

Contextual Propagation of Properties for Knowledge Graphs

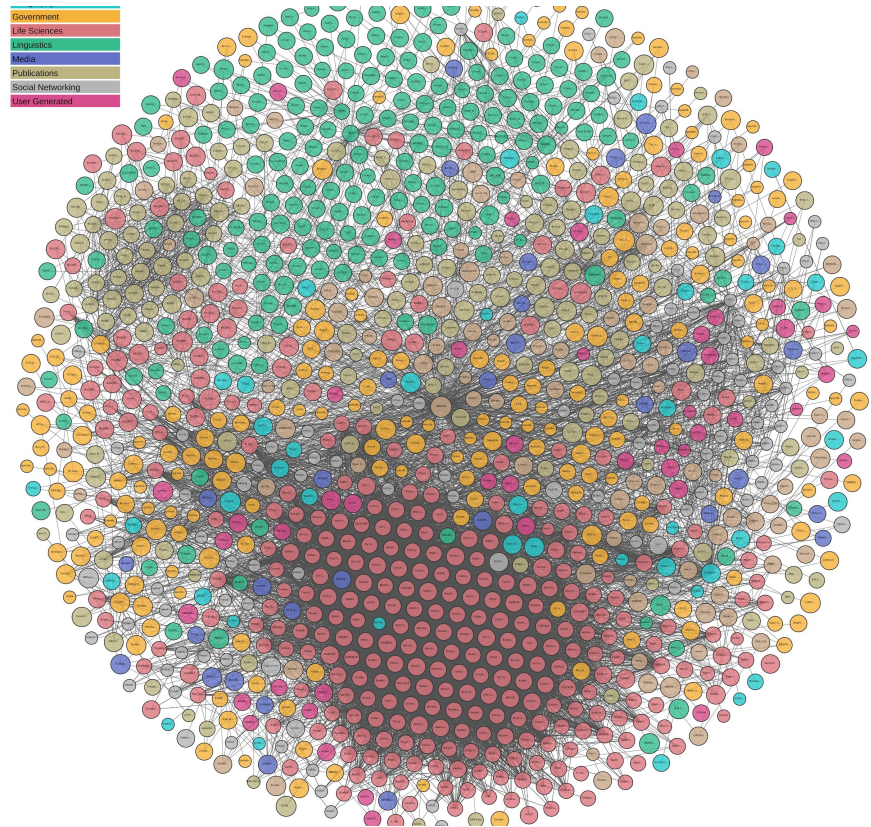
A Sentence Embedding Based Approach

Dr Pierre-Henri Paris, Dr Fayçal Hamdi, Dr Nopal Niraula, Pr Samira Si-Said Cherfi

The Linked Open Data

The principles are:

1. Use **URIs** as names for things
2. Use **HTTP URIs** so that people can look up those names
3. When someone looks up a URI, provide **useful information**
4. Include links to other URIs so that they can **discover more things**



Identity: A Very Old Problem

Antiquities:

- Theseus paradox:



- The law of identity: $A = A$

Leibniz's law of Identity

Identity of indiscernibles:

$$\forall x, \forall y (\forall p, \forall o, (\langle x, p, o \rangle \text{ and } \langle y, p, o \rangle) \rightarrow x = y)$$

Indiscernibility of identicals:

$$\forall x, \forall y (x = y \rightarrow \forall p, \forall o, (\langle x, p, o \rangle \rightarrow \langle y, p, o \rangle))$$

owl:sameAs semantics is based on these principles

⇒ property-value couples can be **propagated** from one entity to another identical entity and thus, **increase completeness**

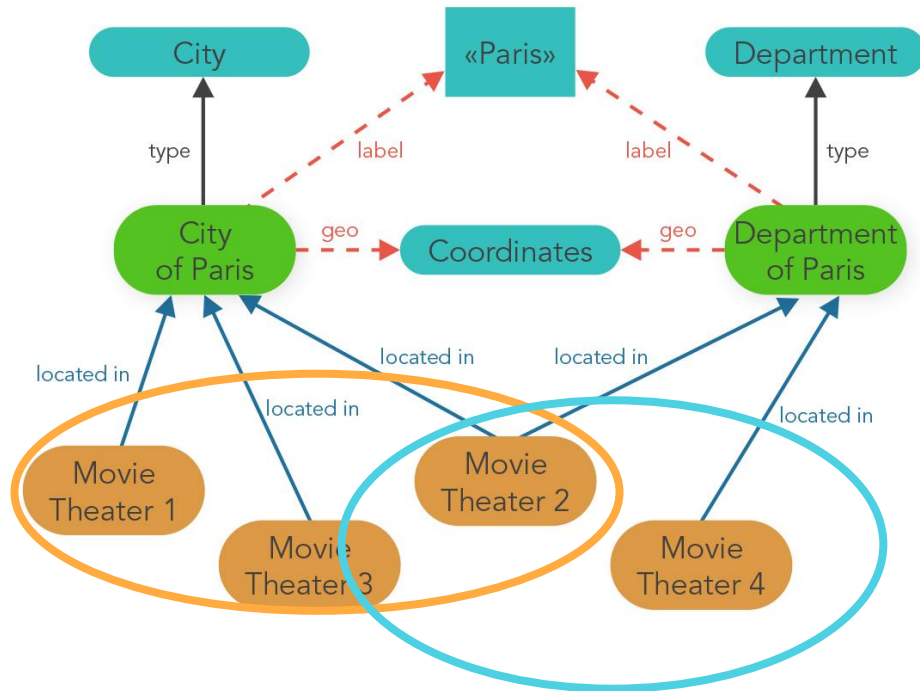
Identity Crisis

- Guarino, N., & Welty, C. (2002). Evaluating ontological decisions with OntoClean. *Communications of the ACM*, 45(2), 61-65.
- Bouquet, P., Stoermer, H., & Giacomuzzi, D. (2007). OKKAM: Enabling a Web of Entities. *13*, 5, 7.
- Halpin, H., Hayes, P. J., McCusker, J. P., McGuinness, D. L., & Thompson, H. S. (2010, November). When owl: sameas isn't the same: An analysis of identity in linked data. In *International semantic web conference* (pp. 305-320). Springer, Berlin, Heidelberg.
- Ding, L., Shinavier, J., Finin, T., & McGuinness, D. L. (2010, April). owl: sameAs and Linked Data: An empirical study. In *Proceedings of the Second Web Science Conference*.

Objective

- Both the city and the department of Paris are different in a **legal context**
- But, they are identical in a **geographical context**
- What if a user want to retrieve movie theaters in Paris?
 - Only 3 are connected to the city
 - Only 2 are connected to the department
- Contextual identity** is a possible answer

⇒ **Contextual identity** must allow the **propagation of properties** in certain cases, but how to find those properties?



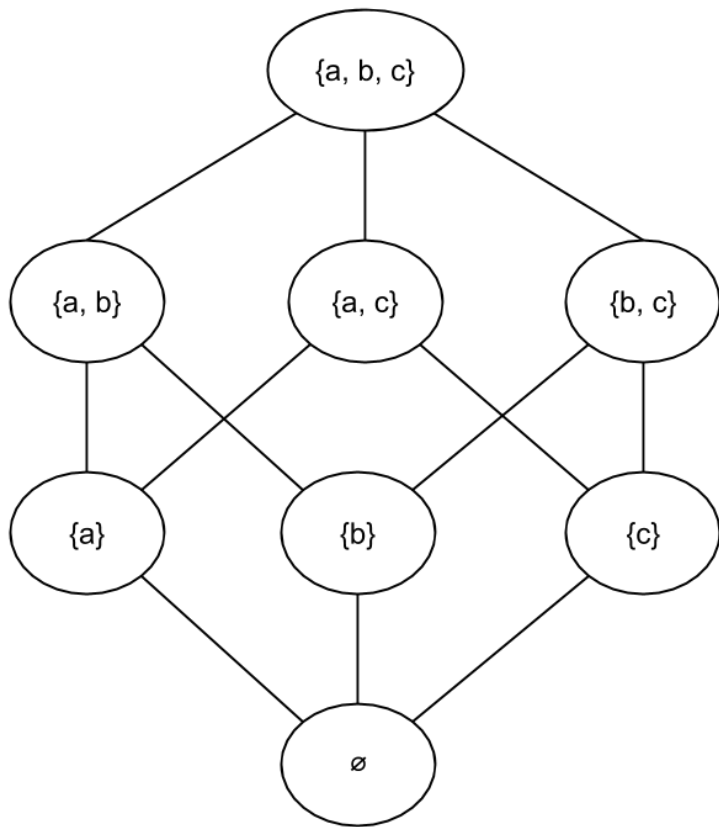
Related work

- Identity context = set of properties (indiscernibility set)
 - entities must share the same value for each property
- Contexts can be represented with a lattice

But there is no clue on what to do with other properties
= **No propagation.**

Suppose **a** is indiscernible w.r.t. several entities, but
what to do with **b** and **c**?

Identity lattice



Related work

Identity context = sub-ontology, which provides an overview of the context.

Context can be automatically computed.

But **propagable properties** don't belong to the framework.

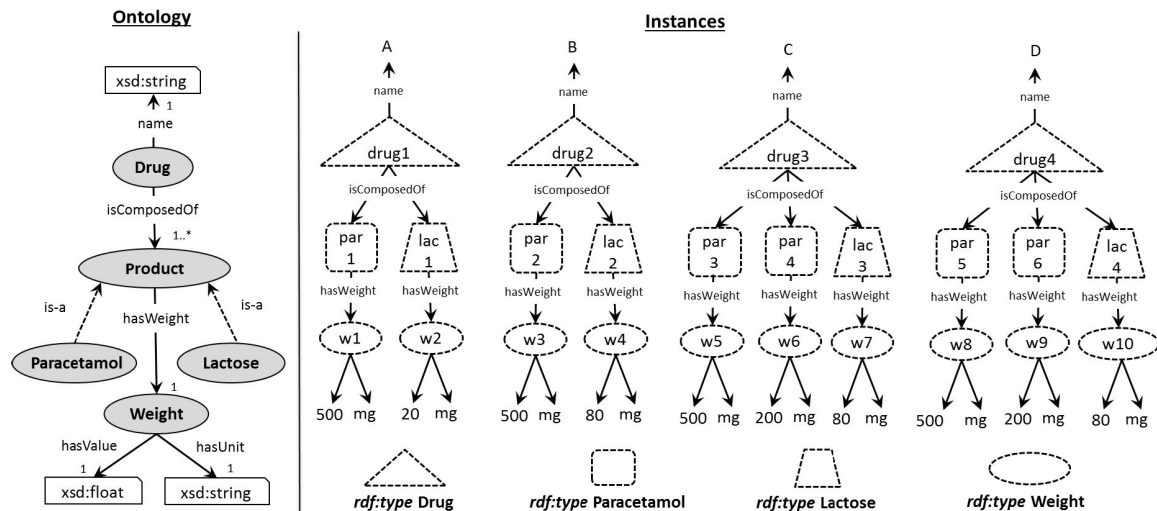


Figure 1: An extract of ontology *O*, four instances *drug1*, *drug2*, *drug3* and *drug4* of the target class *Drug*.

Related work

Identity context = indiscernibility set +
propagation set + alignment procedure.

$$x =_{(\Pi, \Psi, \approx)} y \leftrightarrow \forall (p_1, p_2) \in \Pi^2 \text{ with } p_1 \approx p_2 \\ \text{and } \forall v_1, v_2 \text{ with } v_1 \approx v_2 : \langle x, p_1, v_1 \rangle \leftrightarrow \langle y, p_2, v_2 \rangle$$

But **users must provide everything**.

$$x =_{(\Pi, \Psi, \approx)} y \rightarrow \forall (p_1, p_2) \in \Psi^2 \text{ with } p_1 \approx p_2 \\ \text{and } \forall v_1, v_2 \text{ with } v_1 \approx v_2 : \langle x, p_1, v_1 \rangle \leftrightarrow \langle y, p_2, v_2 \rangle$$

How to find a propagation set of properties?

- Identity context based on Idrissou et al.'s definition
- **Tobler's first law** : *“Everything is related to everything else, but near things are more related than distant things.”*

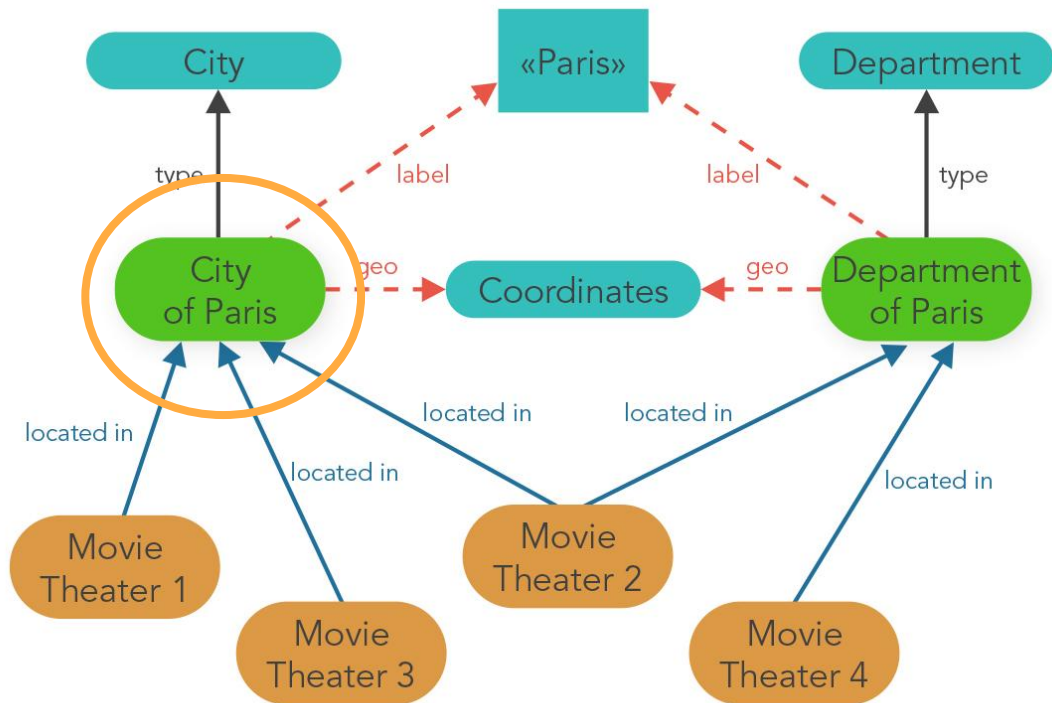
⇒ Propagable properties could be semantically related to indiscernible properties

- Sentences describing properties could be transformed into **numerical vectors**
- Vectors representing **propagable properties must be close to** vectors representing **indiscernible properties**

Property Propagation

Sample knowledge graph about Paris and its movie theaters.

We consider the City of Paris as the **seed** of the identity lattice.

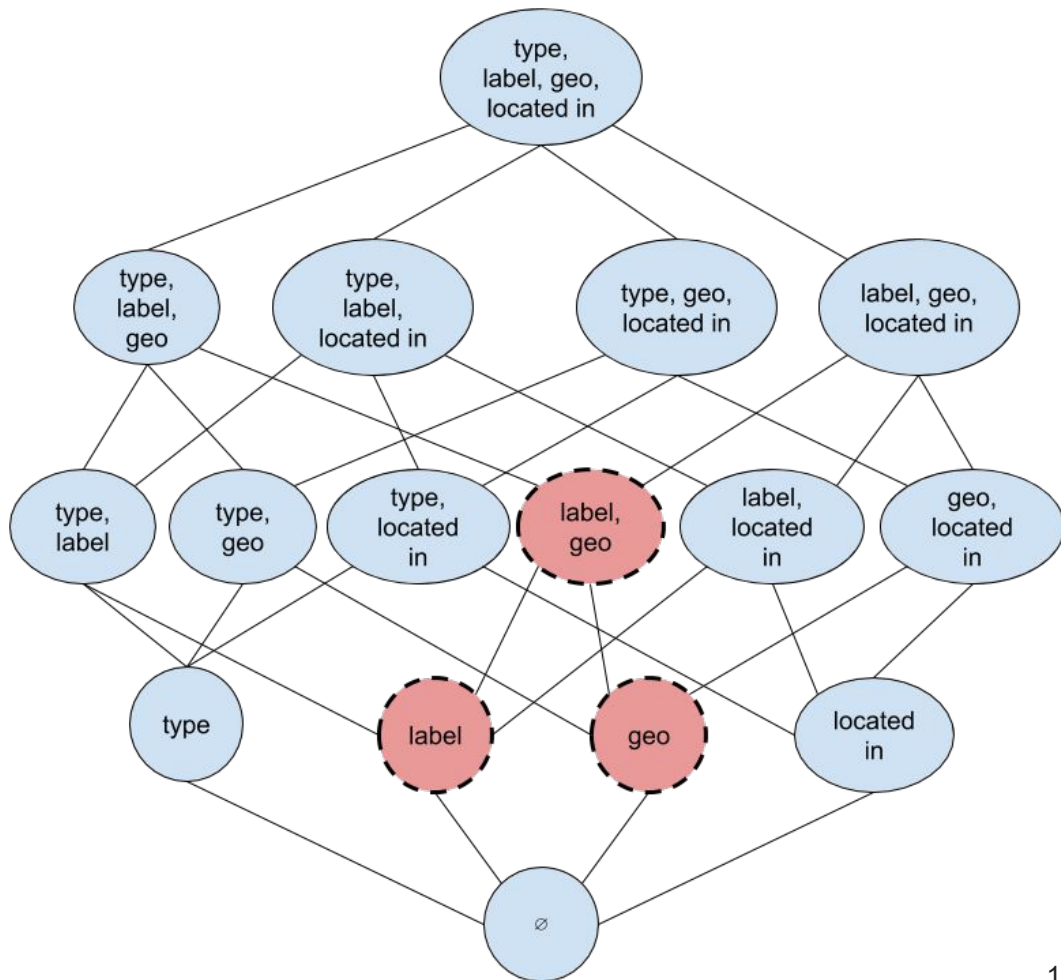


Property Propagation

Simplified identity lattice.

Each node correspond to the an indiscernibility set.

Only red nodes have contextually identical entities.



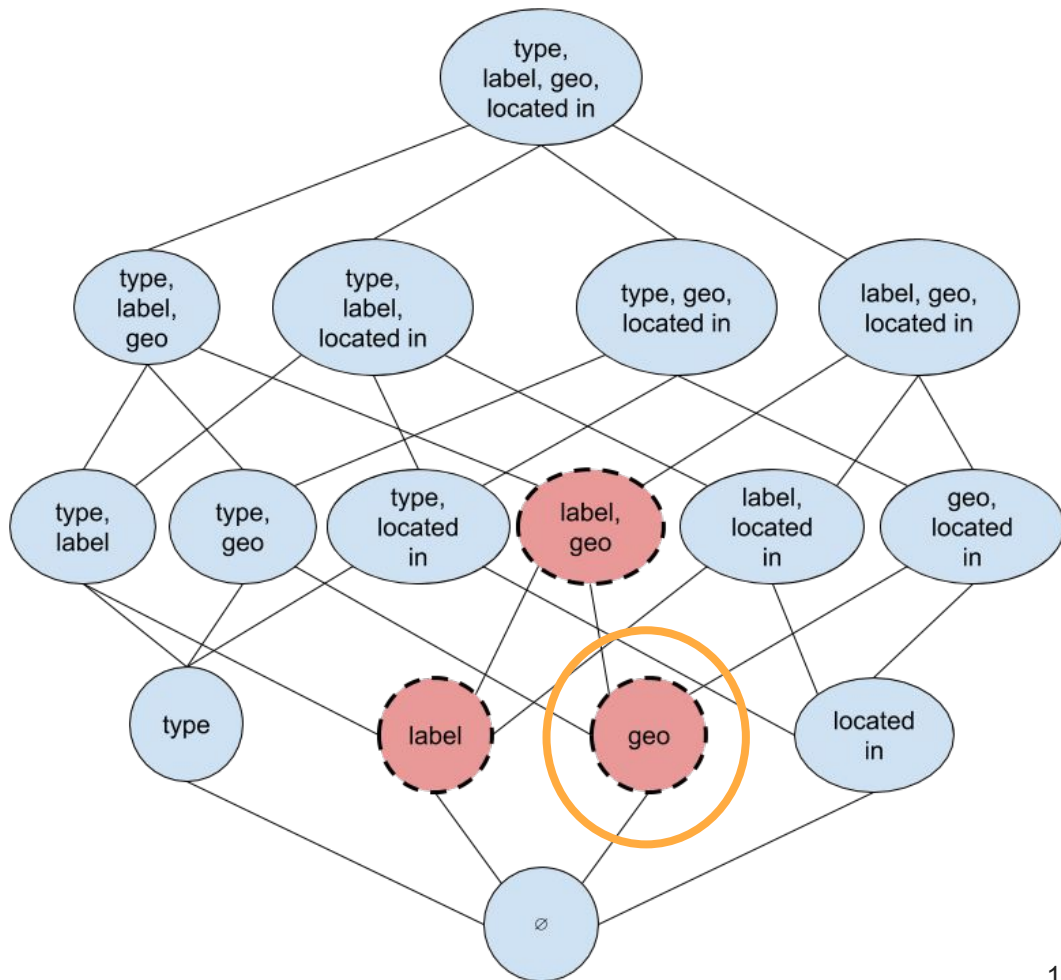
Property Propagation

Candidate properties for propagation =
“type”, “label” and “located in”

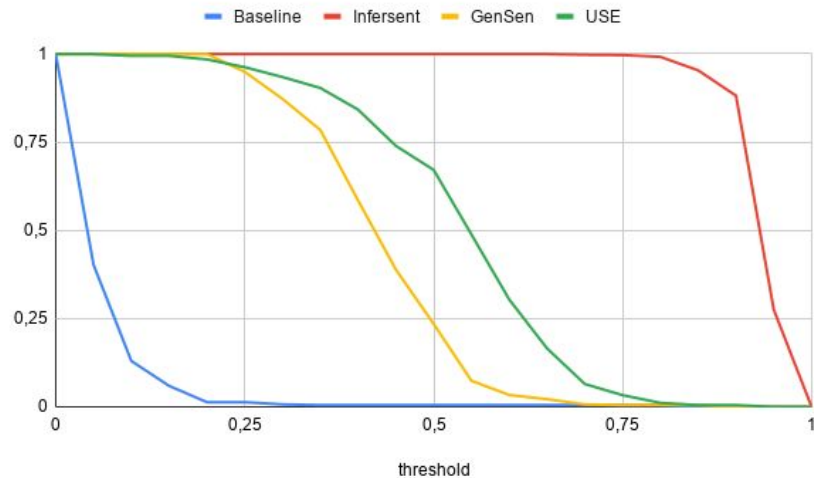
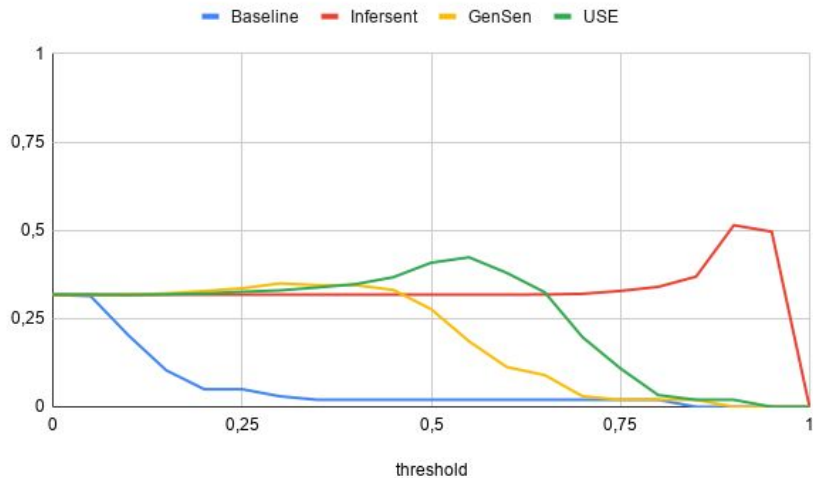
We compute the embeddings of the descriptions of the four properties.

The vector representing “located in” is
close to the vector representing “geo”

⇒ “located in” can be
propagated for the
indiscernibility set {geo}

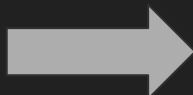


Results



Gold standard = 100 entities (5 classes)

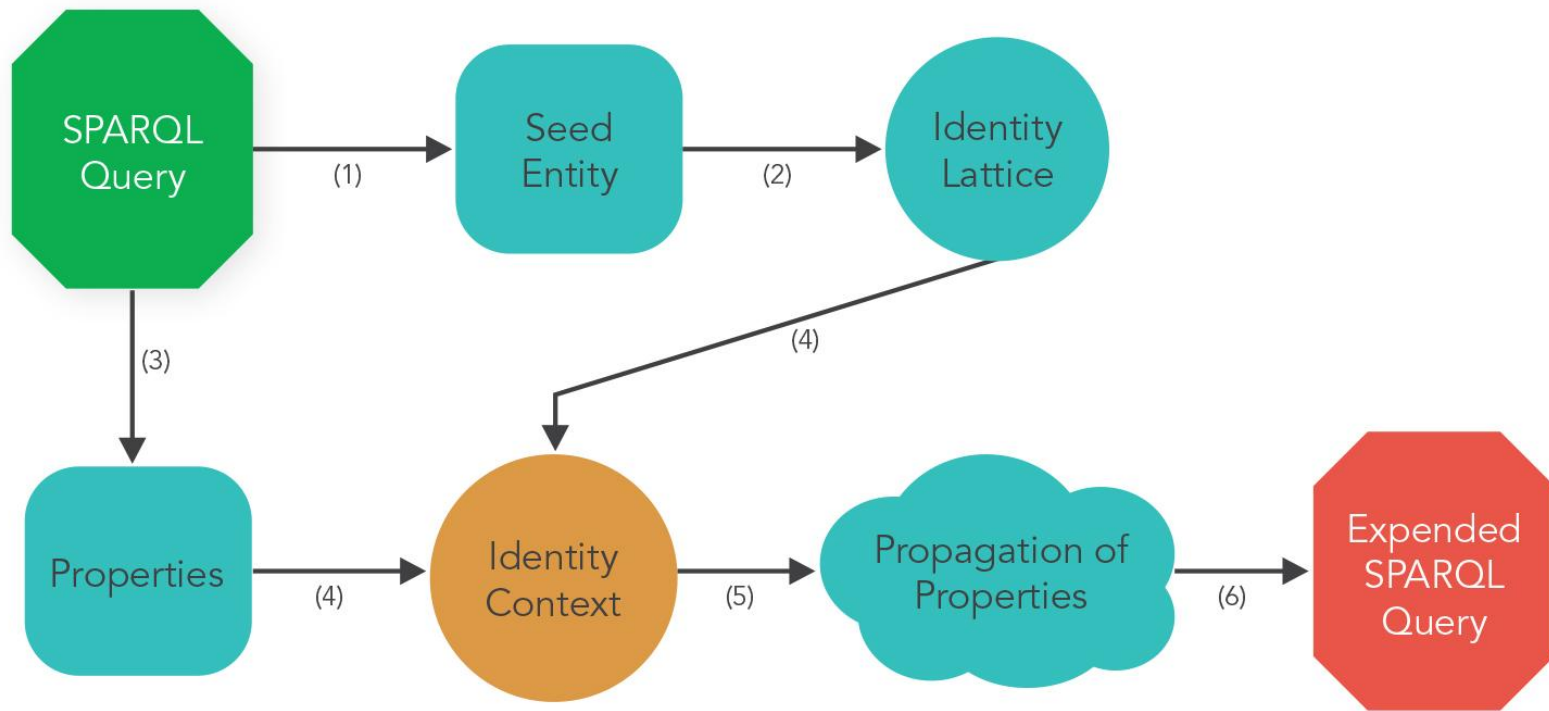
Baseline vs Infersent vs GenSen vs USE
⇒ The winner is Infersent



Textual descriptions are useful to
discover properties that are propagable.

Highly dependent on the encoder.

Framework for Propagation of Properties



Example

Who are the convicted members of Les
Républicains?

```
SELECT DISTINCT ?politician ?crime
```

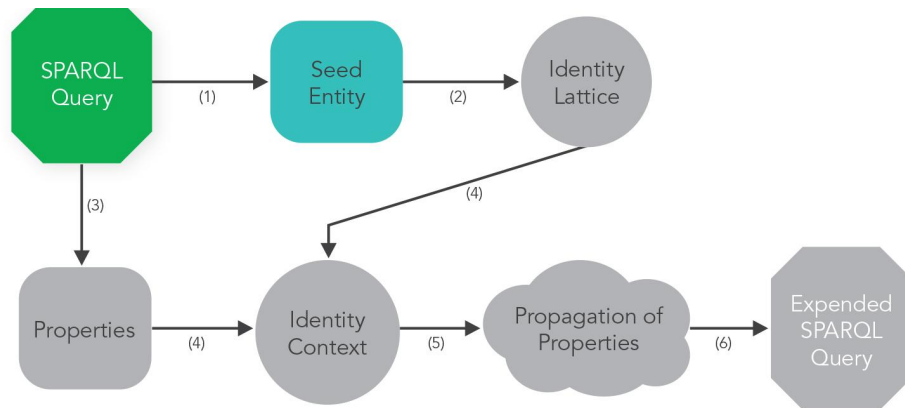
```
WHERE {
```

```
  ?politician :memberOf :TheRepublicans ;
```

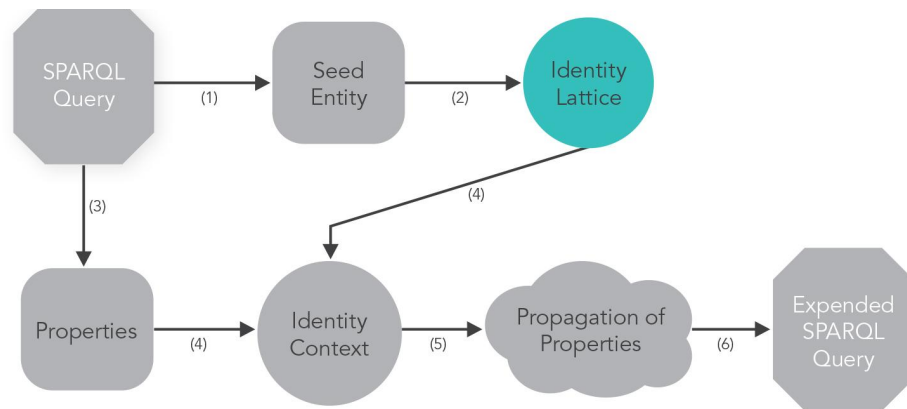
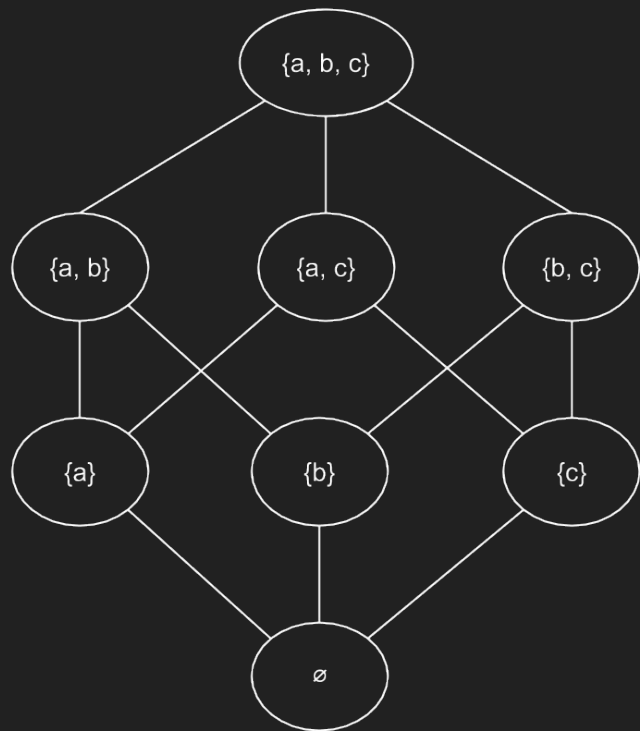
```
    :convictedOf ?crime .
```

```
}
```

# of results w/o context	2
--------------------------	---



Example



Example

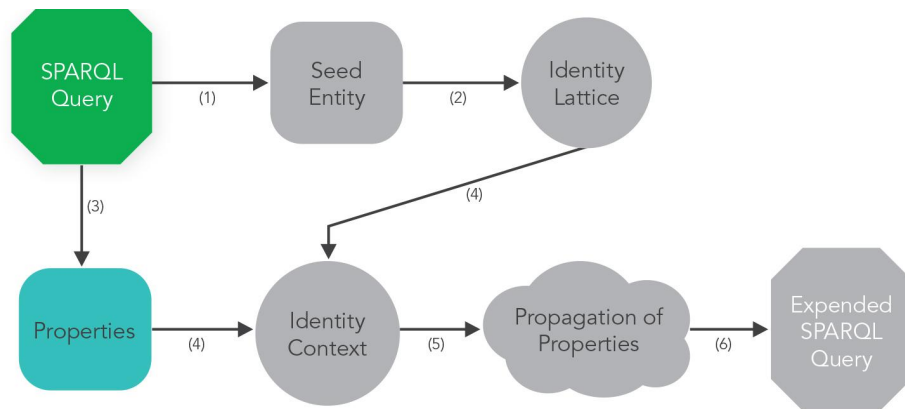
```
SELECT DISTINCT ?politician ?crime
```

```
WHERE {
```

```
  ?politician :memberOf :TheRepublicans ;
```

```
    :convictedOf ?crime .
```

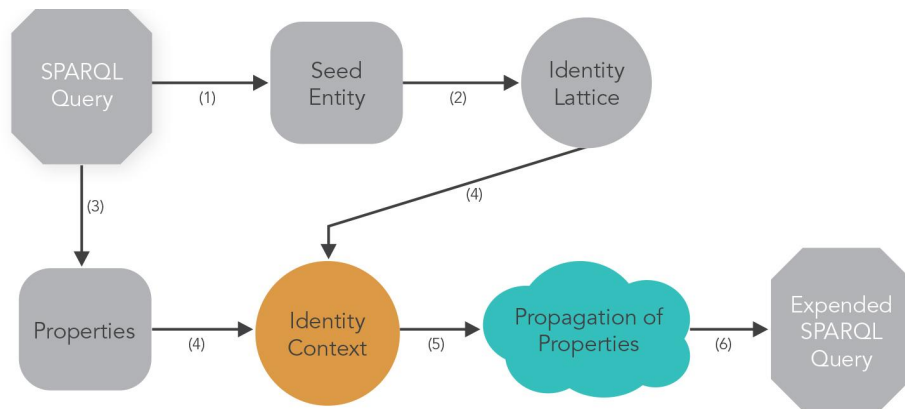
```
}
```



Example

The user must **choose** the most appropriate identity context among those proposed.

Seed	The Republicans
ψ	member of , political party
Π	country, political, ideology
Contextually identical entities	UMP, RPR, UDR, UNR



Example

SELECT DISTINCT ?politician ?crime

WHERE {

VALUES (?party) {

(:TheRepublicans) (:UMP) (:RPR) (:UDR)

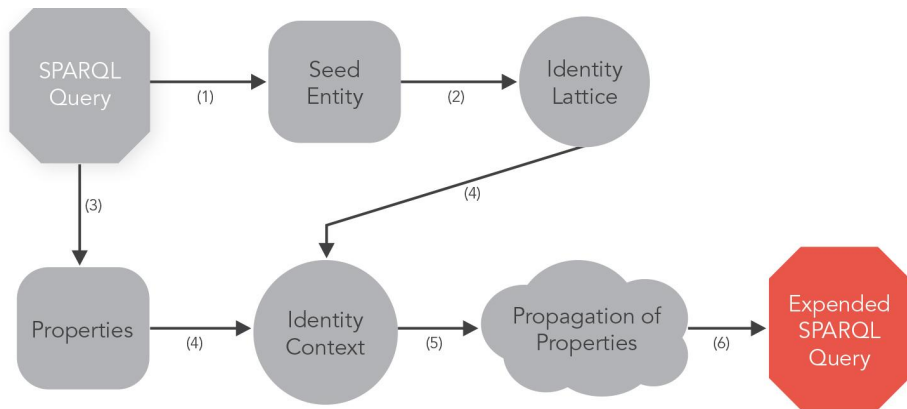
(:UNR)

}

?politician :memberOf ?party ;

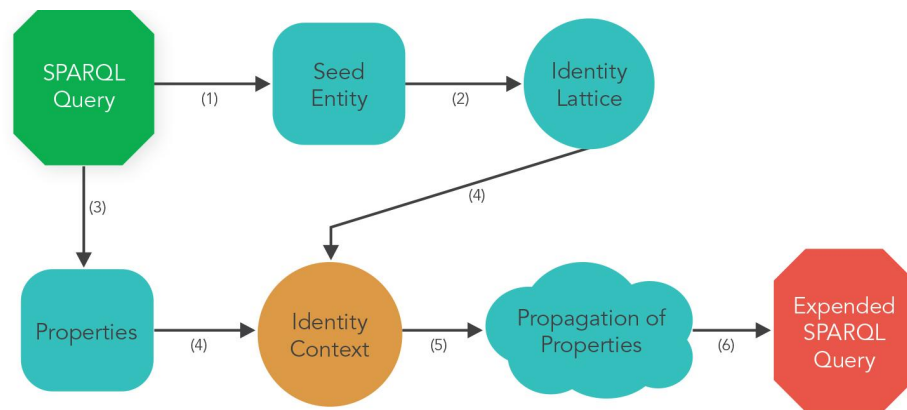
:convictedOf ?crime .

}



Example

Seed	The Republicans
ψ	member of, political party
Π	country, political, ideology
Contextually identical entities	UMP, RPR, UDR, UNR
# of results w/o context	2
# of results w/ context	13



Conclusions

*“While entity disambiguation and resolution is an active research area in the semantic Web, and now in knowledge graphs for several years, it is almost surprising that it continues to be **one of the top challenges in the industry** almost across the board.”*

*“How can **identity** be described in a way that different teams can **agree on it** and know what the other teams are describing?”*

Natasha Noy et al., « Industry-scale knowledge graphs: Lessons and challenges », *Queue* 17, n° 2 (2019): 48–75.

⇒ **Identity is often context-dependant.**

⇒ **Propagation is essential for identity and contextual identity.**

Conclusions

An approach **to compute a set of propagable properties** given a set of indiscernible properties:

- Based on Tobler's first law and sentence embedding
- A gold standard
- A full framework to increase completeness of SPARQL queries

Future Work

Propagation of properties:

- Not rely only on description of properties
- Try to use values of properties or semantic features of the property
- Challenge our work with a combination of distinct KGs
- Consider cases where the seed is the value of a triple
- Finalize the full prototype

Related work

Paper	Identity Context	Advantages	Drawbacks
Beek et al.	Set of properties	Contexts can be represented with a lattice	No propagation
Raad et al.	Sub-ontology	Clear delimitation of the context, which can be calculated automatically	Propagation is not part of the framework
Idrissou et al.	Indiscernibility set + propagation set + alignment procedure	Propagation is part of the framework	User must provide everything