

AGENCE NATIONALE DE LA RECHERCHE
ANR

cap-digital

Comparaison d'ontologies géographiques

Ammar Mechouche



IGN / COGIT

Comparaison d'ontologies géographiques

- Etat de l'art très pauvre sur la comparaison globale d'ontologies

- Méthodes intéressantes trouvées

- Le Duy Ngan, Angela Eck Soong Goh, Le Quang Hung: Comparing two ontologies. Int. J. Web Eng. Technol. 5(1): 48-68 (2009) : *comapre le nombre de propriétés communes qui relie des concepts communs. Pas d'alignement préalable.*

- Alexander Maedche, Steffen Staab: Measuring Similarity between Ontologies. EKAW 2002: 251-263. *basée sur le nombre de concepts communs entre les voisinages de deux concepts communs. Pas d'alignement préalable.*

- Ce qui est présenté ici :

- Test de la méthode de Maedche & S. Staab 2002 ;
 - Analyse de ce type de méthodes ;
 - Nouvelles pistes.



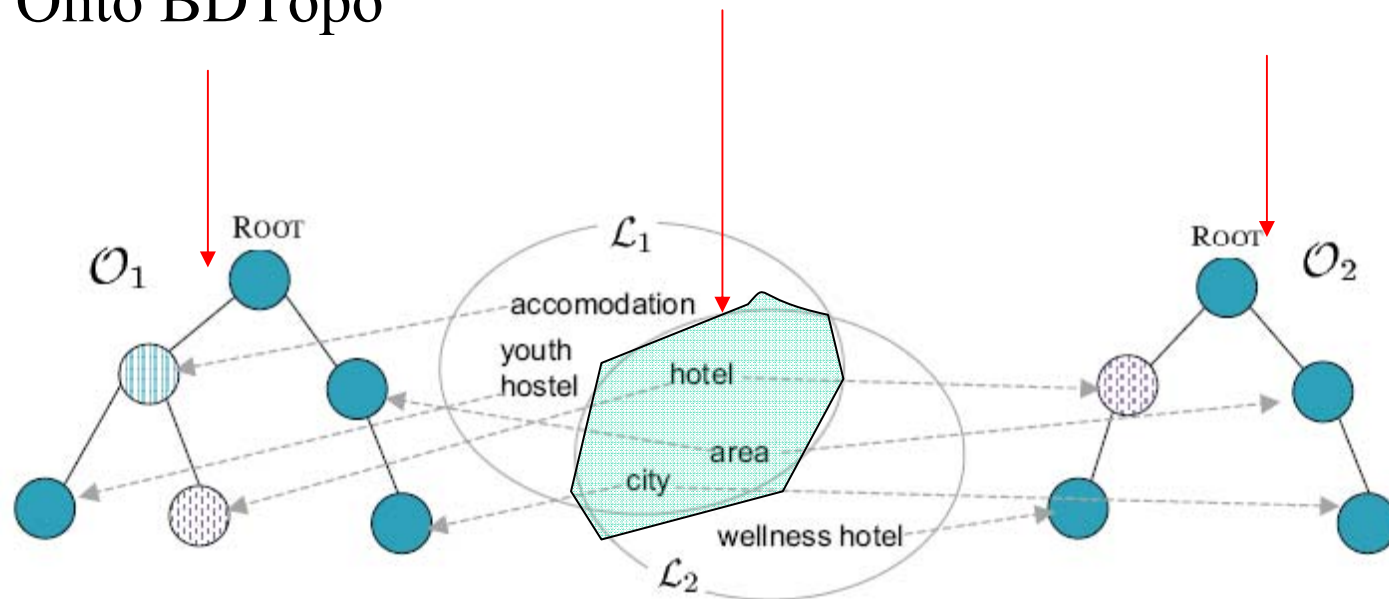
Comparaison d'ontologies géographiques

-Test de la méthode de Maedche & S. Staab 2002

Résultats de
L'alignement

Onto BDTopo

Onto BDCarto



Comparaison d'ontologies géographiques

- Test de la méthode de Maedche & S. Staab 2002

The taxonomic overlap (TO) between \mathcal{H}_1 and \mathcal{H}_2 as seen from the concepts referred to by L may then be computed by following \mathcal{F}_1^{-1} and \mathcal{F}_2^{-1} back to the common lexicon.

$$TO'(L, \mathcal{O}_1, \mathcal{O}_2) := \frac{\overset{M}{A} |\mathcal{F}_1^{-1}(SC(\mathcal{F}(\{L\}), \mathcal{H}_1)) \cap \mathcal{F}_2^{-1}(SC(\mathcal{F}(\{L\}), \mathcal{H}_2))|}{\underset{X}{L' \text{ e alignement } (L)} |\mathcal{F}_1^{-1}(SC(\mathcal{F}(\{L\}), \mathcal{H}_1)) \cup \mathcal{F}_2^{-1}(SC(\mathcal{F}(\{L\}), \mathcal{H}_2))|}$$

the simplest assumption is that the L is simply missing from \mathcal{L}_2^c , but when comparing the two hierarchies the optimistic taxonomic approximation is the one that searches for the maximum overlap given a fictive membership of L to \mathcal{L}_2^c by

$$TO''(L, \mathcal{O}_1, \mathcal{O}_2) := \max_{C \in \mathcal{L}_2^c} \left\{ \frac{\overset{M}{A} |\mathcal{F}_1^{-1}(SC(\mathcal{F}(\{L\}), \mathcal{H}_1)) \cap \mathcal{F}_2^{-1}(SC(C), \mathcal{H}_2)|}{\underset{X}{C' \text{ e alignement } (C)} |\mathcal{F}_1^{-1}(SC(\mathcal{F}(\{L\}), \mathcal{H}_1)) \cup \mathcal{F}_2^{-1}(SC(C), \mathcal{H}_2)|} \right\}$$

Given these premises the averaged similarity \overline{TO} between two taxonomies ($\mathcal{H}_1, \mathcal{H}_2$) of ontologies ($\mathcal{O}_1, \mathcal{O}_2$) may then be defined by:

$$\overline{TO}(\mathcal{O}_1, \mathcal{O}_2) := \frac{1}{|\mathcal{L}_1^c|} \sum_{L \in \mathcal{L}_1^c} TO(L, \mathcal{O}_1, \mathcal{O}_2), \text{ with}$$

$$TO(L, \mathcal{O}_1, \mathcal{O}_2) := \begin{cases} TO'(L, \mathcal{O}_1, \mathcal{O}_2) & \text{if } L \in \mathcal{L}_2^c \\ TO''(L, \mathcal{O}_1, \mathcal{O}_2) & \text{if } L \notin \mathcal{L}_2^c \end{cases}$$

$TO(O_BDTopo, O_BDCarto) = \mathbf{21\%}$ - dépendance de l'alignement



Comparaison d'ontologies géographiques

La méthode de Maedche & S. Staab considère les ontologies à comparer dans leur globalité, alors que :

- On peut avoir à comparer une ontologie géographique (qui décrit un seul thème) avec une autre qui décrit plusieurs thèmes (peut être avec un niveau de détail différent)
- Cette méthode compare uniquement les structures des deux ontologies. Pas de critère de comparaison (par exemple niveau de détail)
- → (1) besoin de partitionner les ontologies à comparer ; (2) ensuite comparer les différentes parties selon un critère donné.



Comparaison d'ontologies géographiques

Méthode :

1. Utiliser l'alignement pour partitionner les deux ontologies
2. Déterminer les parties des deux ontologies qui sont fortement corrélées
3. Comparer ces parties corrélées selon le critère niveau de détail

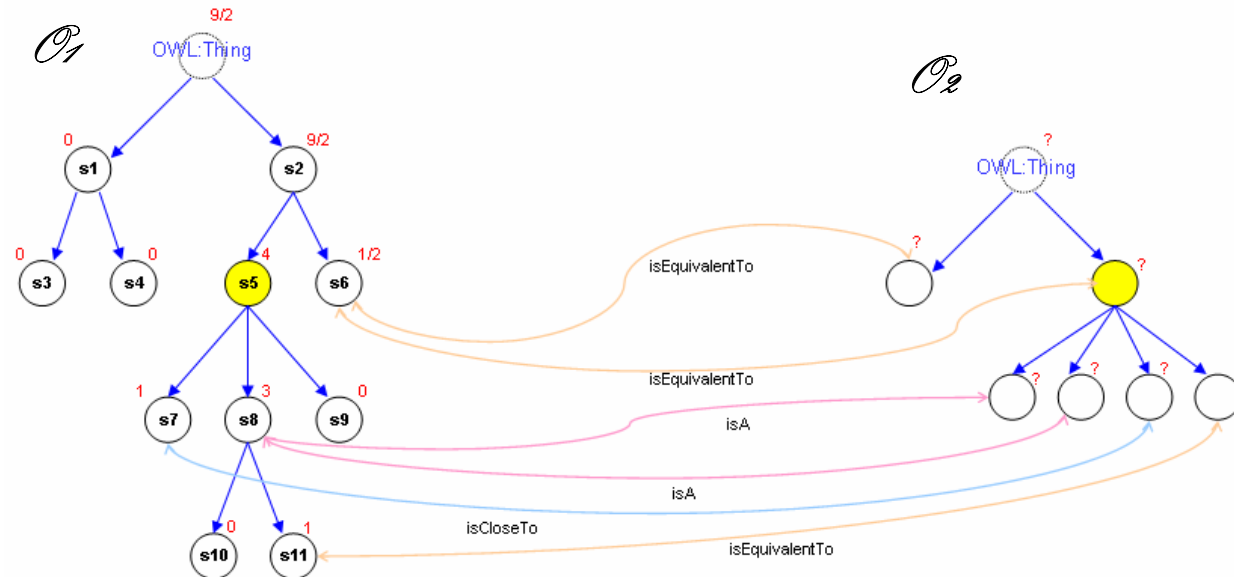


Comparaison d'ontologies géographiques

1. Partitionnement des ontologies :

- Résultats de l'alignement (O1, O2) (**asymétrique**)
 - $\langle t1, t1', isA \rangle$
 - $\langle t1, t1', isEquivalentTo \rangle$
 - $\langle t1, t1', isCloseTo \rangle$

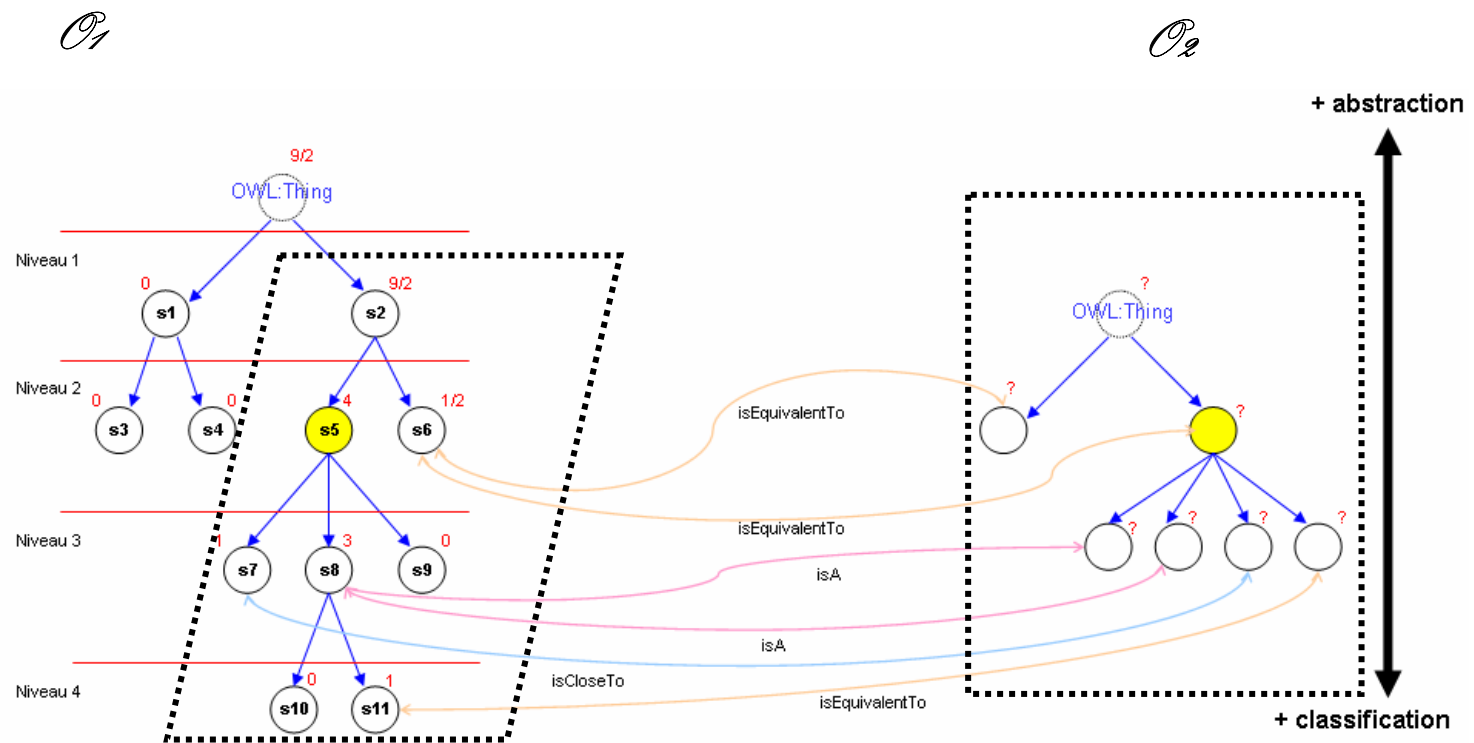
- Affecter des poids aux sommets de O1 en se basant sur les mappings précédents



Comparaison d'ontologies géographiques

1. Partitionnement des ontologies :

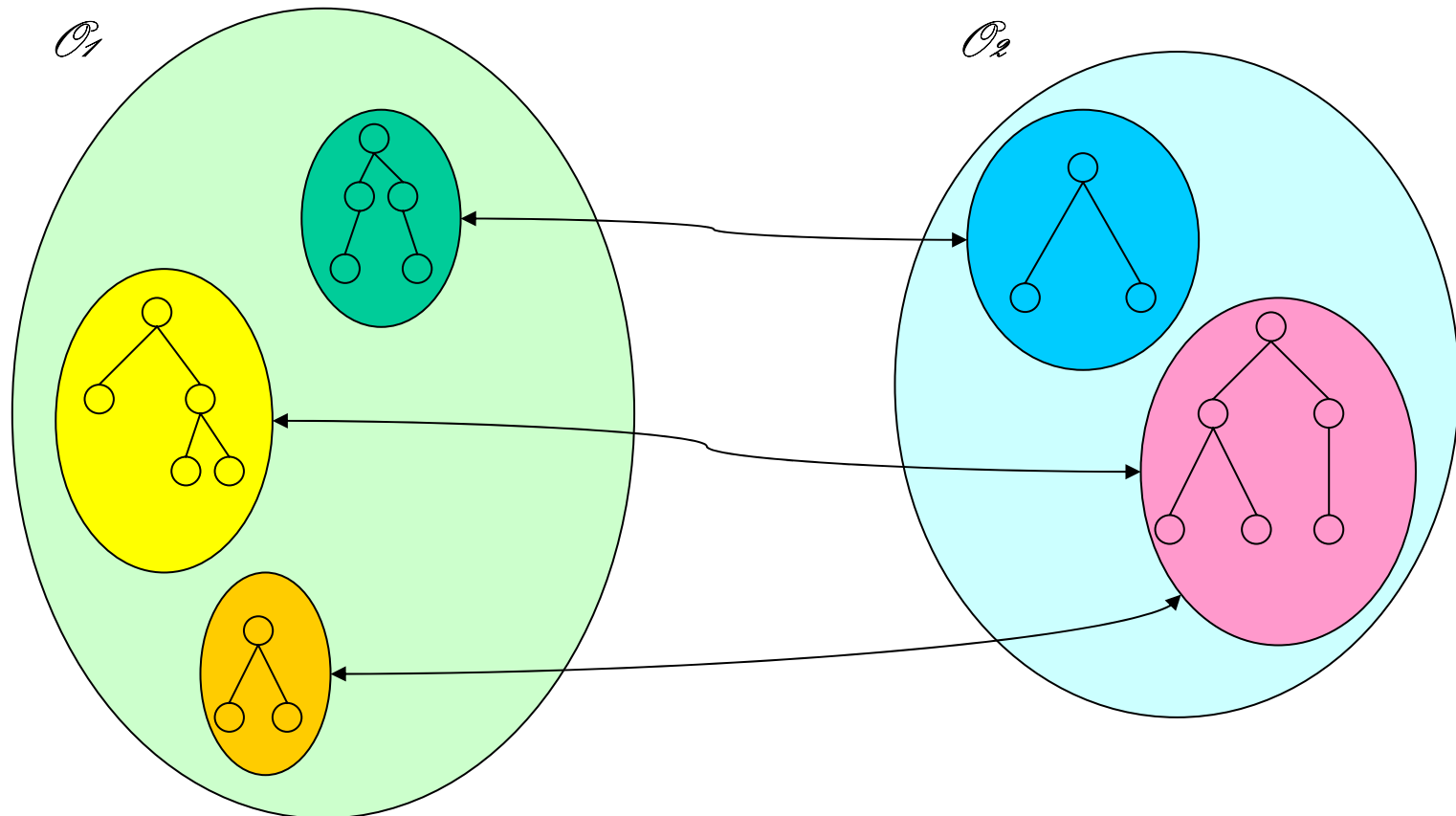
- Déterminer les sommets importants en tenant compte de la profondeur, du poids et de l'environnement de chaque sommet



Comparaison d'ontologies géographiques

2. Calcul des corrélations entre partitions :

- Adaptation du modèle vectoriel (connu en recherche d'information)



Ensuite...

1. Déterminer l'ontologie la plus riche et la plus détaillée en se basant sur le nombre de fils / frères de deux concepts alignés et appartenant à deux partitions mises en correspondance
2. Utiliser les partitions calculées et les correspondances entre elles dans la visualisation des mappings entre les deux ontologies (ZOOM)



Petit test sur deux ontologies réelles

1. Ontologies Géonto (dans sa globalité) - nb concepts == 766)
2. Bout d'ontologie du thème « hydrographie » (provenant de ordonance survey) - nb concepts == 119
3. Alignement entre les deux (merci Nathalie) - nb mappings == 54

Petit résultat

Le bout d'ontologie des Anglais correspond à deux parties dans l'ontologie Géonto : **élément_hydrographique_terrestre** (avec un fort degré) + **infrastructure_de_transport** (faiblement).

