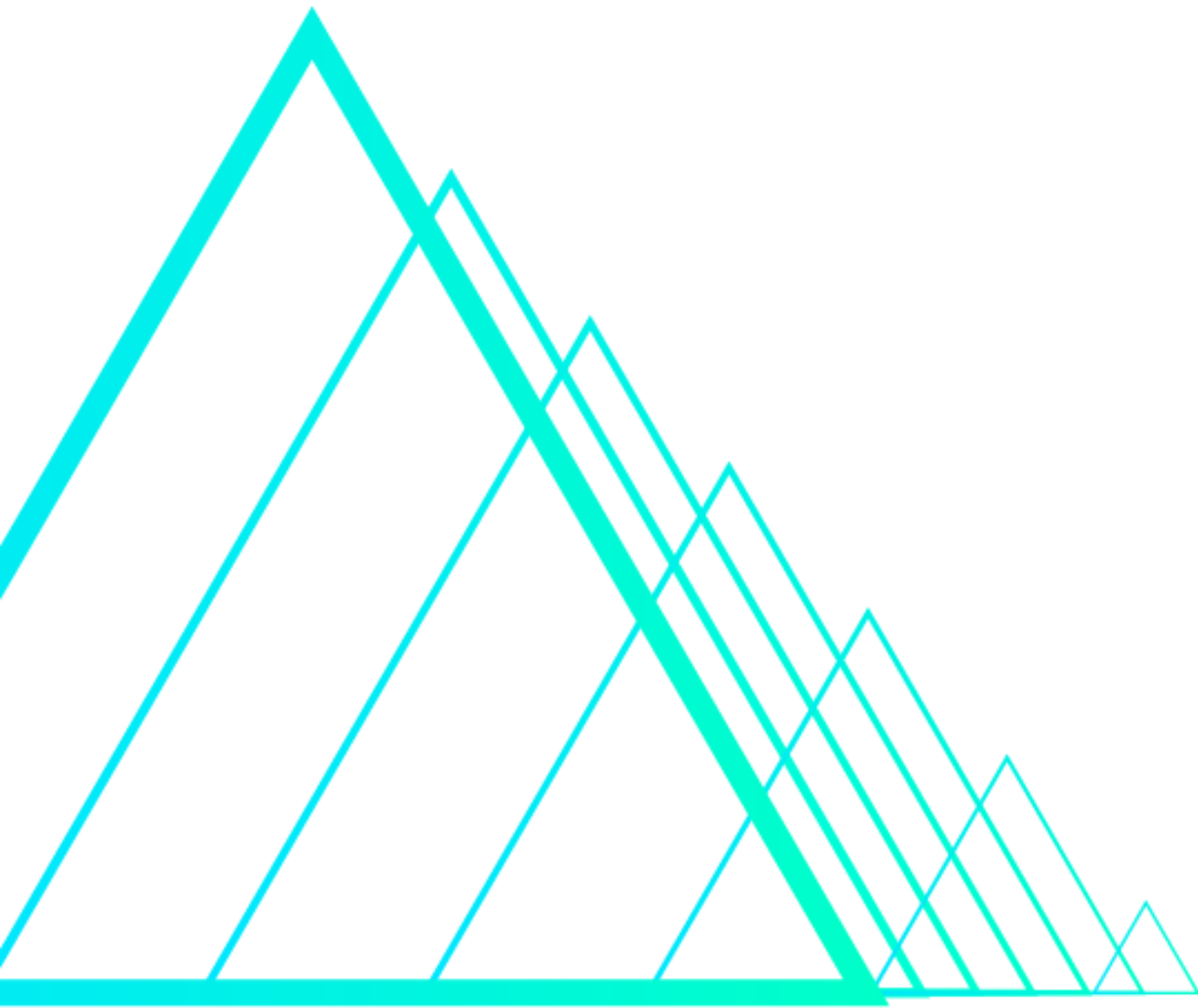
통합검색 전체 질의어 중 약 30% 대응 중.



# 맺음말







**Thank You**

