

## Proportion analogique

Ce que cela signifie :

$A : B :: C : D$  («  $A$  est à  $B$  ce que  $C$  est à  $D$  »)

Habituellement 3 postulats :

- $A : B :: A : B$  (réflexivité)
- $A : B :: C : D \rightarrow C : D :: A : B$  (symétrie)
- $A : B :: C : D \rightarrow A : C :: B : D$  (permutation centrale)

Ici uniquement 2 postulats :

- $A : B :: A : B$  (réflexivité)
- $A : B :: C : D \rightarrow C : D :: A : B$  (symétrie)

## Modèle RDF et Graphe de connaissances

Le **modèle RDF** (Resource Description Framework) permet d'écrire des informations de manière systématique interprétable par l'humain et la machine :

```
@prefix gbo: <http://gbo#> .
@prefix rvd: <http://data.rvdata.us/id/> .

rvd:cruise/AT7-21 rdf:type owl:NamedIndividual ,
    gbo:Cruise ;
    gbo:hasChiefScientist rvd:person/100026 ;
    gbo:hasCruiseType rvd:cruise/op_science ;
    gbo:hasEndPortCall rvd:portcall/1033835343 ;
    gbo:hasIdentifier _:genid786 ;
    ...
```

Un **graphe de connaissances** permet de représenter ces données sous forme d'un **graphe dirigé étiqueté**

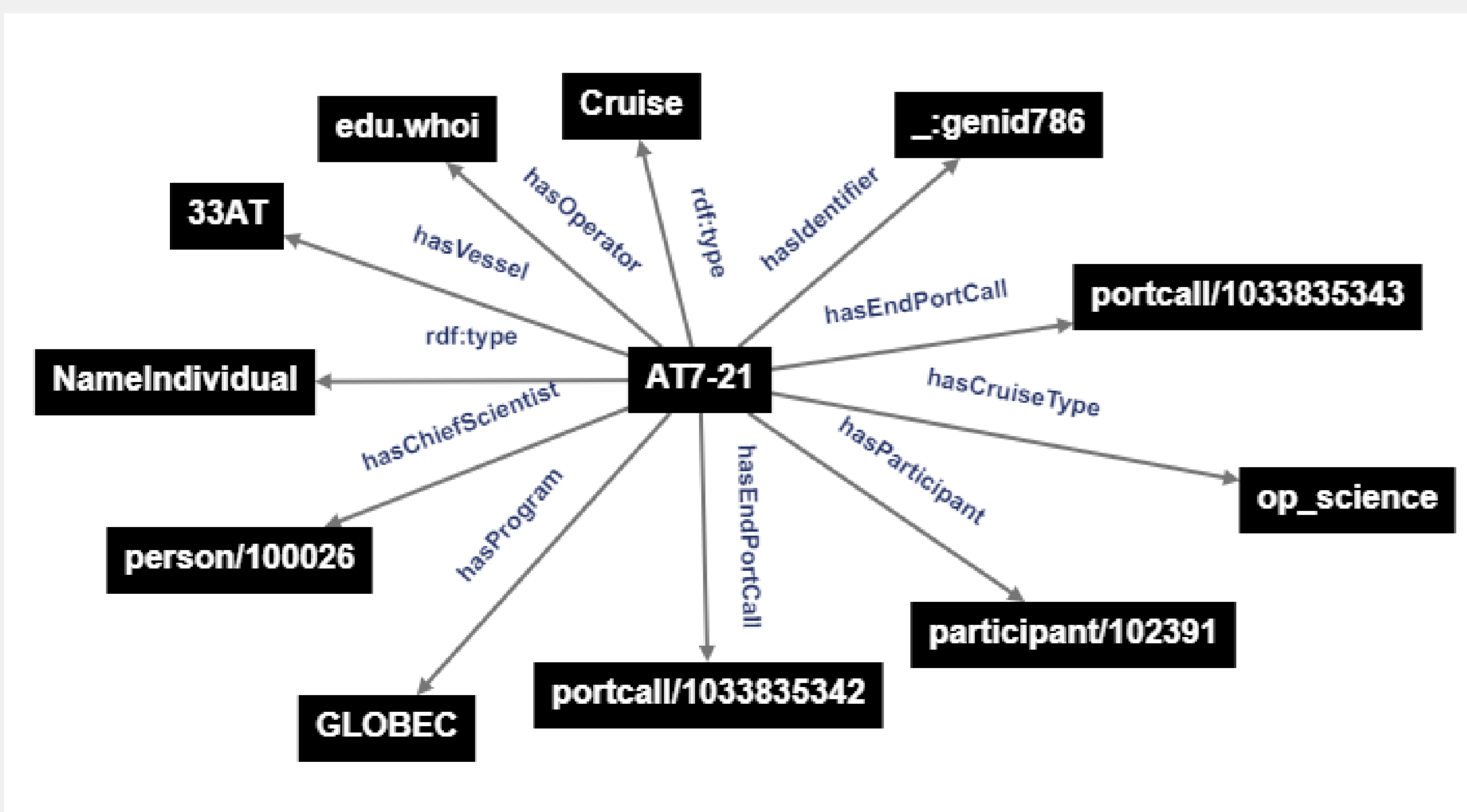


Figure 1. Graphe de connaissances du voisinage de l'entité AT7-21

## Exemple de tâches d'alignement

Le but du projet est d'explorer l'utilisation des proportions analogiques dans ou entre des graphes de connaissances pour réaliser par exemple des alignements.

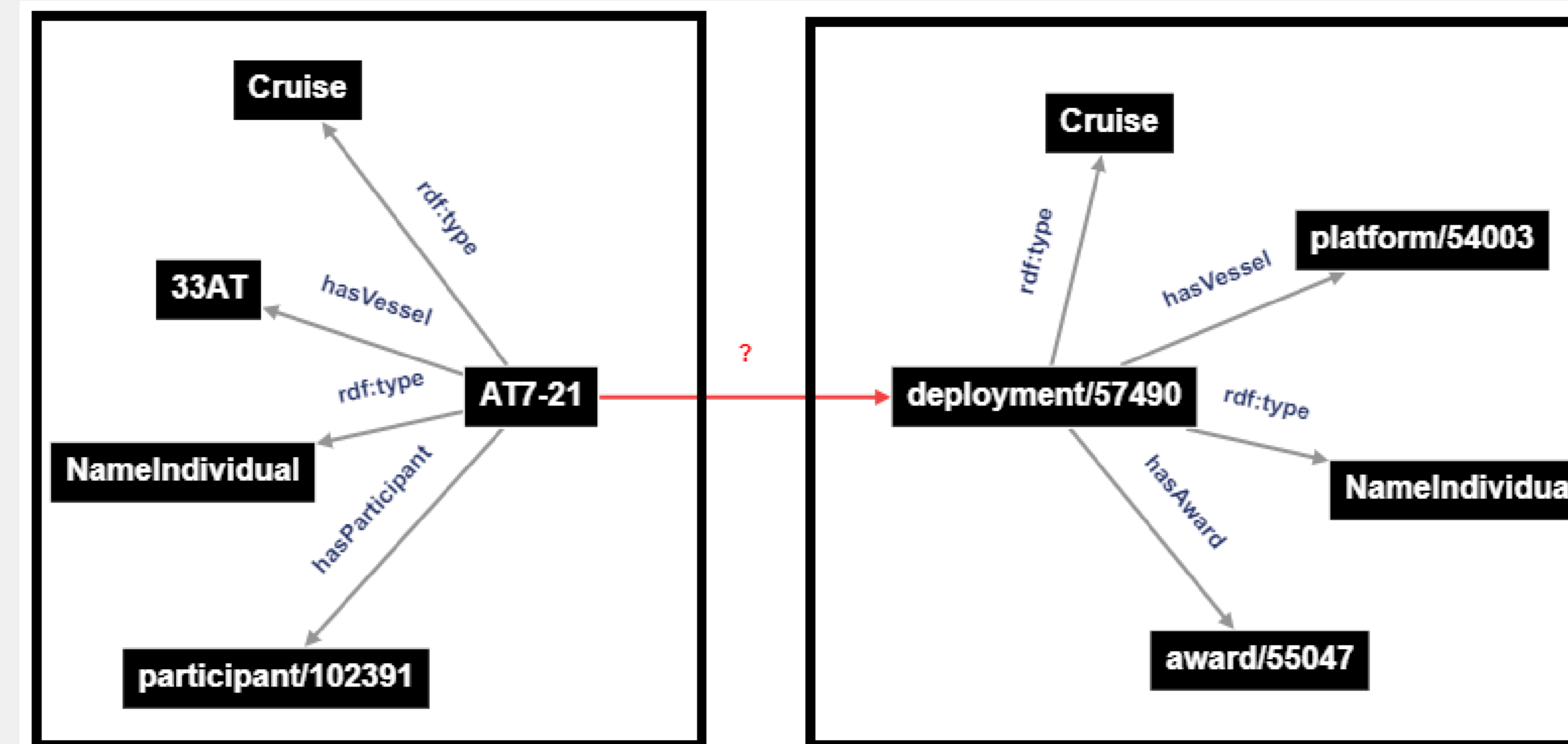


Figure 2. Tâche d'alignement AT7-21

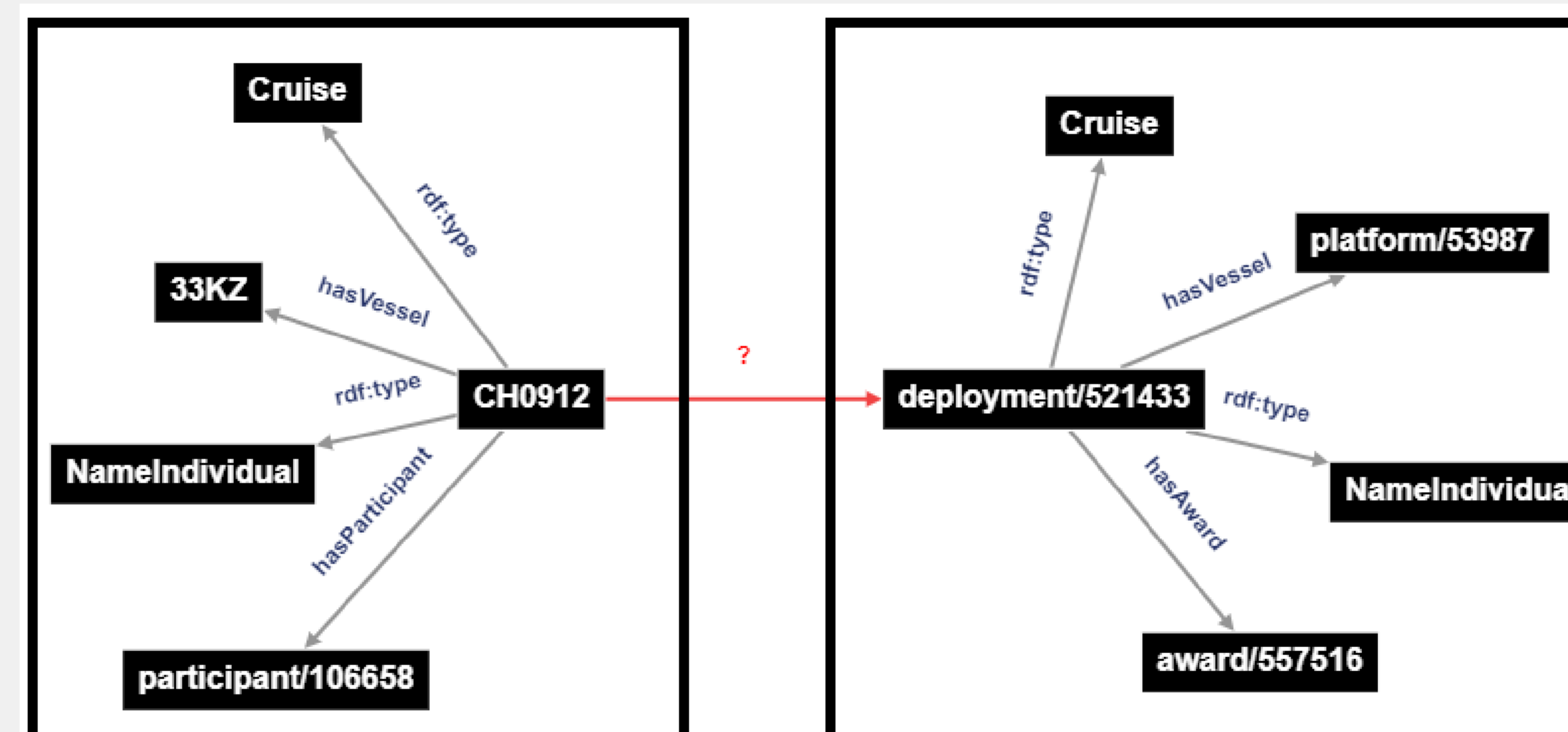


Figure 3. Tâche d'alignement CH0912

## Algorithme

L'algorithme utilise différentes mesures de similarité *dans* un graphe de connaissances ou *entre* graphes de connaissances. Une compare le **type** des entités. Une autre regarde les **littéraux** présents dans les voisinages de chaque entité pour voir à quel point ils sont proches. Une dernière regarde le nombre de **prédicats en commun** présents dans le voisinage des entités.

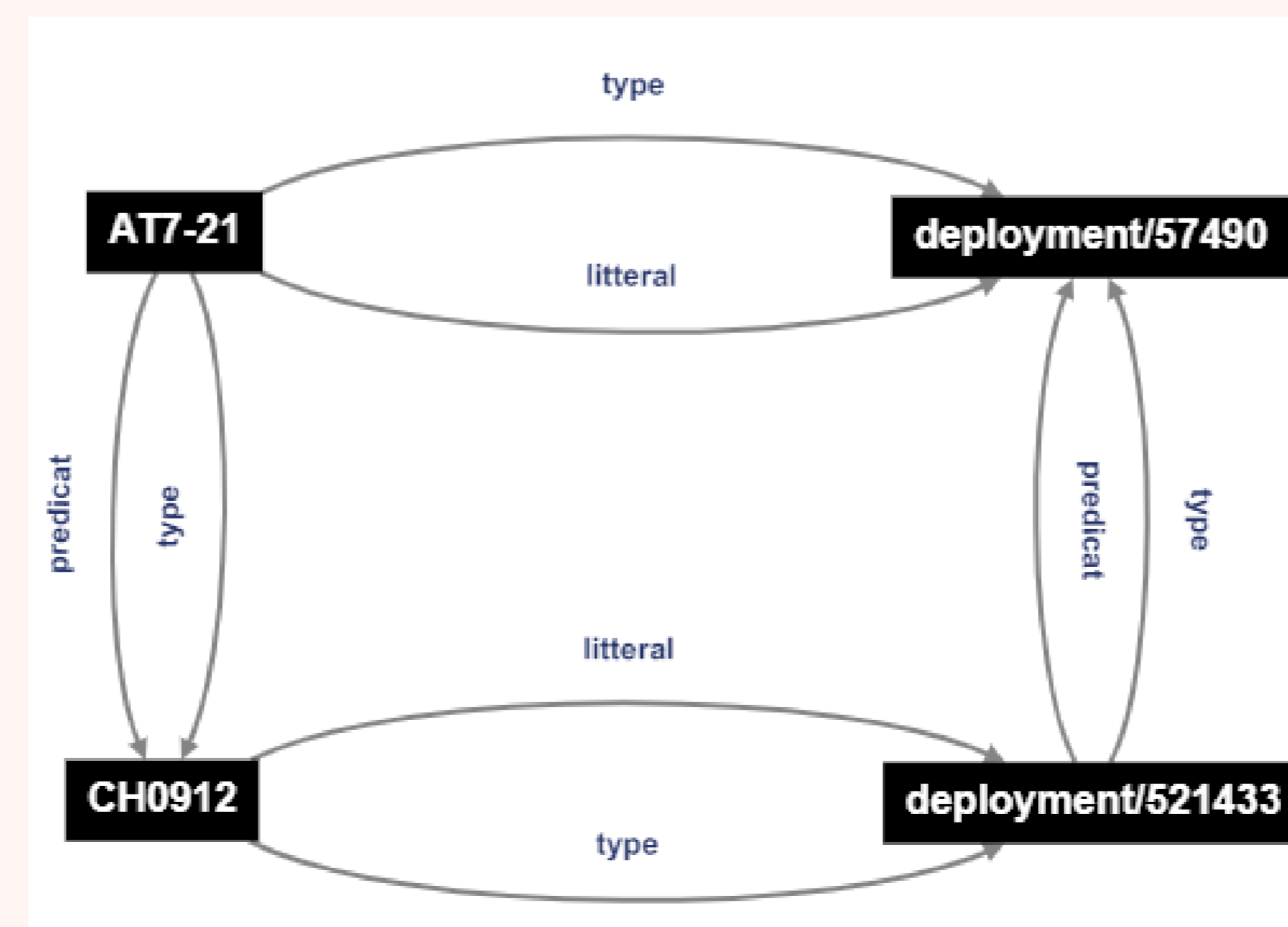


Figure 4. Schéma récapitulatif de l'algorithme

## Formule d'optimisation

Afin de trouver le bon mapping, le but de l'algorithme est de **maximiser la similarité** entre les différents éléments des couples d'entités :

$$T^* = \underset{T}{\operatorname{argmax}} \sum_{a,b \in \mathcal{E}_1} \operatorname{sim}^*(a,b,T(a),T(b)),$$

où  $\operatorname{sim}^*$  est la somme des différents scores de l'algorithme (types, littéraux, prédicats). Ce mapping  $T^*$  correspond à la transformation qu'il faut appliquer à deux éléments en relation  $a$  et  $b$  d'un graphe de connaissances pour obtenir une relation équivalente entre deux éléments  $T(a) = c$  et  $T(b) = d$  d'un autre graphe de connaissances.

## Table de proportions analogiques trouvées

Pour les jeux de données de Geolink<sup>a</sup>, les 225 alignements désirés ont été trouvés.

indice	proportion analogique
81	CH0912 : EN531 :: deployment/521433 : deployment/521426
82	AT18-08 : CH0912 :: deployment/568087 : deployment/521433
83	CH0912 : MV1101 :: deployment/521433 : deployment/473222
84	AE1314 : CH0912 :: deployment/543580 : deployment/521433
85	AE0810 : CH0912 :: deployment/58062 : deployment/521433
86	CH0912 : HRS070714AB :: deployment/521433 : deployment/58668
87	CH0912 : NH1006 :: deployment/521433 : deployment/59062
88	CH0912 : KM0919 :: deployment/521433 : deployment/58874
89	AT26-15 : CH0912 :: deployment/517377 : deployment/521433
90	<b>AT7-21 : CH0912 :: deployment/57490 : deployment/521433</b>
91	AT21-02 : CH0912 :: deployment/535929 : deployment/521433
92	CH0912 : OC443 :: deployment/521433 : deployment/58725
93	AE1409 : CH0912 :: deployment/565190 : deployment/521433
94	AT29-04 : CH0912 :: deployment/568866 : deployment/521433
95	CH0912 : KN195-02 :: deployment/521433 : deployment/651161

Table 1. 15 proportions analogiques pour Geolink

Comme chaque proportion analogique implique deux alignements gold, les mappings et les relations liant les couples d'entités sont des relations d'équivalence.

## References

- [1] Lisa Ehlringer and Wolfram W6B. Towards a definition of knowledge graphs. In Michael Martin, Marti Cuquet, and Erwin Folmer, editors, *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems - SEMANTICS2016 and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCESS'16) co-located with the 12th International Conference on Semantic Systems (SEMANTICS 2016)*, Leipzig, Germany, September 12-15, 2016, volume 1695 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.
- [2] Henri Prade and Gilles Richard. Analogical proportions and analogical reasoning - an introduction. In David W. Aha and Jean Lieber, editors, *Case-Based Reasoning Research and Development - 25th International Conference, ICCBR 2017, Trondheim, Norway, June 26-28, 2017, Proceedings*, volume 10339 of *Lecture Notes in Computer Science*, pages 16-32. Springer, 2017.
- [3] Oren Sultan and Dafna Shahaf. Life is a circus and we are the clowns : Automatically finding analogies between situations and processes. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 3547-3562. Association for Computational Linguistics, 2022.

a. <https://oaei.ontologymatching.org/2020/geolinkcruise/index.html>