



# Propagation de propriétés entre entités contextuellement identiques

Pierre-Henri Paris

Télécom Paris

Atelier DECADE, le 28 juin 2022

<https://phparis.net>



# ■ Un peu d'histoire et de contexte





*“Le président est arrivé !”*



*“Le président est arrivé !”*

**De quel président s’agit-il ?**



*“Le président est arrivé !”*

De quel président s’agit-il ?



Emmanuel Macron,  
France, 2022.

*“Le président est arrivé !”*

De quel président s’agit-il ?



Emmanuel Macron,  
France, 2022.



Donald Trump,  
USA, 2016-2020.

*“Le président est arrivé !”*

De quel président s’agit-il ?



Emmanuel Macron,  
France, 2022.



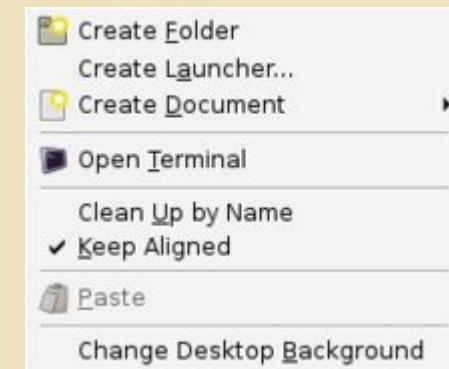
Donald Trump,  
USA, 2016-2020.



Maxime Lefrancois,  
PFIA, 2022.



- Linguistique et communication
- Psychologie et sociologie
- Histoire et géographie
- Informatique
- ...









*Ceci n'est pas un menu contextuel.*


- Logiques classiques  $\rightarrow$  domaine, interprétation et modèle

$D = \{ \text{Emmanuel Macron}, \text{Joe Biden}, \text{Donald Trump} \} : I_1(\text{President}) = \{ \text{Emmanuel Macron}, \text{Donald Trump} \}$  ou  $I_2(\text{President}) = \{ \text{Joe Biden} \}$







- Logiques classiques  $\rightarrow$  domaine, interprétation et modèle

$D = \{ \text{ , \text{ , \text{ } : I_1(\text{President}) = \{ \text{ , \text{ } \text{ ou } I_2(\text{President}) = \{ \text{ } \}$


- Logiques contextuelles  $\rightarrow$  John McCarthy

$\forall c, \text{Country}(c, \text{France}) \wedge \text{Event}(c, \text{PFIA}, 2022) \rightarrow$   
 $\text{ist}(c, \forall h, \text{Human}(h) \wedge \text{President}(h) \rightarrow \text{Name}(h, \text{}))$

- Logiques classiques  $\rightarrow$  domaine, interprétation et modèle

$D = \{ \text{ , \text{ , \text{ } : I_1(\text{President}) = \{ \text{ , \text{ } \text{ ou } I_2(\text{President}) = \{ \text{ } \}$

- Logiques contextuelles  $\rightarrow$  John McCarthy

$\forall c, \text{Country}(c, \text{France}) \wedge \text{Event}(c, \text{PFIA}, 2022) \rightarrow$   
 $\text{ist}(c, \forall h, \text{Human}(h) \wedge \text{President}(h) \rightarrow \text{Name}(h, \text{}))$

- Logiques modales

$\Box((\neg \text{french} \vee \text{convicted}) \rightarrow \bigcirc \neg \text{being\_elected\_president})$

## Le paradoxe de Thésée



Le paradoxe  
de Thésée



Principe  
d'identité



Le principe d'identité des indiscernables

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

Le principe d'indiscernabilité des identiques

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

La sémantique de *owl:sameAs* est fondée sur ces principes

Le principe d'identité des indiscernables

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

Le principe d'indiscernabilité des identiques

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

La sémantique de *owl:sameAs* est fondée sur ces principes

⇒ Les couples propriété-valeur peuvent être **propagés** d'une entité à une autre entité identique



Le principe d'identité des indiscernables

$$\forall x, \forall y \left( \forall p, \forall o, (\langle x, p, o \rangle \wedge \langle y, p, o \rangle) \Rightarrow x = y \right)$$

Le principe d'indiscernabilité des identiques

$$\forall x, \forall y \left( x = y \Rightarrow \forall p, \forall o, (\langle x, p, o \rangle \Rightarrow \langle y, p, o \rangle) \right)$$

La sémantique de *owl:sameAs* est fondée sur ces principes

⇒ Les couples propriété-valeur peuvent être **propagés** d'une entité à une autre entité identique

**+** **Accroît** l'exhaustivité des entités et la découverte d'informations

Le principe d'identité des indiscernables

$$\forall x, \forall y \left( \forall p, \forall o, (\langle x, p, o \rangle \wedge \langle y, p, o \rangle) \Rightarrow x = y \right)$$

Le principe d'indiscernabilité des identiques

$$\forall x, \forall y \left( x = y \Rightarrow \forall p, \forall o, (\langle x, p, o \rangle \Rightarrow \langle y, p, o \rangle) \right)$$

La sémantique de *owl:sameAs* est fondée sur ces principes

⇒ Les couples propriété-valeur peuvent être **propagés** d'une entité à une autre entité identique

- + **Accroît** l'exhaustivité des entités et la découverte d'informations
- Propagation de **mauvaises informations** lorsque les liens sont faux

Le principe d'identité

$\forall x$



cerna



$\langle o, (\langle x, p, o \rangle \wedge \langle y, p, o \rangle) \Rightarrow \langle x, y, o \rangle$



Le principe d'irréflexivité

$\forall x$

es identiques



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

La sémantique de *owl:sameAs* est fondée sur ces principes

⇒ Les couples propriété-valeur peuvent être **propagés** d'une entité à une autre entité identique

- + **Accroît** l'exhaustivité des entités et la découverte d'informations
- Propagation de **mauvaises informations** lorsque les liens sont faux

# | Le principe d'identité de Leibniz

Le principe d'identité

$\forall x$



concerna

$\langle o, (\langle x, p, o \rangle \wedge \langle y, p, o \rangle) \Rightarrow \langle x, p, o \rangle \wedge \langle y, p, o \rangle$



Le principe d'irréflexivité

$\forall x$

es identiques

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

La sémantique de *owl:sameAs* est fondée sur ces principes

⇒ Les couples propriété-valeur peuvent être **propagés** d'une entité à une autre entité identique

**+** Acc



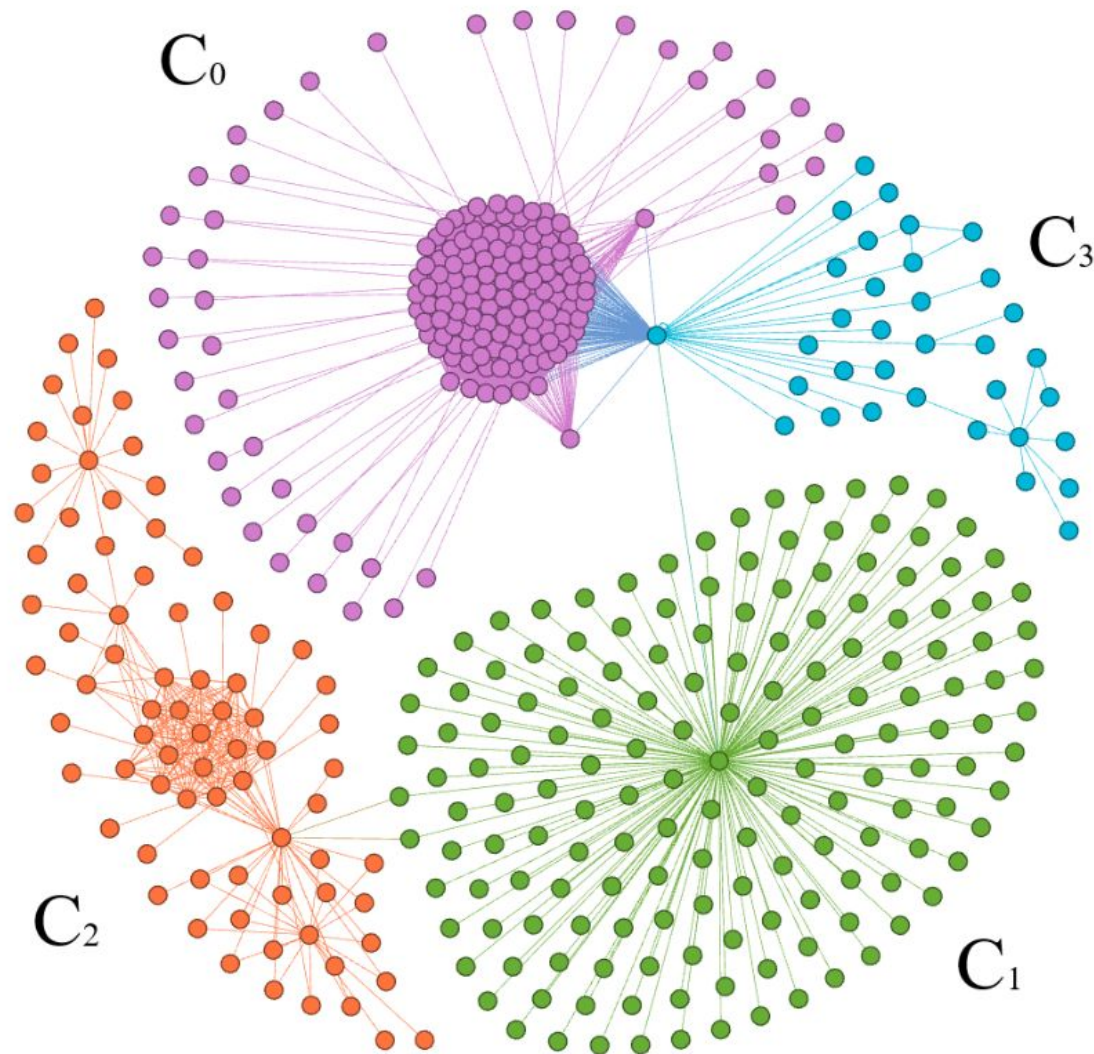
entités la découvert

**-** Prop

es introductions lorsqu



faux

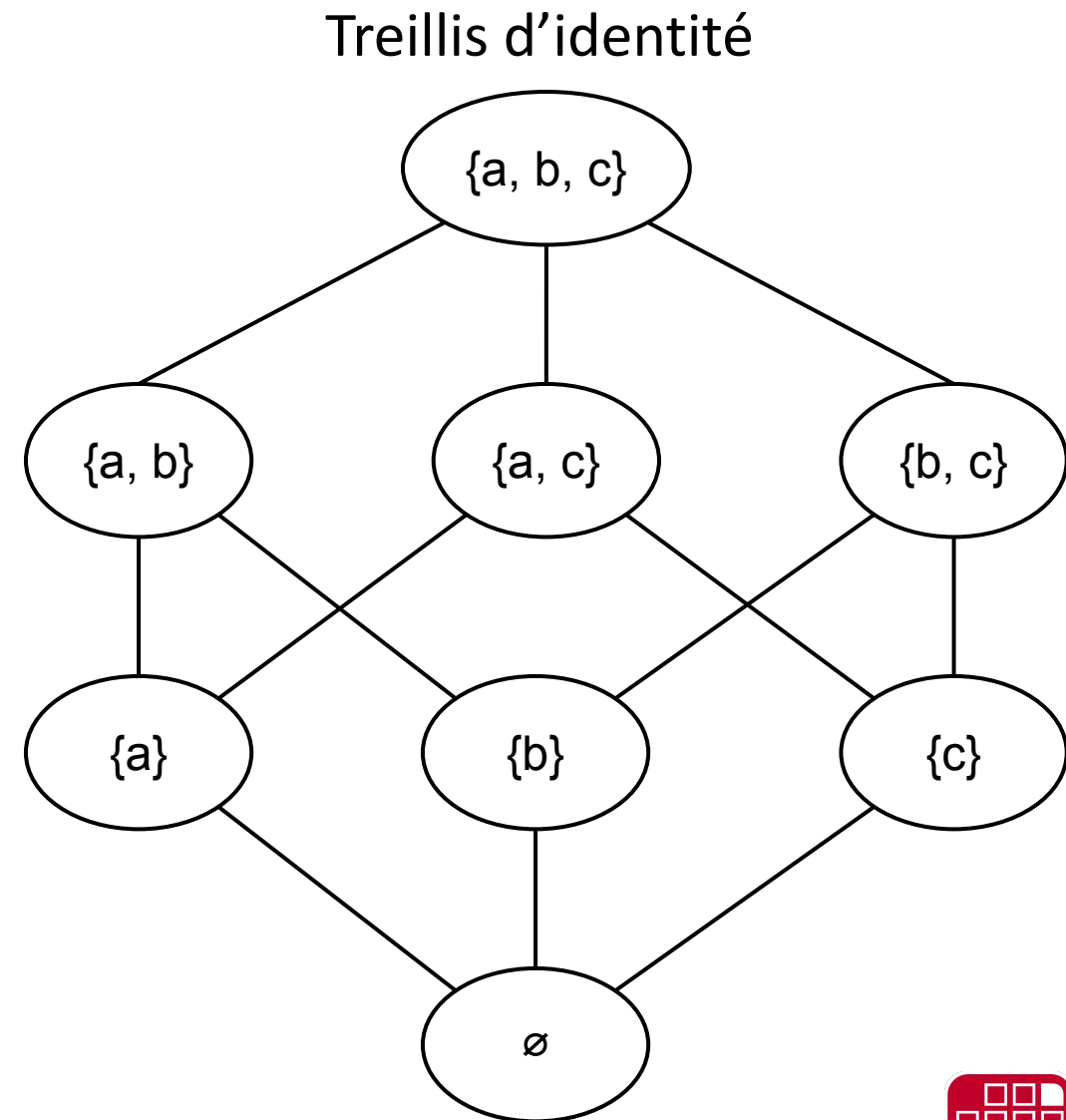


Les communautés détectées à partir de l'ensemble d'égalité contenant le terme `dbp:Barack_Obama` en utilisant l'algorithme de Louvain.  
Les 4 communautés détectées sont distinguées par la couleur de leurs noeuds.

**I Comment  
obtenir une notion d'identité  
moins rigide ?**



- **Contexte d'identité** = ensemble de propriétés (ensemble d'indiscernabilité)
  - Les entités doivent **partager la même valeur pour chaque propriété**
- Les contextes peuvent être représentés par un **treillis**.

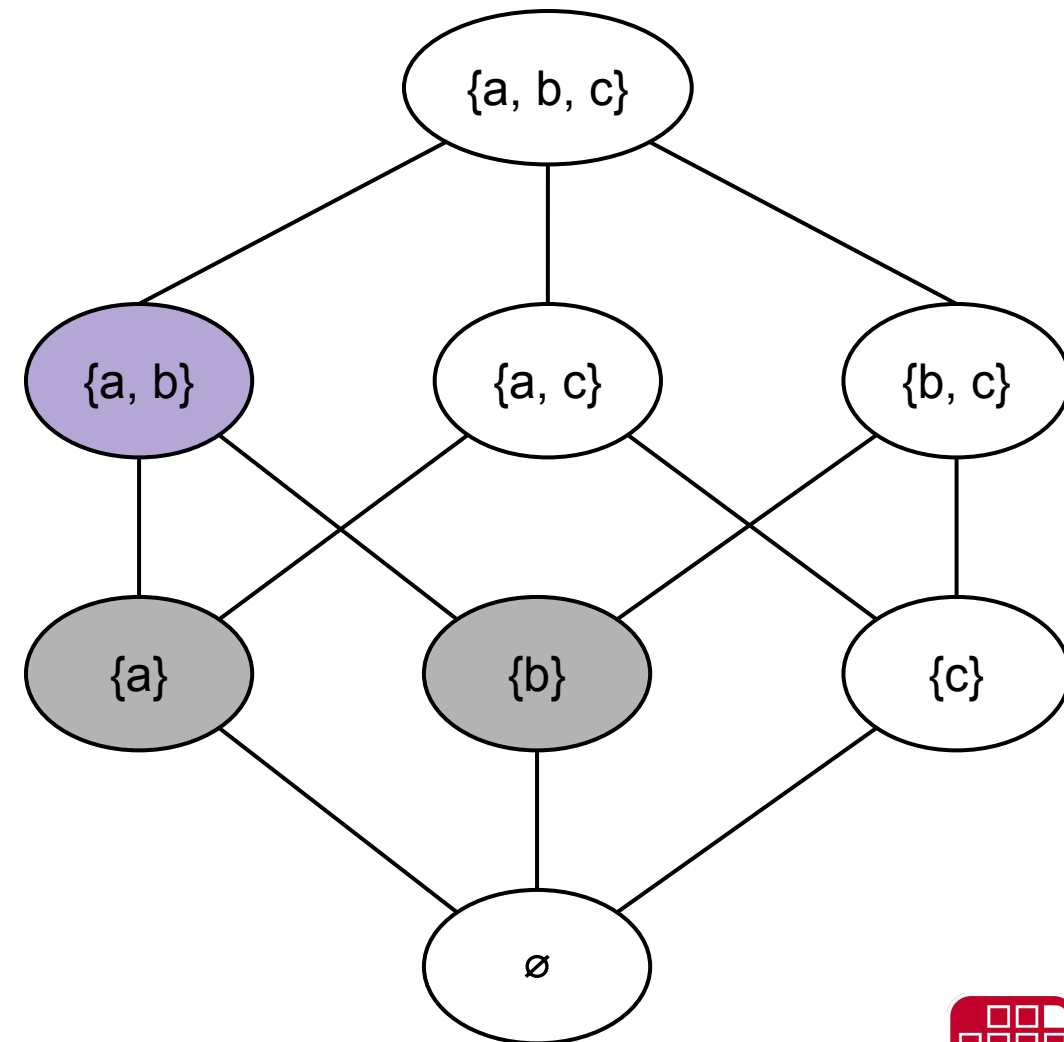


- **Contexte d'identité** = ensemble de propriétés (ensemble d'indiscernabilité)
  - Les entités doivent partager la même valeur pour chaque propriété
- Les contextes peuvent être représentés par un **treillis**.

$\langle x, a, o1 \rangle \wedge \langle \text{graine}, a, o1 \rangle$

$\langle x, b, o2 \rangle \wedge \langle \text{graine}, b, o2 \rangle$

Treillis d'identité





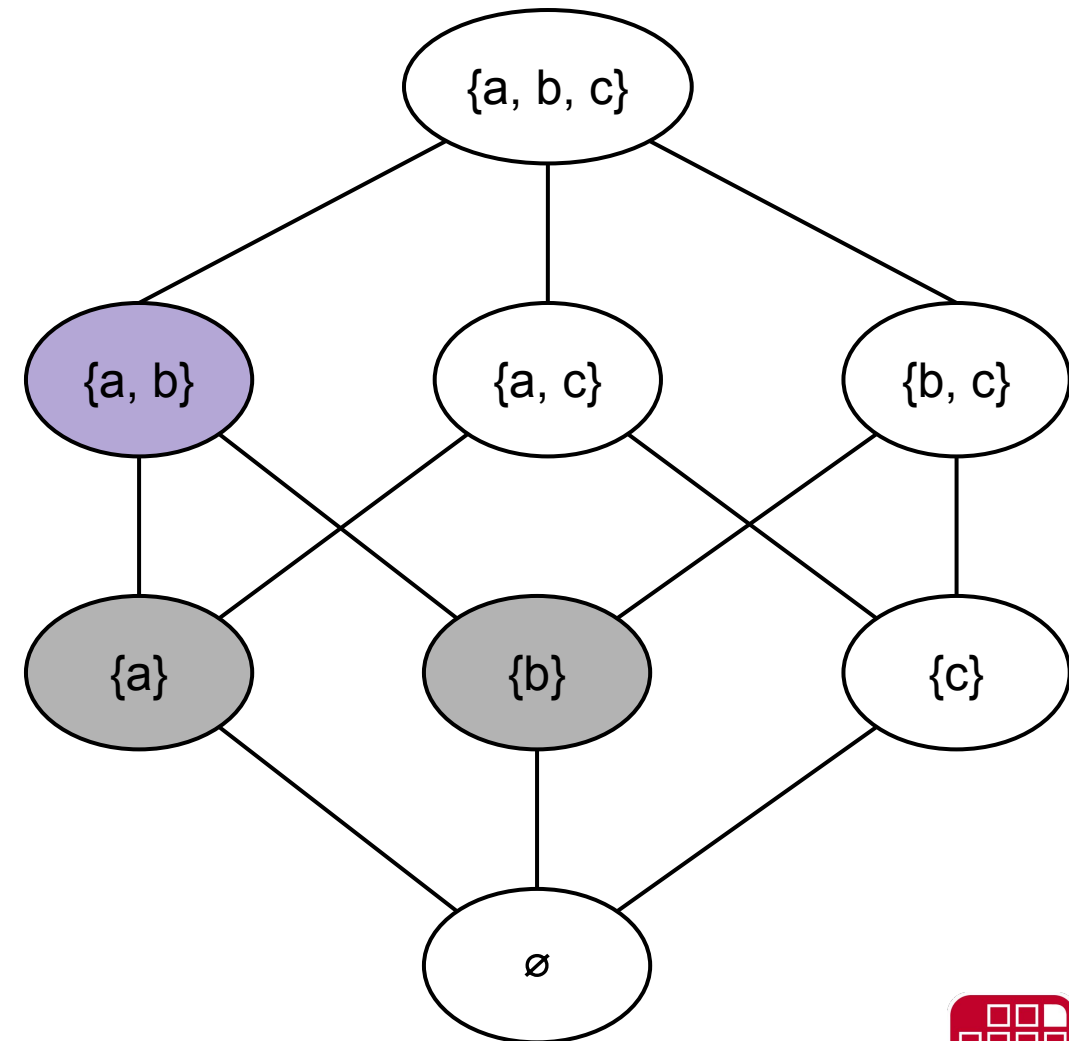
- **Contexte d'identité** = ensemble de propriétés (ensemble d'indiscernabilité)
  - Les entités doivent partager la même valeur pour chaque propriété
- Les contextes peuvent être représentés par un **treillis**.

$\langle x, a, o1 \rangle \wedge \langle \text{graine}, a, o1 \rangle$

$\langle x, b, o2 \rangle \wedge \langle \text{graine}, b, o2 \rangle$

Si  $\langle x, e, o3 \rangle$ , a-t-on  $\langle \text{graine}, e, o3 \rangle$  ?

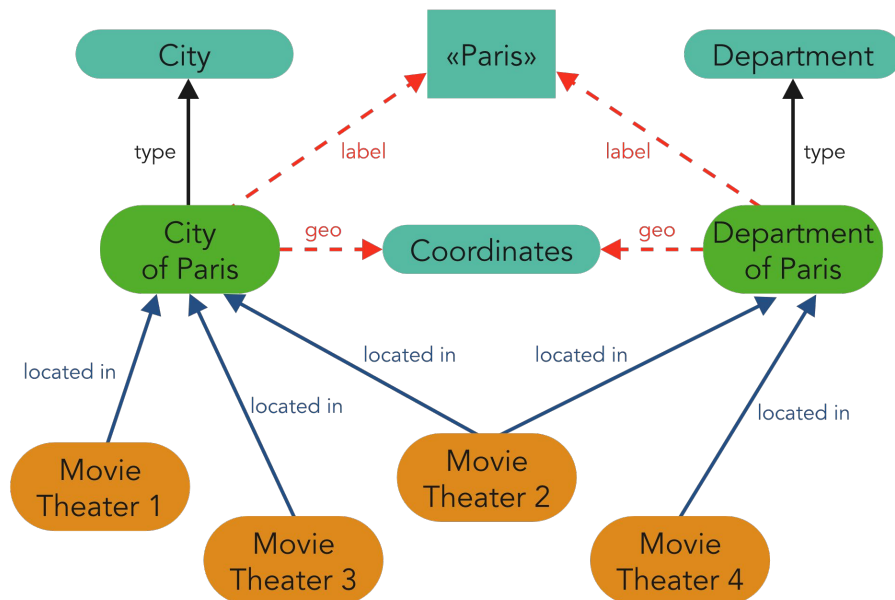
Treillis d'identité



# **I Comment propager des informations avec la notion d'identité contextuelle: L'approche**

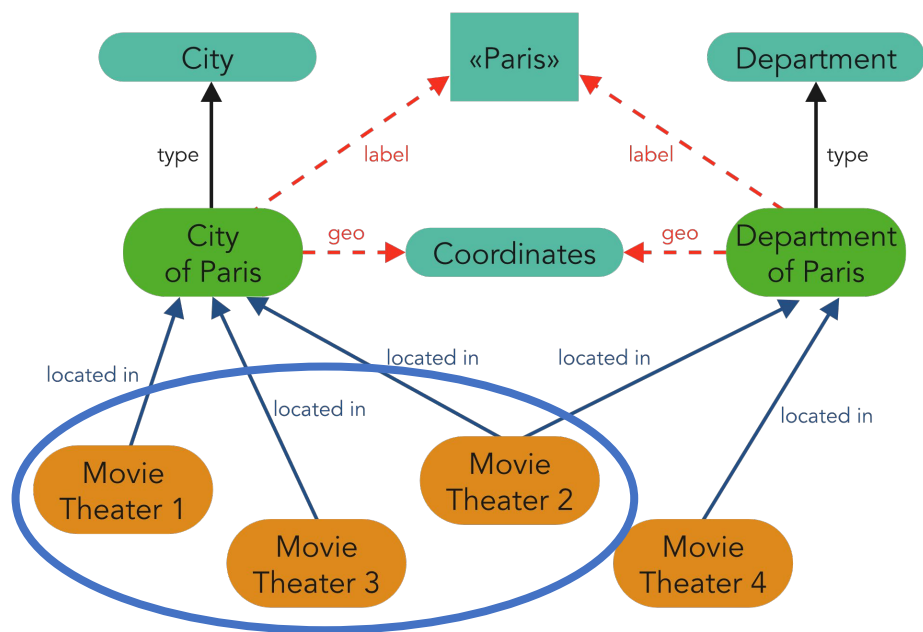


## Identité contextuelle [IC'20, ISWC'20]

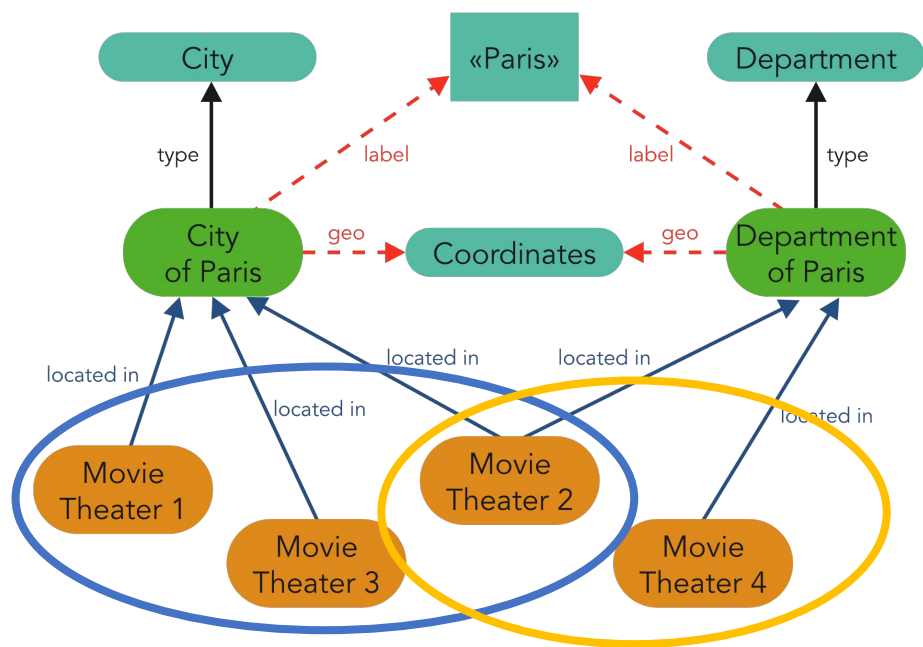


- Calculer semi-automatiquement les propriétés propageables
- Fondée sur les plongements de phrases des descriptions longues des propriétés

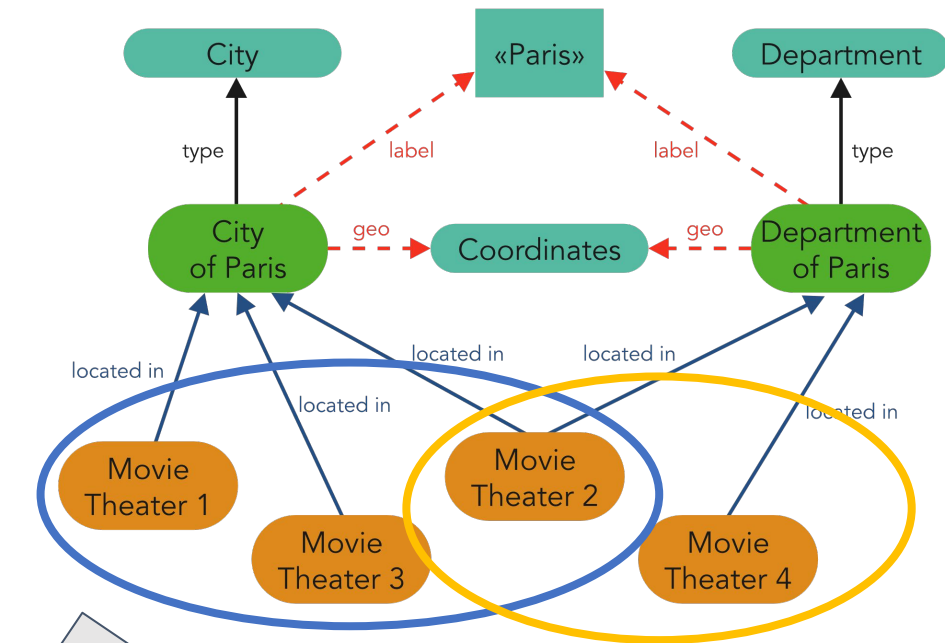
## Identité contextuelle [IC'20, ISWC'20]



## Identité contextuelle [IC'20, ISWC'20]



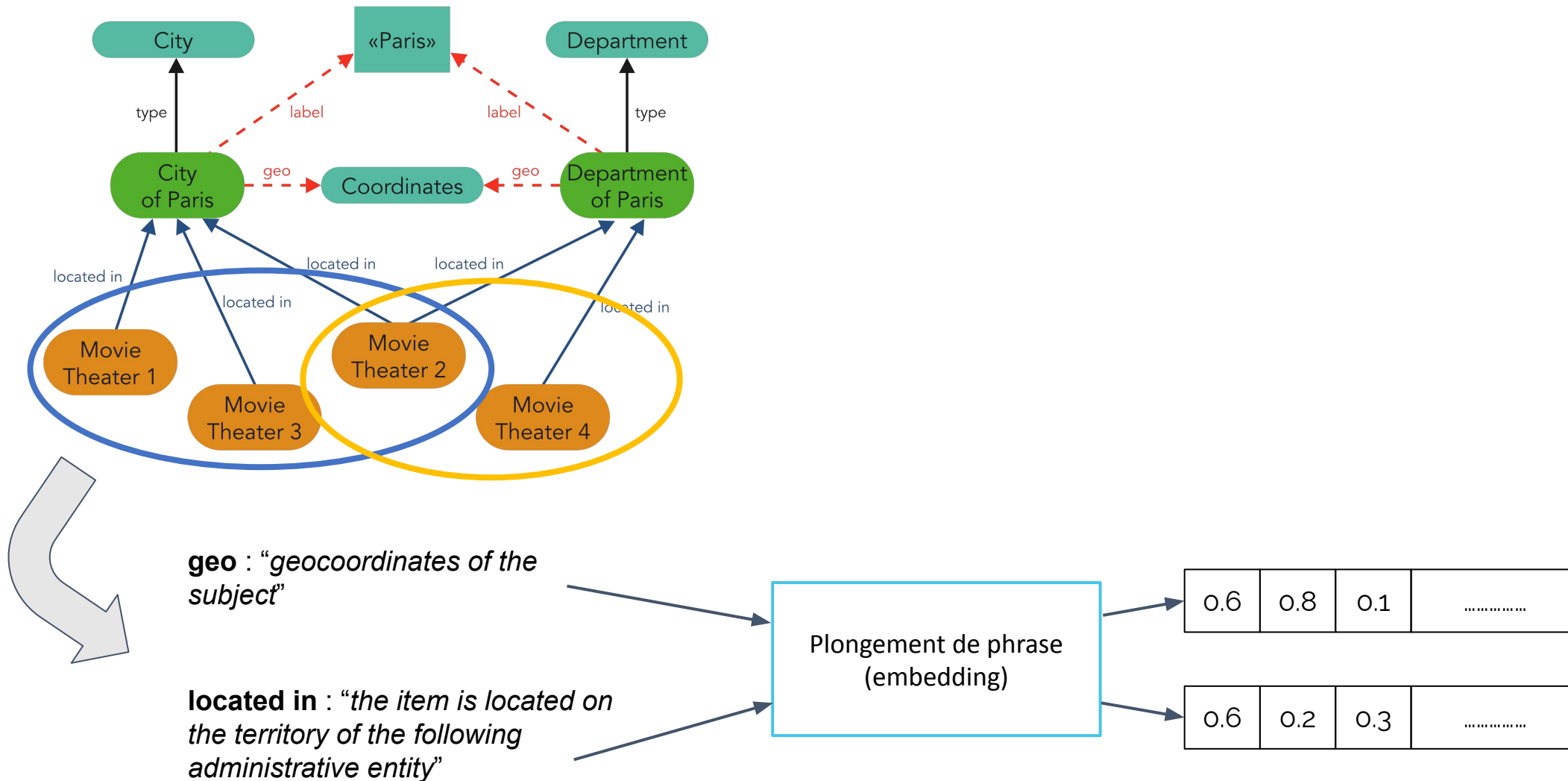
## Identité contextuelle [IC'20, ISWC'20]



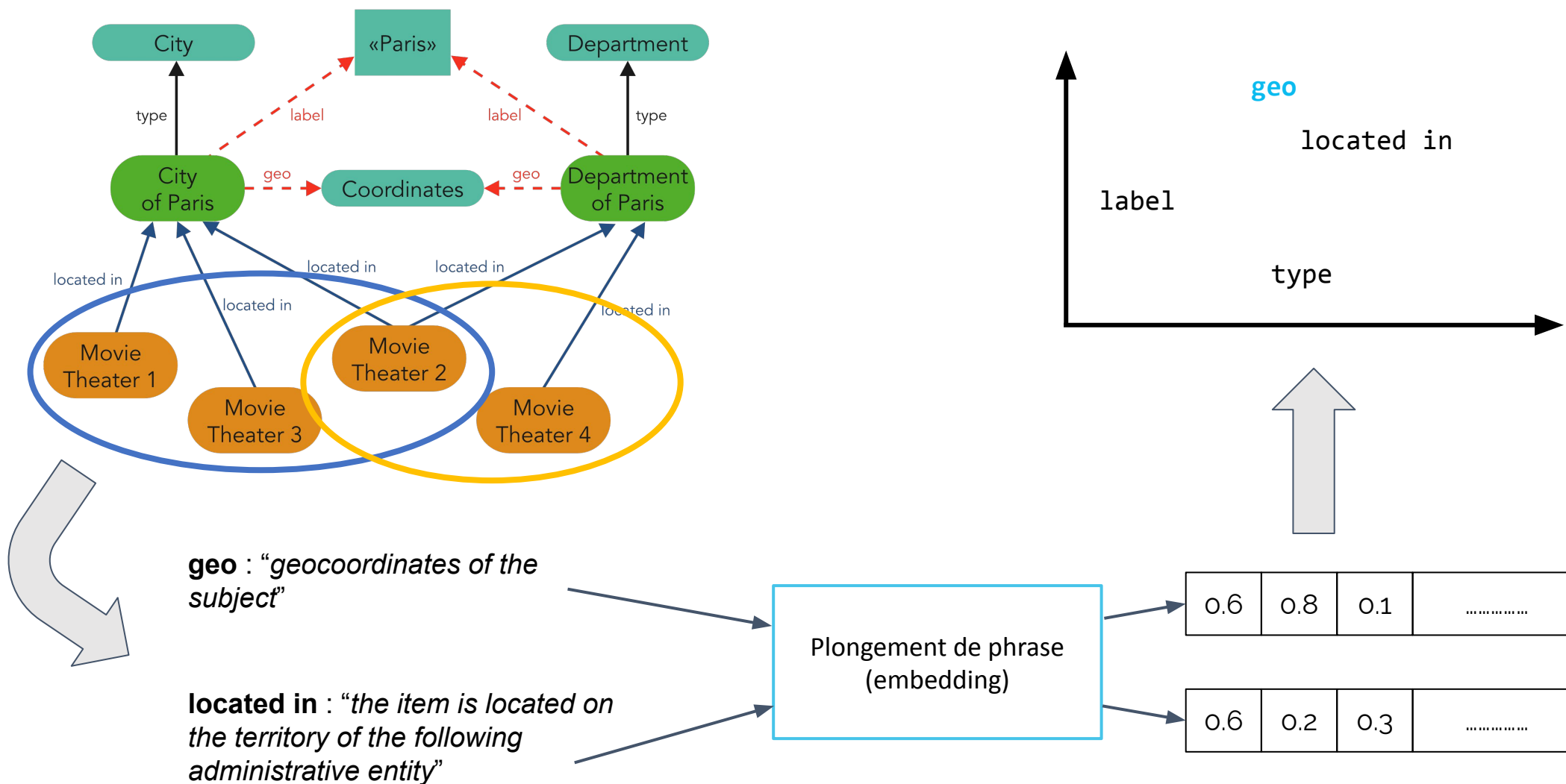
**geo** : “geocoordinates of the subject”

**located in** : “the item is located on the territory of the following administrative entity”

## Identité contextuelle [IC'20, ISWC'20]

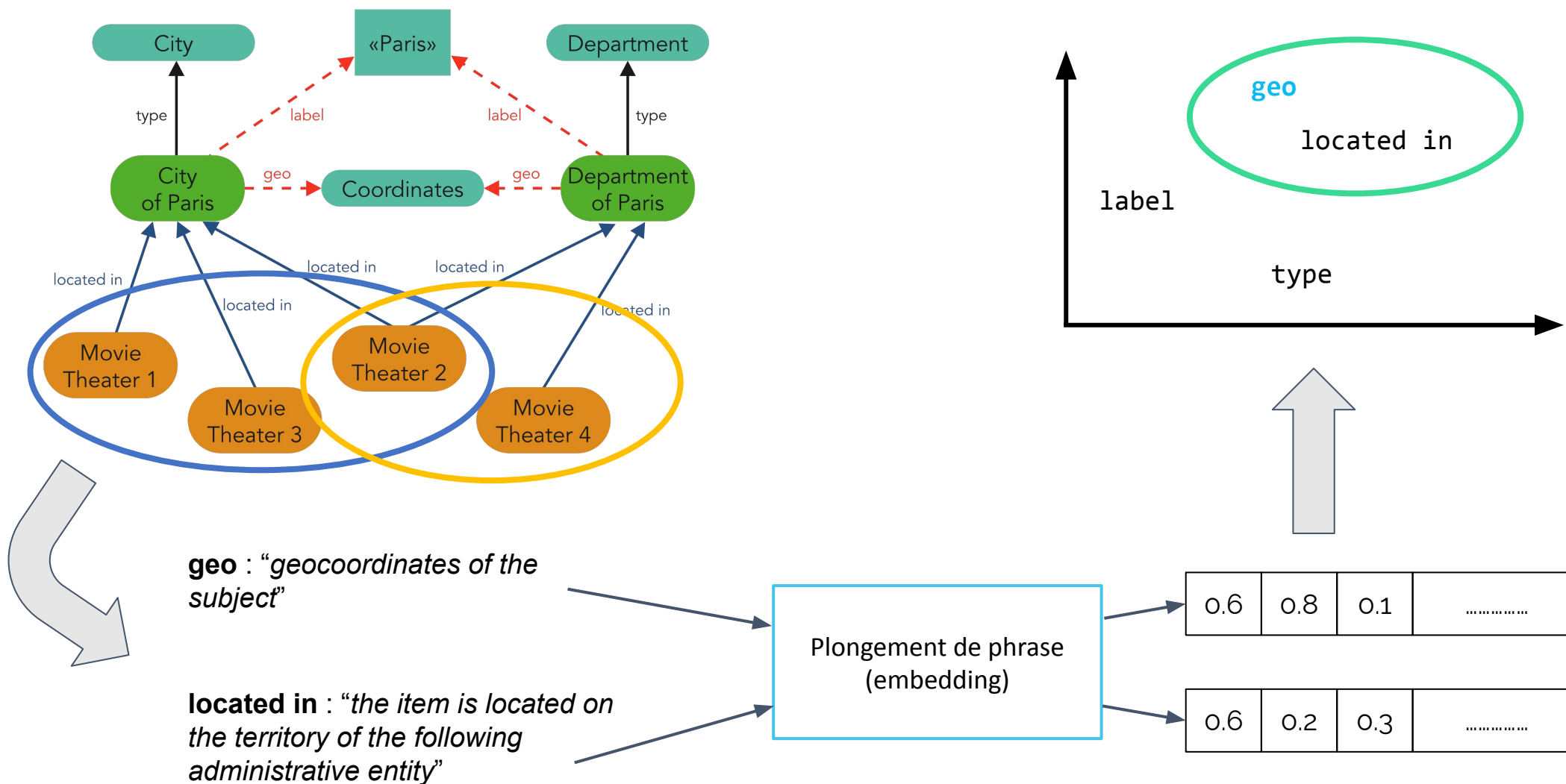


## Identité contextuelle [IC'20, ISWC'20]

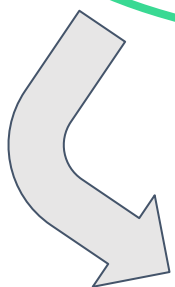
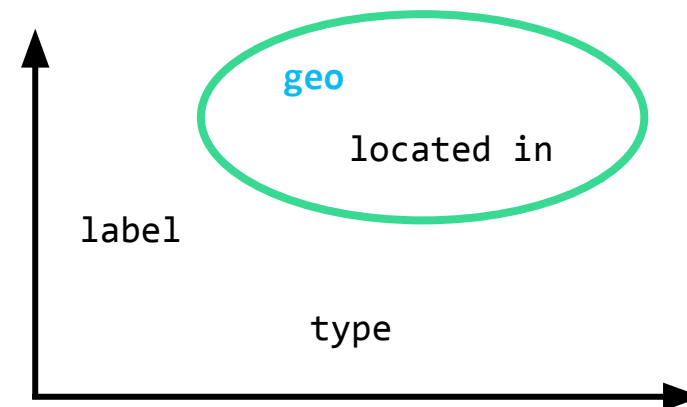
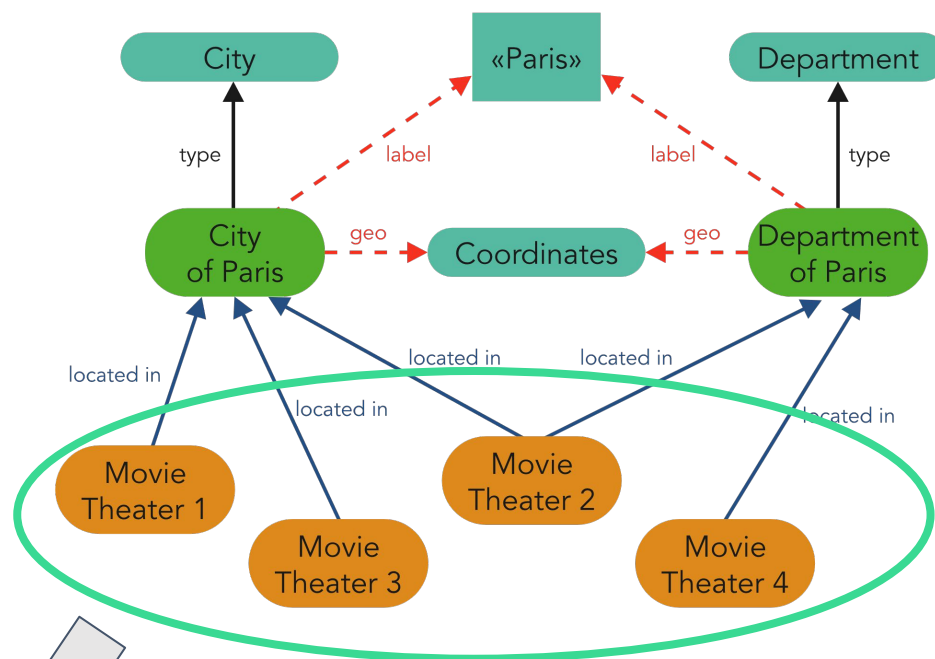




## Identité contextuelle [IC'20, ISWC'20]

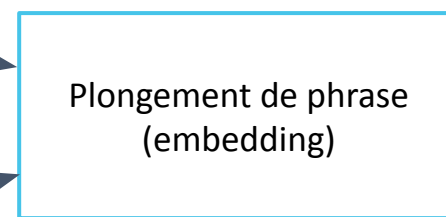


## Identité contextuelle [IC'20, ISWC'20]



**geo** : "geocoordinates of the subject"

**located in** : "the item is located on the territory of the following administrative entity"

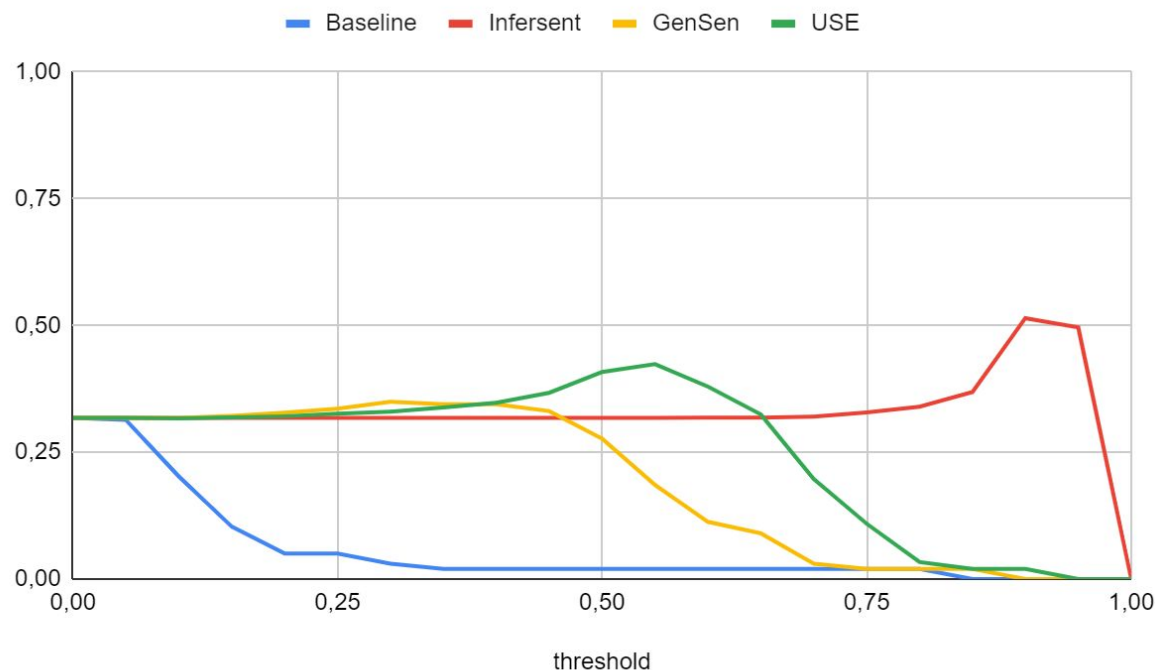


0.6	0.8	0.1	.....
-----	-----	-----	-------

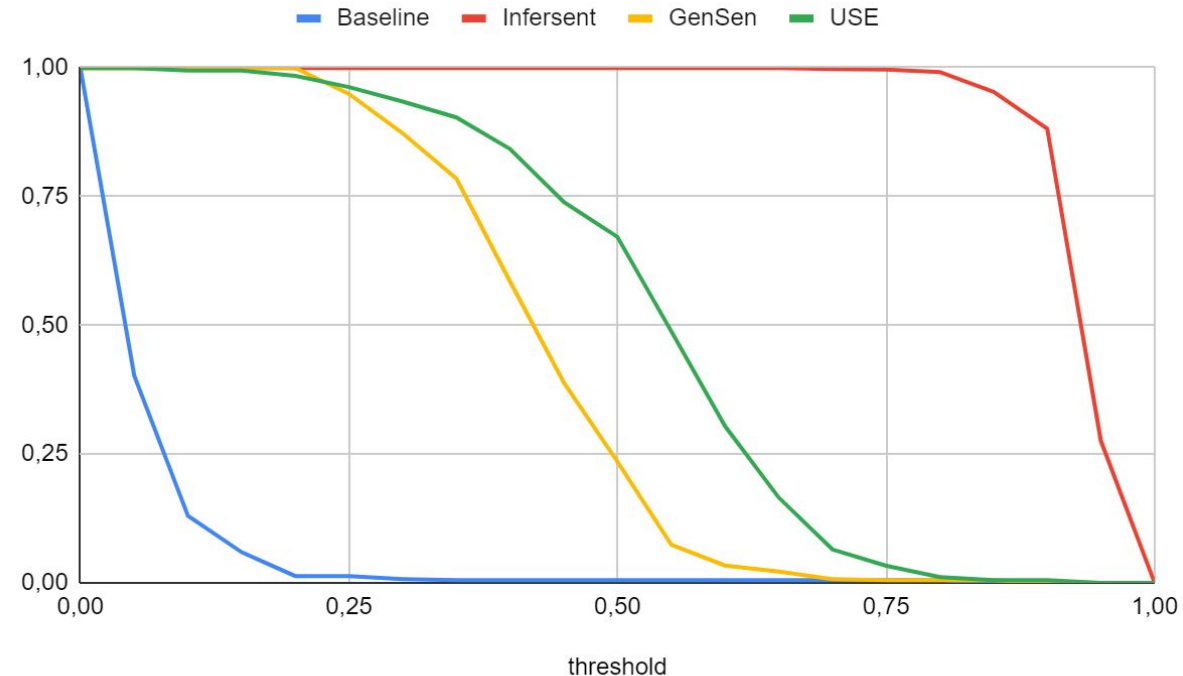
0.6	0.2	0.3	.....
-----	-----	-----	-------



## Précision



## Rappel



Gold standard = 100 entités (5 classes)

Baseline vs Infersent vs GenSen vs USE  
⇒ Le gagnant est Infersent



Les descriptions textuelles sont utiles pour découvrir les propriétés qui sont propageables.

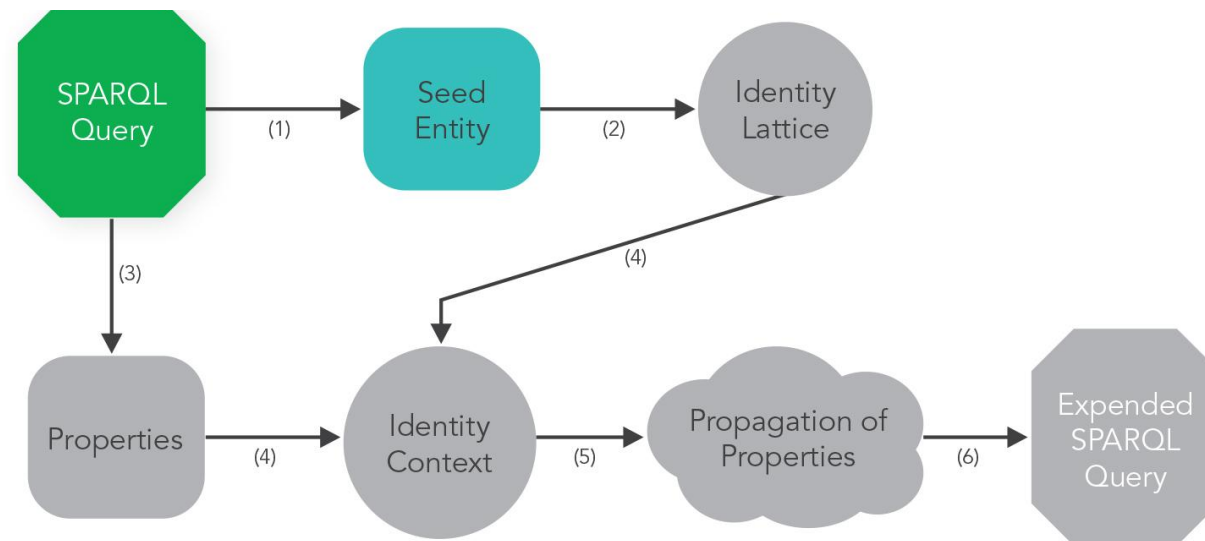
Très dépendant de l'encodeur.

# ■ **Un exemple**

**(attention, certaines images peuvent choquer)**

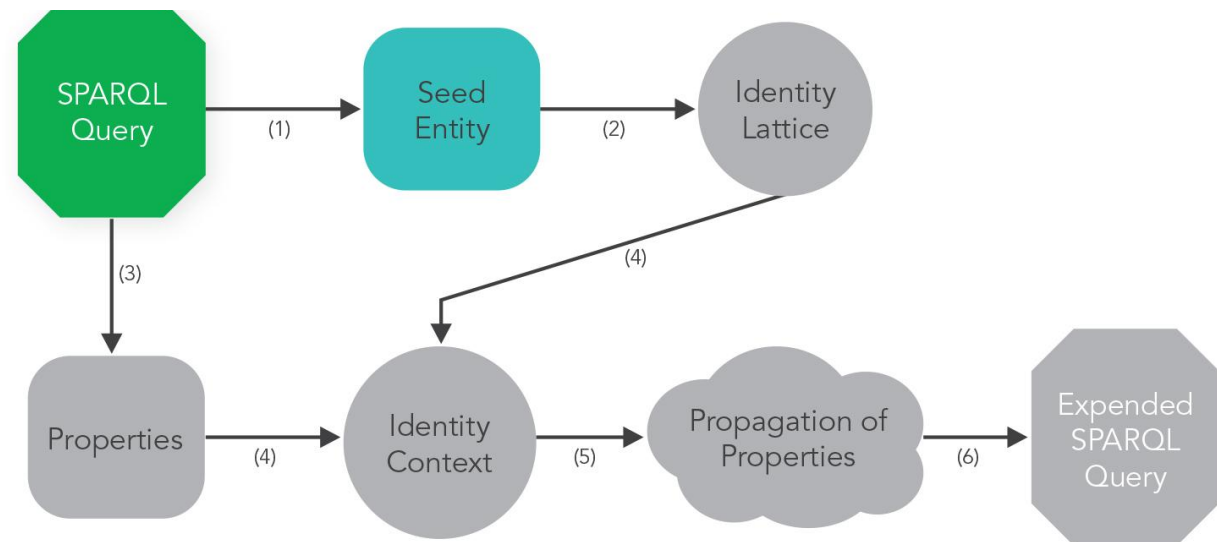


Nombre de républicains condamnés ?



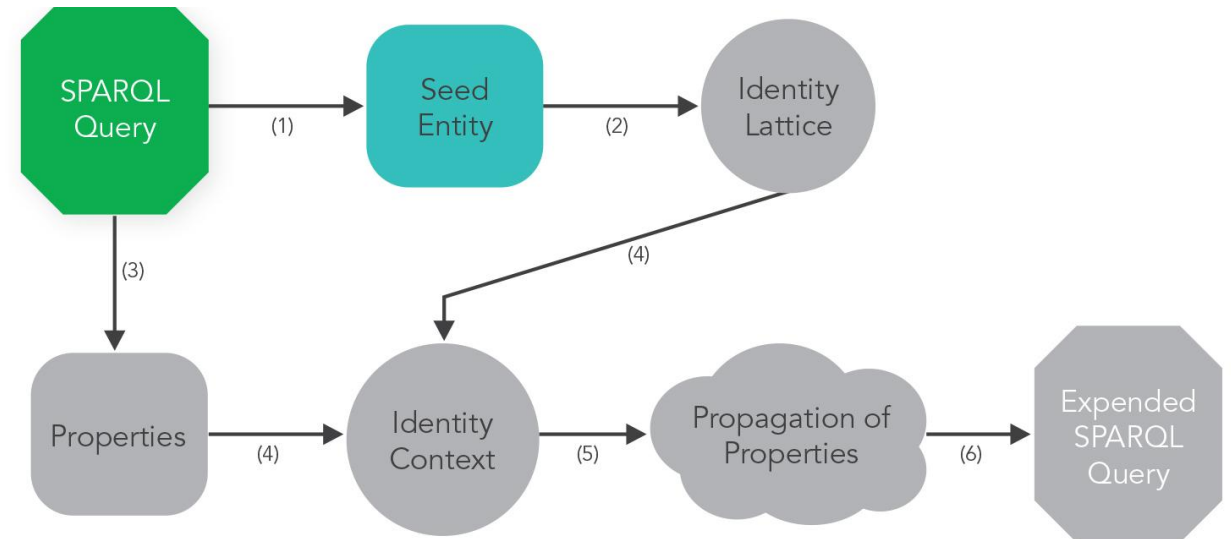
## Nombre de républicains condamnés ?

```
SELECT DISTINCT ?politician ?crime
WHERE {
  ?politician :memberOf :TheRepublicans ;
  :convictedOf ?crime .
}
```



## Nombre de républicains condamnés ?

```
SELECT DISTINCT ?politician ?crime
WHERE {
  ?politician :memberOf :TheRepublicans ;
  :convictedOf ?crime .
}
```



## Changement de nom du parti :

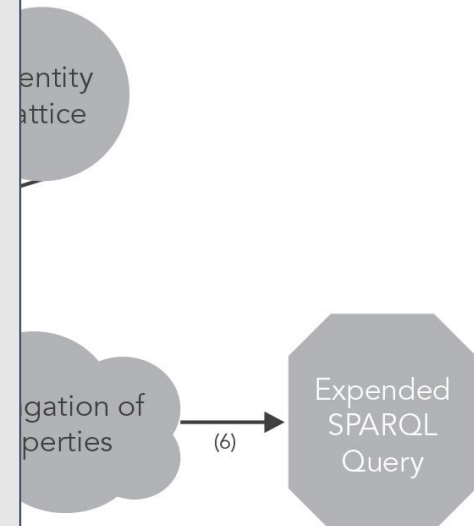
# de résultats sans contexte	2
------------------------------	---

Nombre

```
SELECT DI  
WHERE {  
  ?politi  
  :co  
}
```

Changer

# de résultat  
contexte





Nombre

```
SELECT DI  
WHERE {  
  ?politi  
  :co  
}
```

Changer

# de résultat  
contexte



entity  
attice

igation of  
perties

(6)

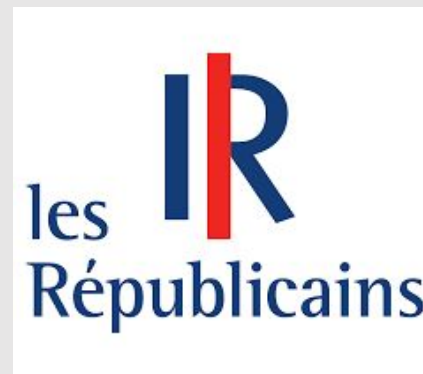
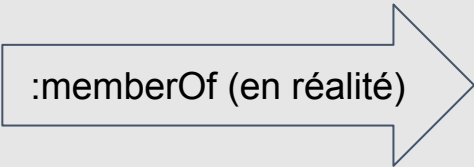
Expended  
SPARQL  
Query

Nombre

```
SELECT DI  
WHERE {  
  ?politi  
  :co  
}
```

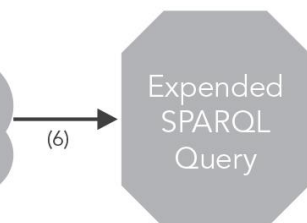
Changer

# de résult  
contexte



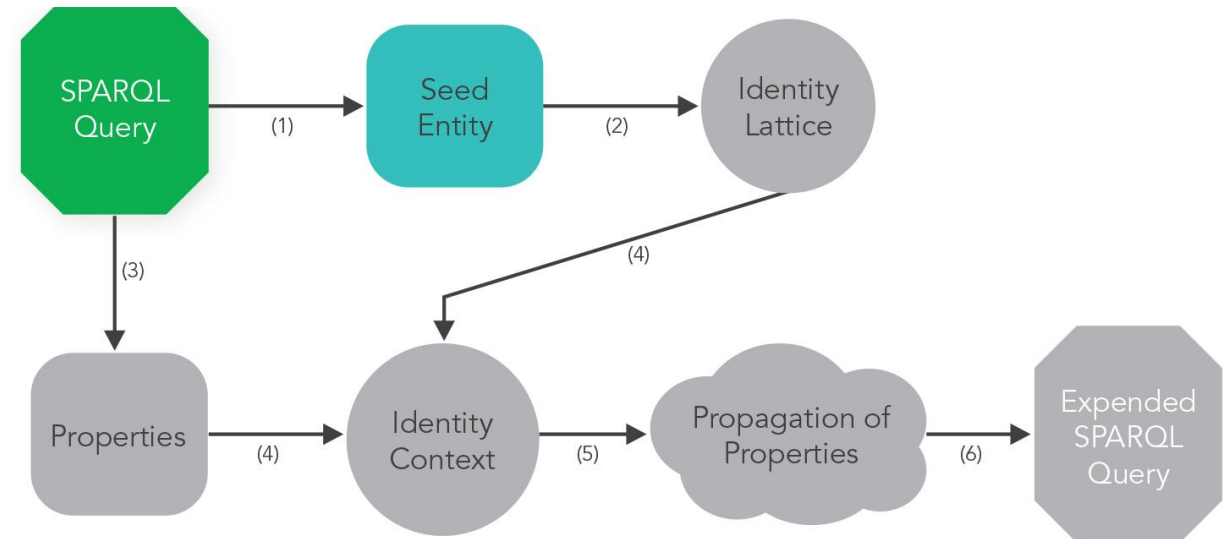
entity  
attice

gation of  
perties



## Nombre de républicains condamnés ?

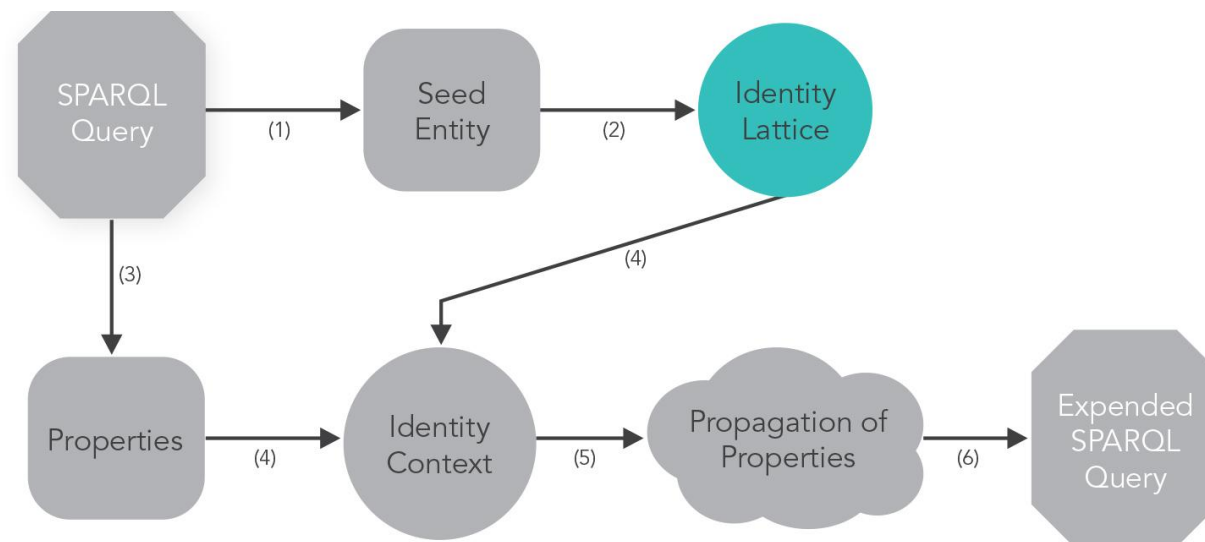
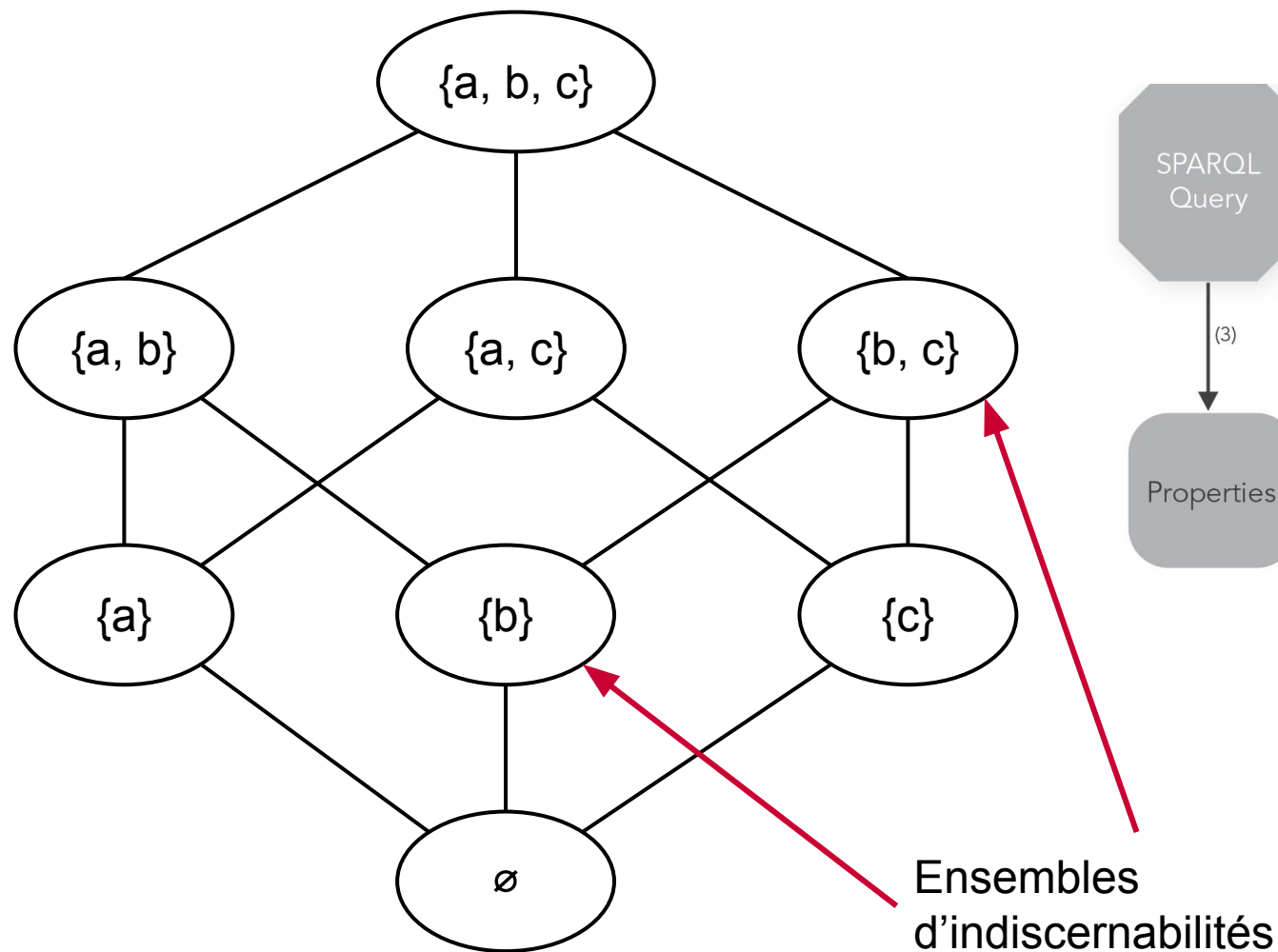
```
SELECT DISTINCT ?politician ?crime
WHERE {
  ?politician :memberOf :TheRepublicans ;
  :convictedOf ?crime .
}
```



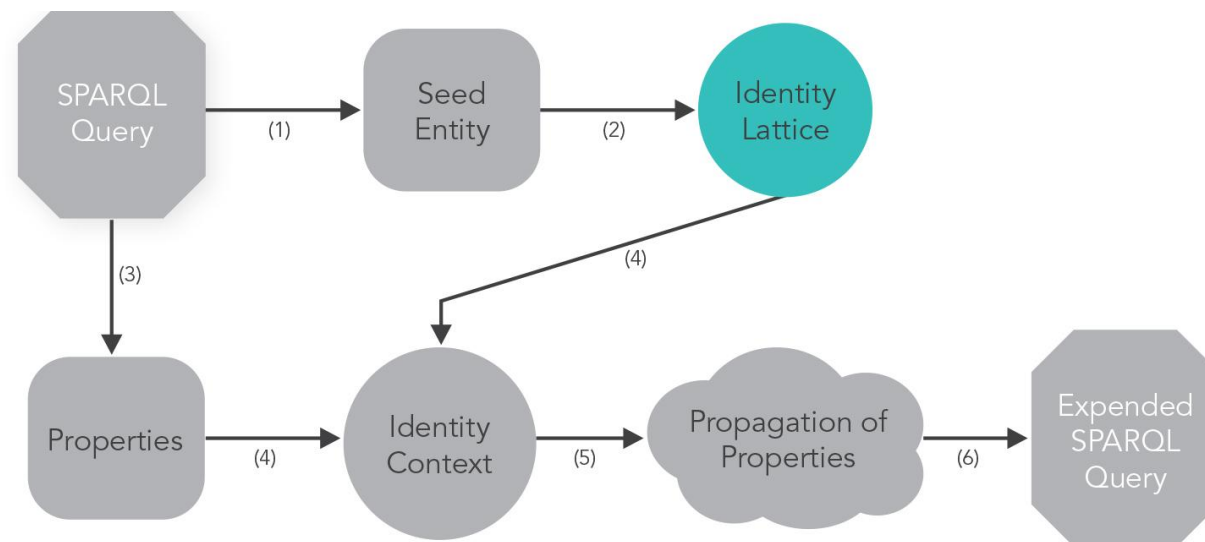
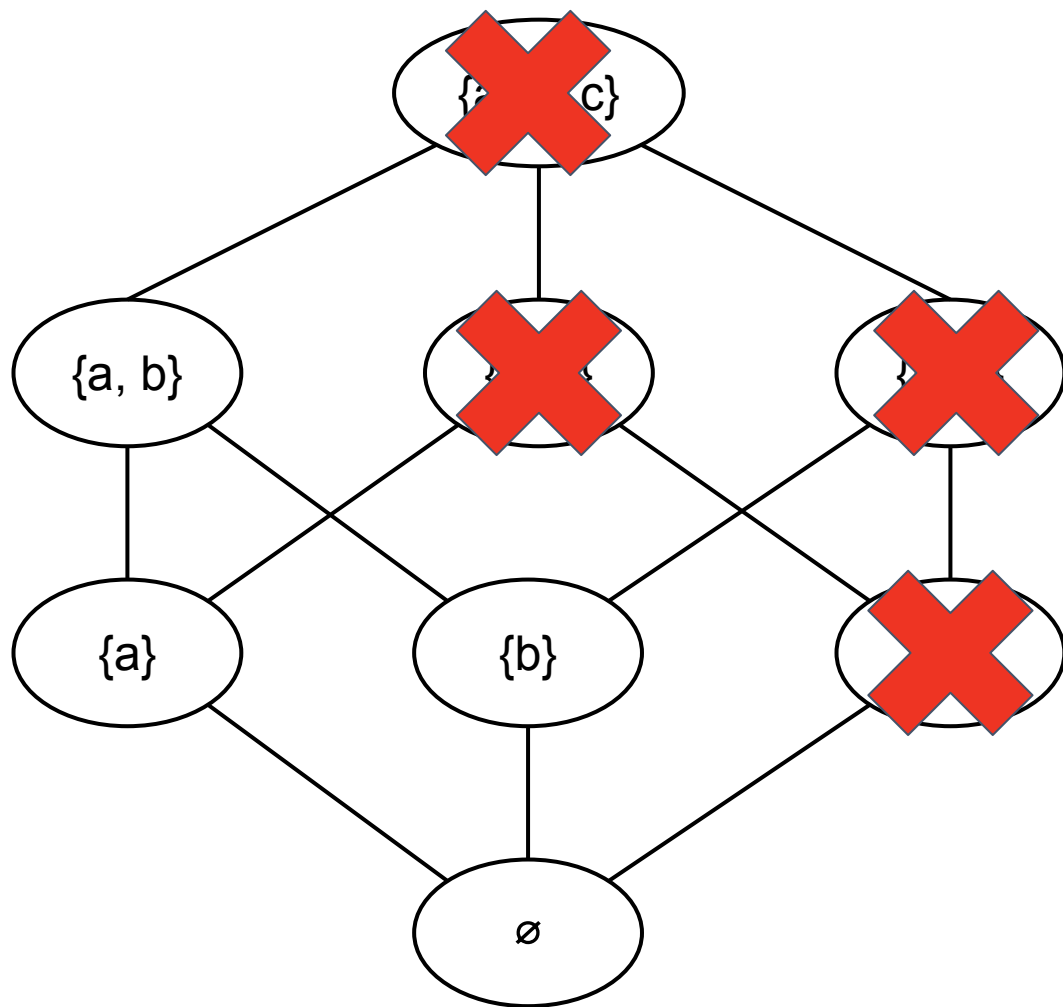
## Changement de nom du parti :

# de résultats sans contexte	2
------------------------------	---

## Construction du treillis d'identité à partir de "Les Républicains"

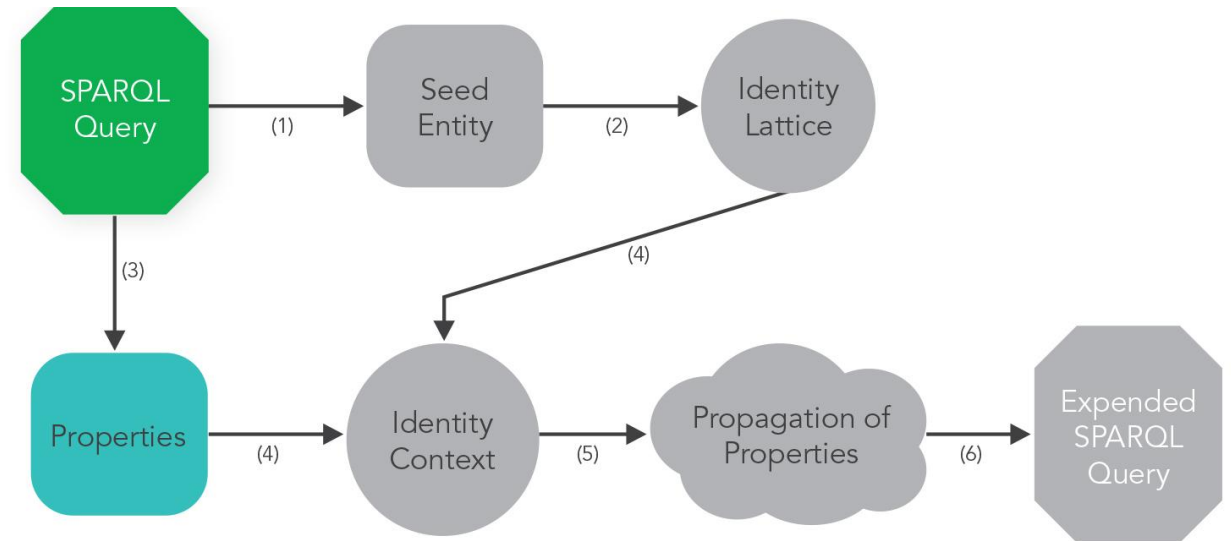


**Elagage** du treillis d'identité si pas d'entité contextuellement identique à "Les Républicains"



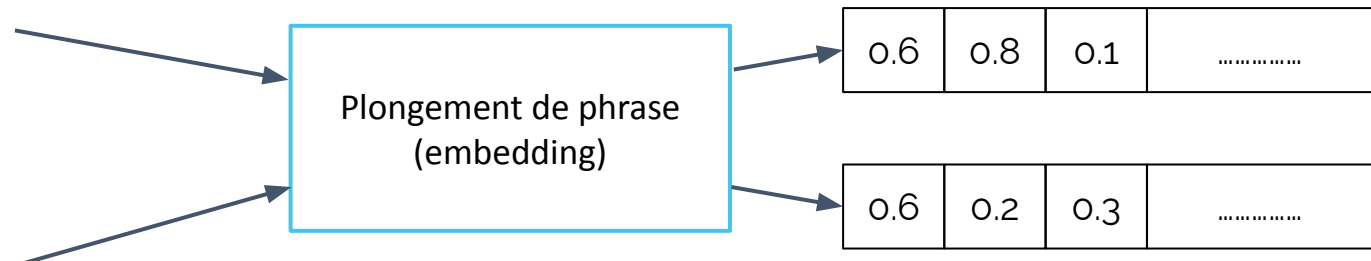
## Extraction de la propriété que l'on souhaite propager

```
SELECT DISTINCT ?politician ?crime
WHERE {
  ?politician :memberOf :TheRepublicans ;
             :convictedOf ?crime .
}
```



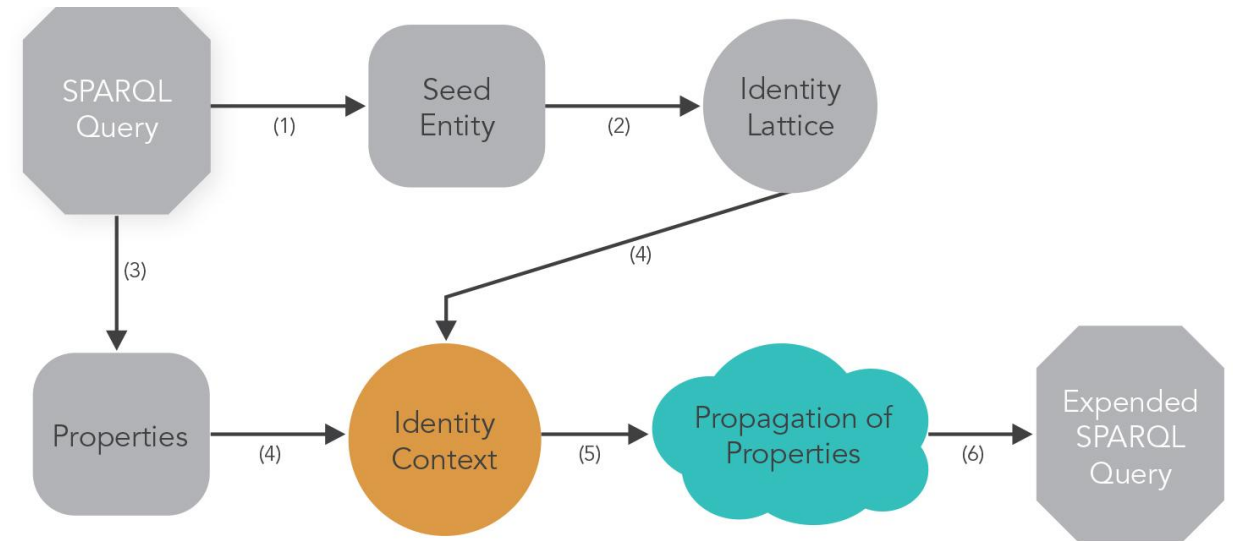
**memberOf** : “the political party of which a person is or has been a member or otherwise affiliated”

**ideology** : “political ideology of an organization or person or of a work (such as a newspaper)”



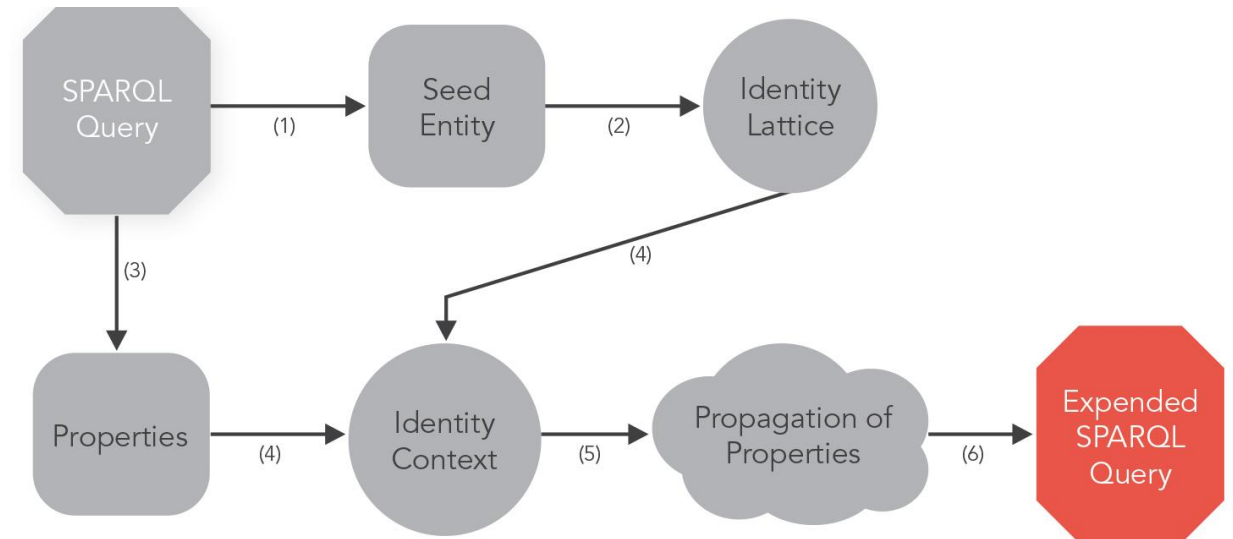
Après **calcul** des ensembles de propagation, on **présente** ceux contenant la propriété à propager

Graine	The Republicans
Ensemble de propagation	<b>member of</b> , political party
Ensemble indiscernable	country, political, ideology
Entités contextuellement identiques	UMP, RPR, UDR, UNR



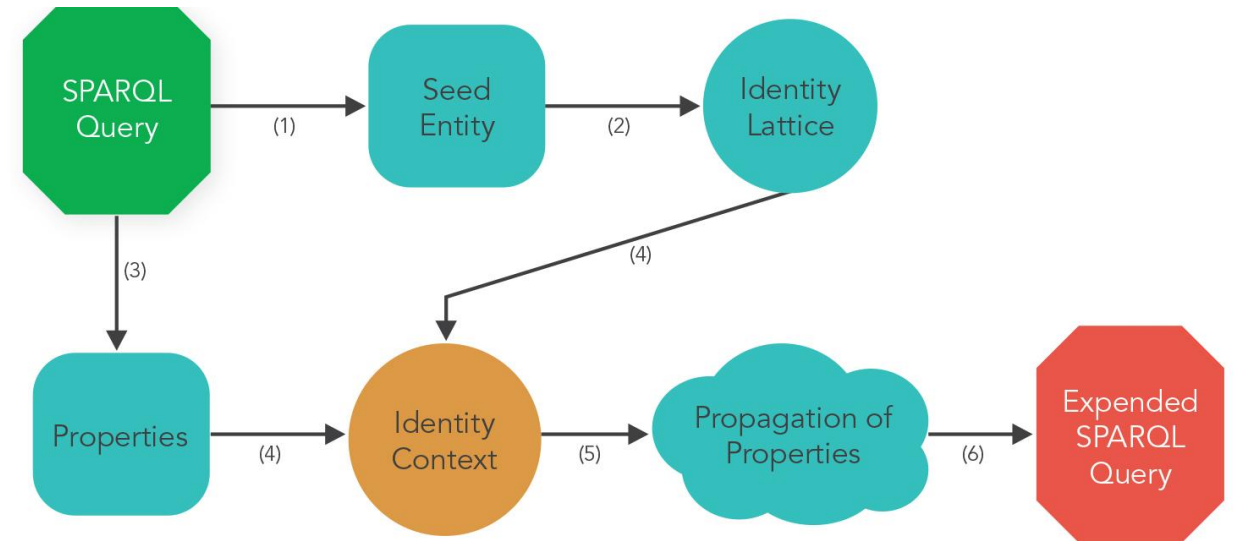
## Réécriture de la requête

```
SELECT DISTINCT ?politician ?crime
WHERE {
  VALUES (?party) {
    (:TheRepublicans) (:UMP) (:RPR)
    (:UDR) (:UNR)
  }
  ?politician :memberOf ?party ;
    :convictedOf ?crime .
}
```





Graine	The Republicans
Ensemble de propagation	member of, political party
Ensemble indiscernable	country, political, ideology
Entités contextuellement identiques	UMP, RPR, UDR, UNR
# de résultats sans contexte	2
# de résultats <b>avec</b> contexte	<b>25</b>



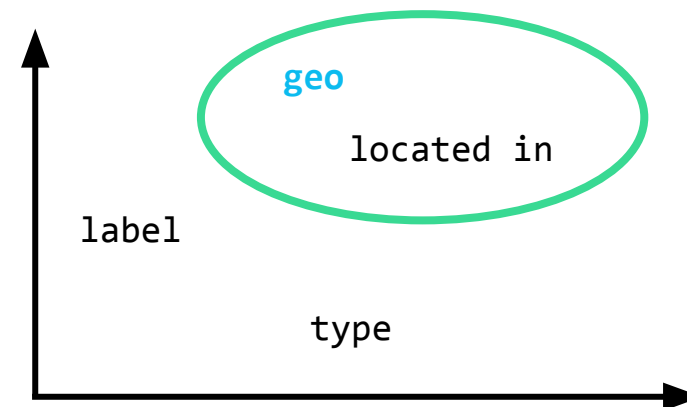
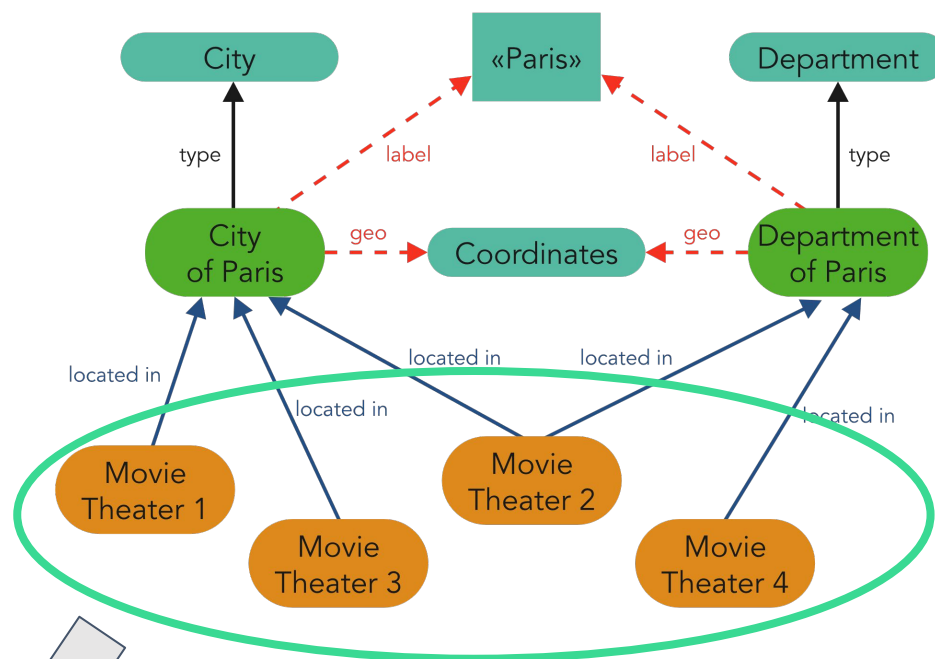
Pour tester la requête sur Wikidata



# ■ **Travaux futurs et conclusion**

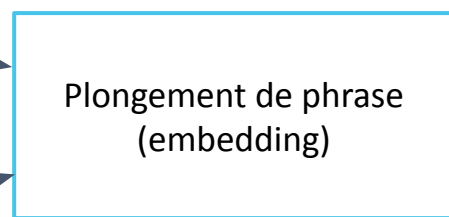


## Identité contextuelle [IC'20, ISWC'20]



**geo** : "geocoordinates of the subject"

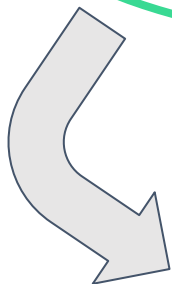
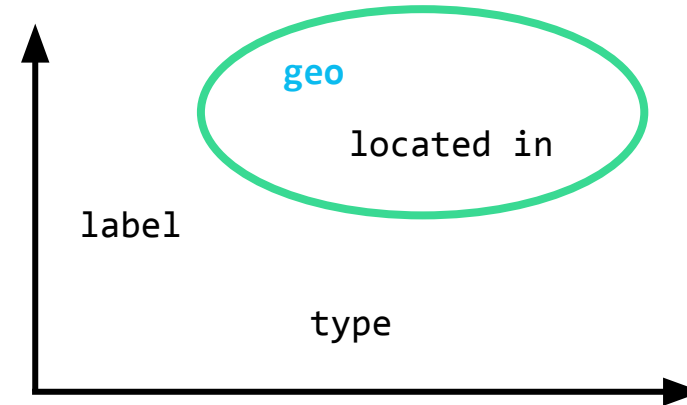
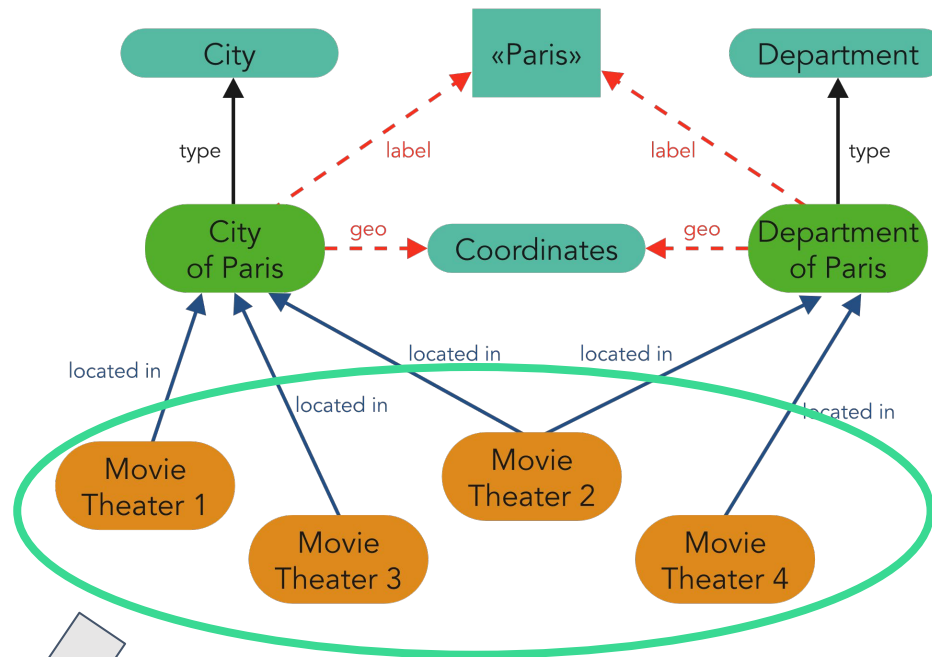
**located in** : "the item is located on the territory of the following administrative entity"



0.6	0.8	0.1	.....
-----	-----	-----	-------

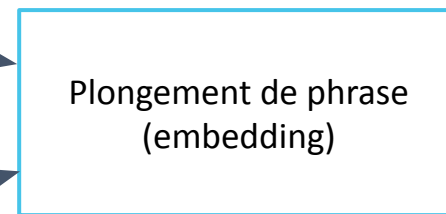
0.6	0.2	0.3	.....
-----	-----	-----	-------

## Identité contextuelle [IC'20, ISWC'20]



**geo** : *"geocoordinates of the subject"*

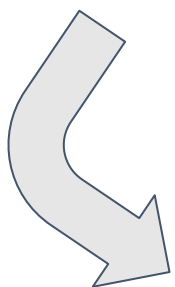
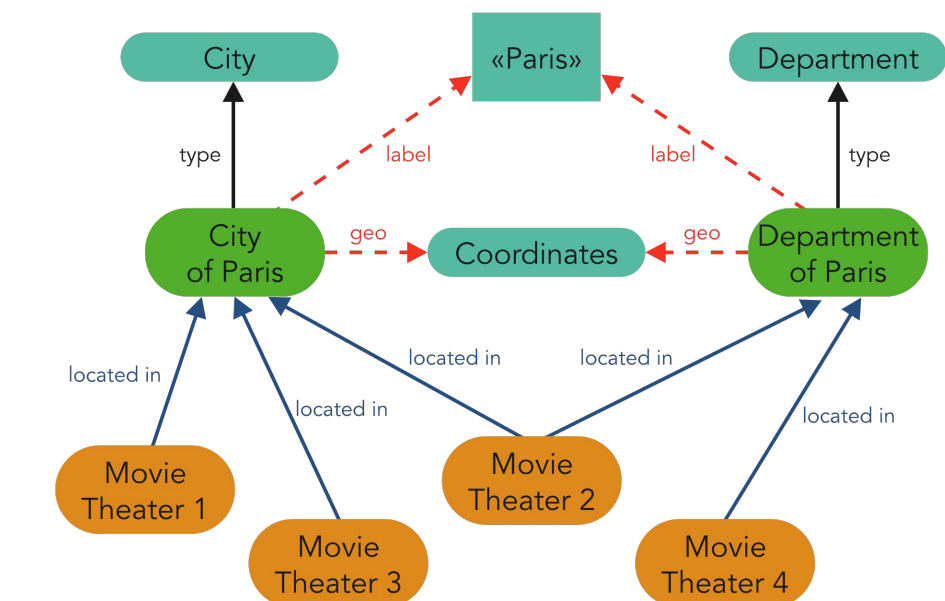
**located in** : *"the item is located on the territory of the following administrative entity"*



0.6	0.8	0.1	.....
-----	-----	-----	-------

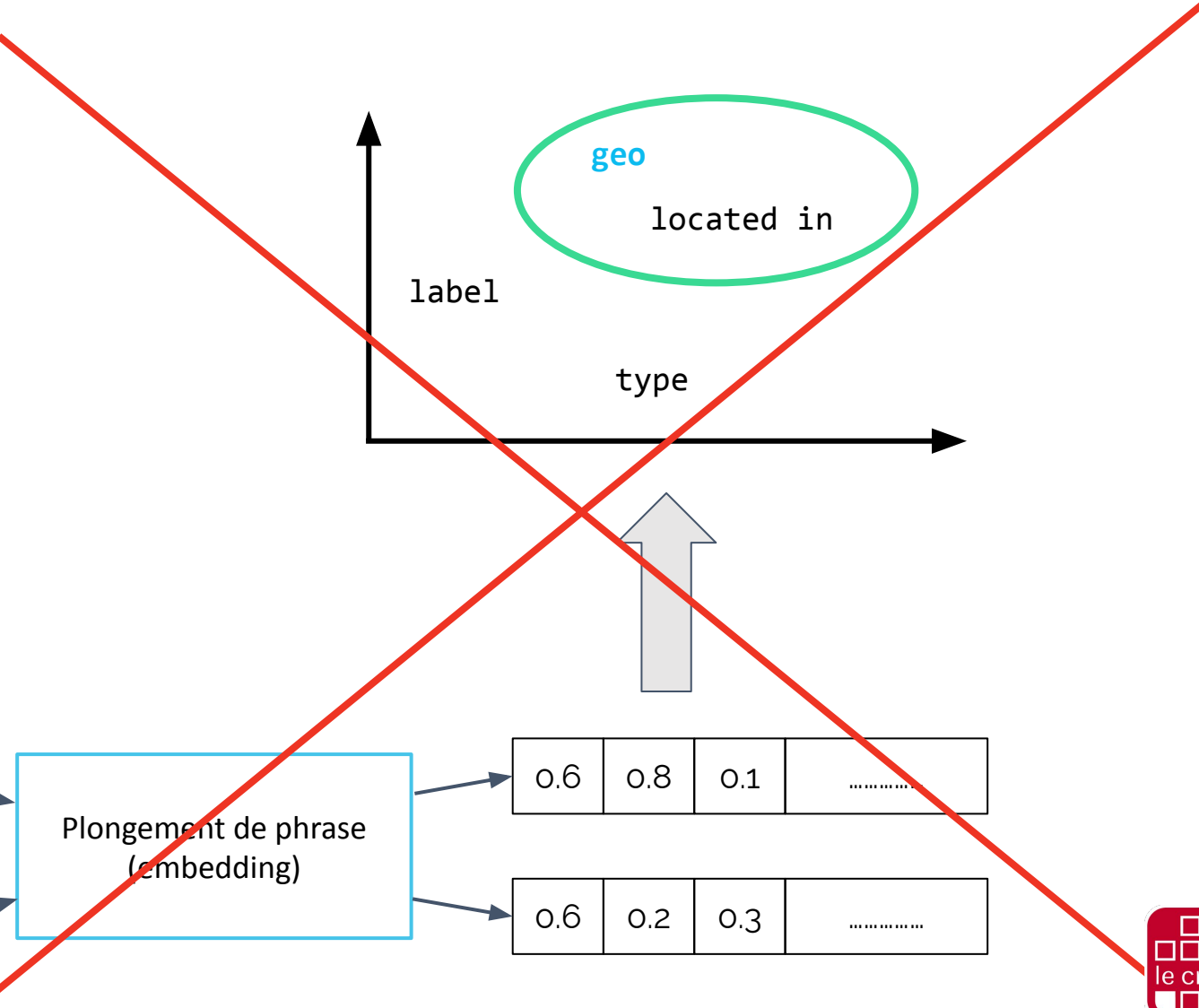
0.6	0.2	0.3	.....
-----	-----	-----	-------

## Identité contextuelle [IC'20, ISWC'20]

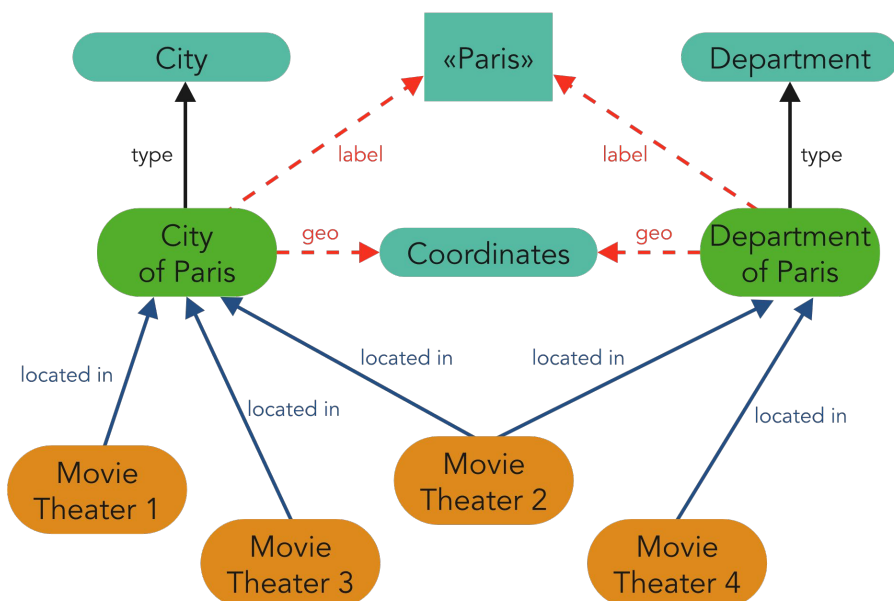


**geo** : "geocoordinates of the subject"

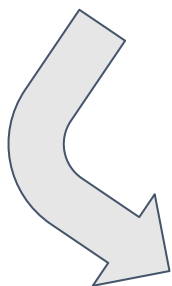
**located in** : "the item is located on the territory of the following administrative entity"



## Identité contextuelle [IC'20, ISWC'20]



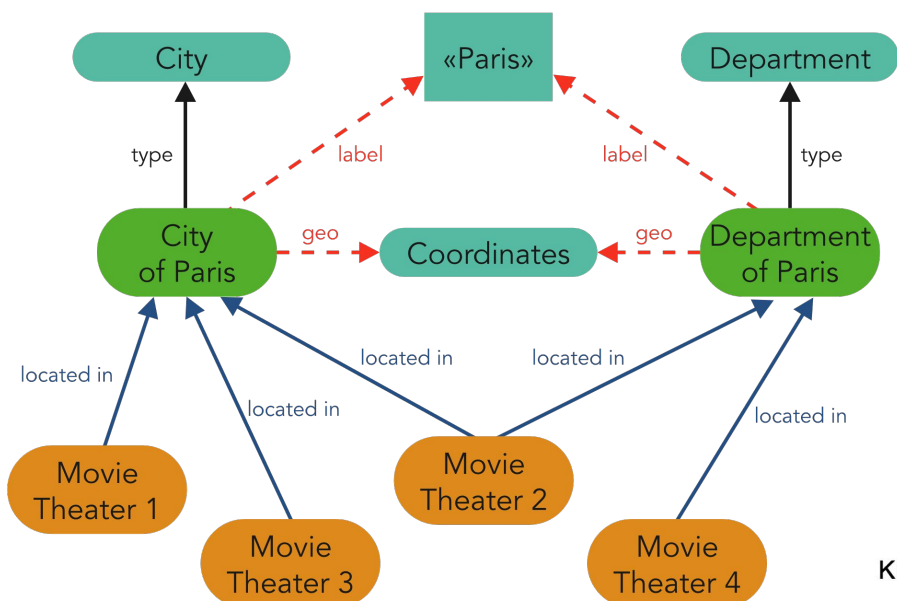
Comment se passer de plongement de phrases ?



**geo** : *"geocoordinates of the subject"*

**located in** : *"the item is located on the territory of the following administrative entity"*

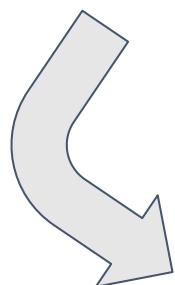
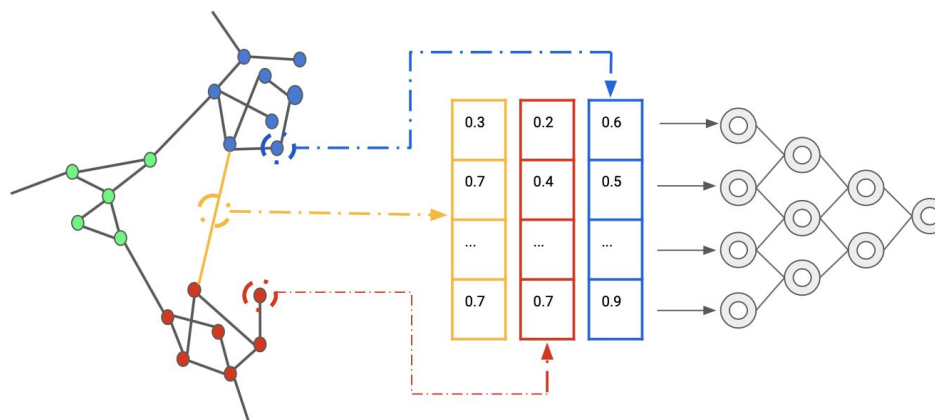
## Identité contextuelle [IC'20, ISWC'20]



Knowledge Graph

Embedded Representation

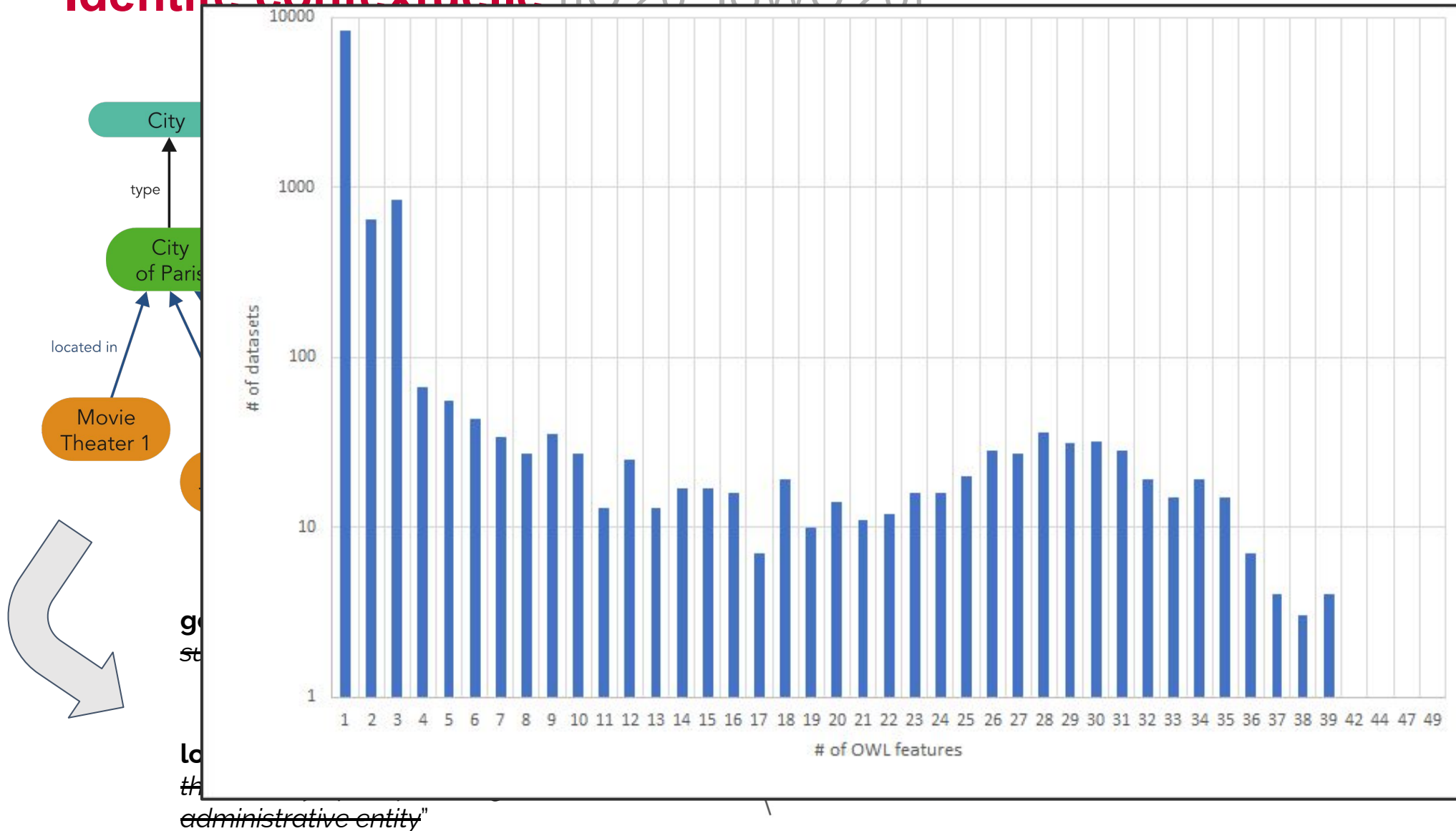
Machine Learning Task



**geo** : *“geocoordinates of the subject”*

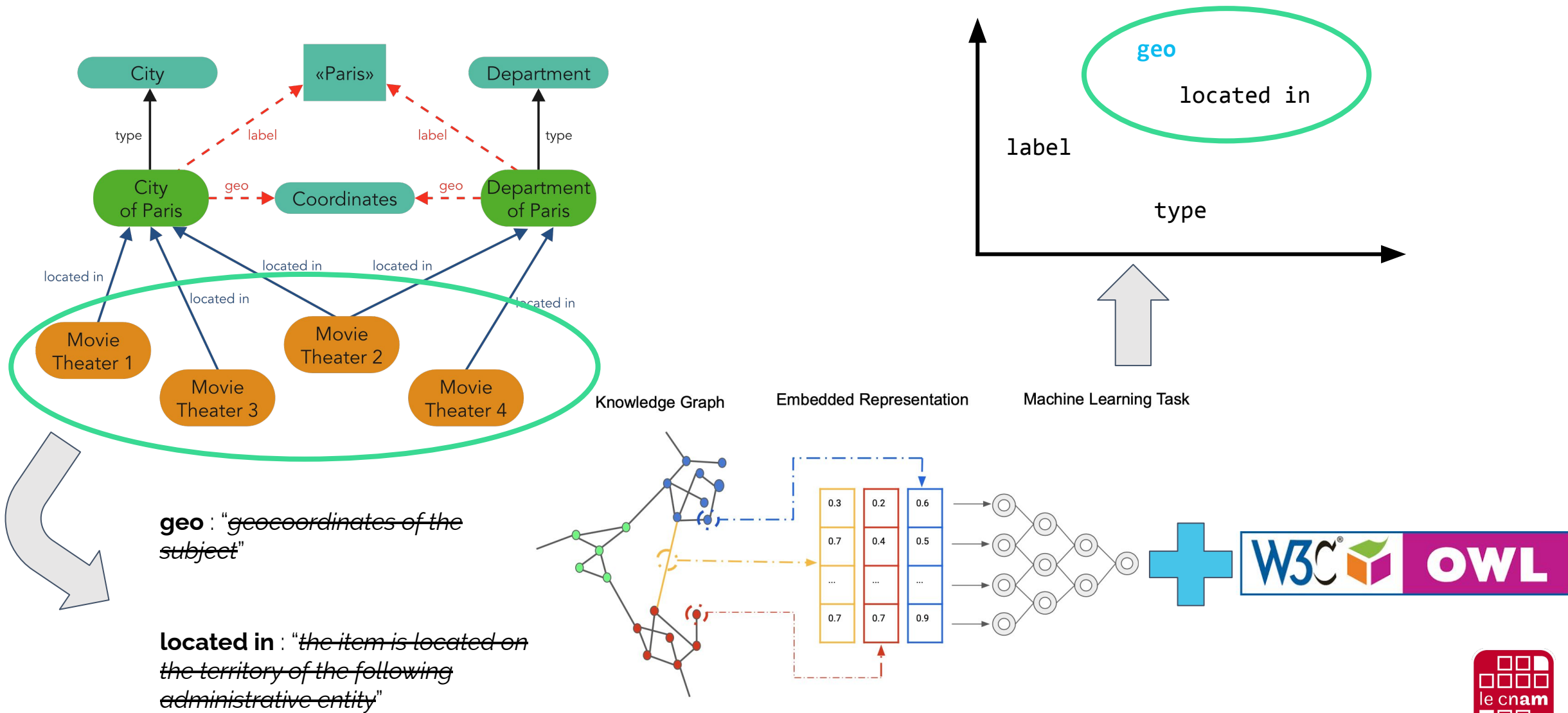
**located in** : *“the item is located on the territory of the following administrative entity”*

## Identité contextuelle IIC'20 ISWC'201





## Identité contextuelle [IC'20, ISWC'20]

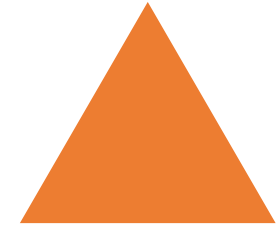
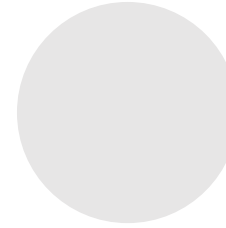




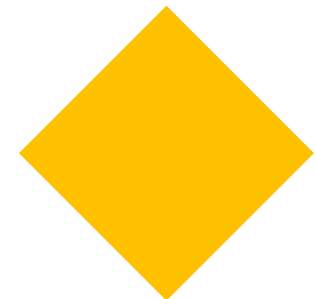
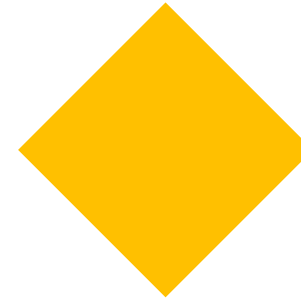
Approche utilisant les plongements de phrases pour trouver des propriétés propageables pour un ensemble d'indiscernabilité donné.



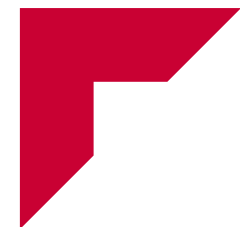
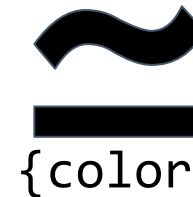
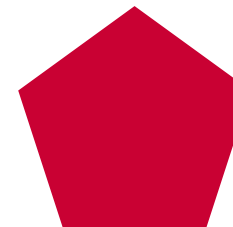
Approche utilisant les plongements de phrases pour trouver des propriétés propageables pour un ensemble d'indiscernabilité donné.



Permet d'utiliser concrètement une version **moins rigide** et donc moins sujette à l'erreur de l'identité.



Le plongement de phrase permet de **découvrir** les propriétés **propageables**, mais les **limites** restent encore floues.



<https://nordf.telecom-paris.fr/en/>

Formalisme fondé sur FOL inspiré de travaux de McCarthy

- Aller plus loin que les simples triplets
- Contexte : temps, espace, croyances, arité  $> 2$ , etc.

“Un *sentiment anti-tabac* s'est développé dans de nombreux *pays* à partir du milieu du *19ème siècle*.”

↑  
Qu'est-ce qu'un  
sentiment  
anti-tabac ?

↑  
Combien ?  
Lesquels ?

↑  
Quand ?