



# Comparaison d'ontologies « distance sémantique entre ontologies »

**Ammar MECHOUCHE**

COGIT - IGN, France

# Objectif

- Exploiter les résultats d'alignement
- Analyser les différences entre deux ontologies
- Dédurre les différences de points de vue
- Aide à la décision (fusion, intégration)
- Recherche d'ontologies sur le Web

# Indices de comparaison recherchés

- Recouvrement thématique
- Indice de niveau de détail / richesse
- Indice de comparaison de structures

# Méthodes proposées

# Recouvrement thématique

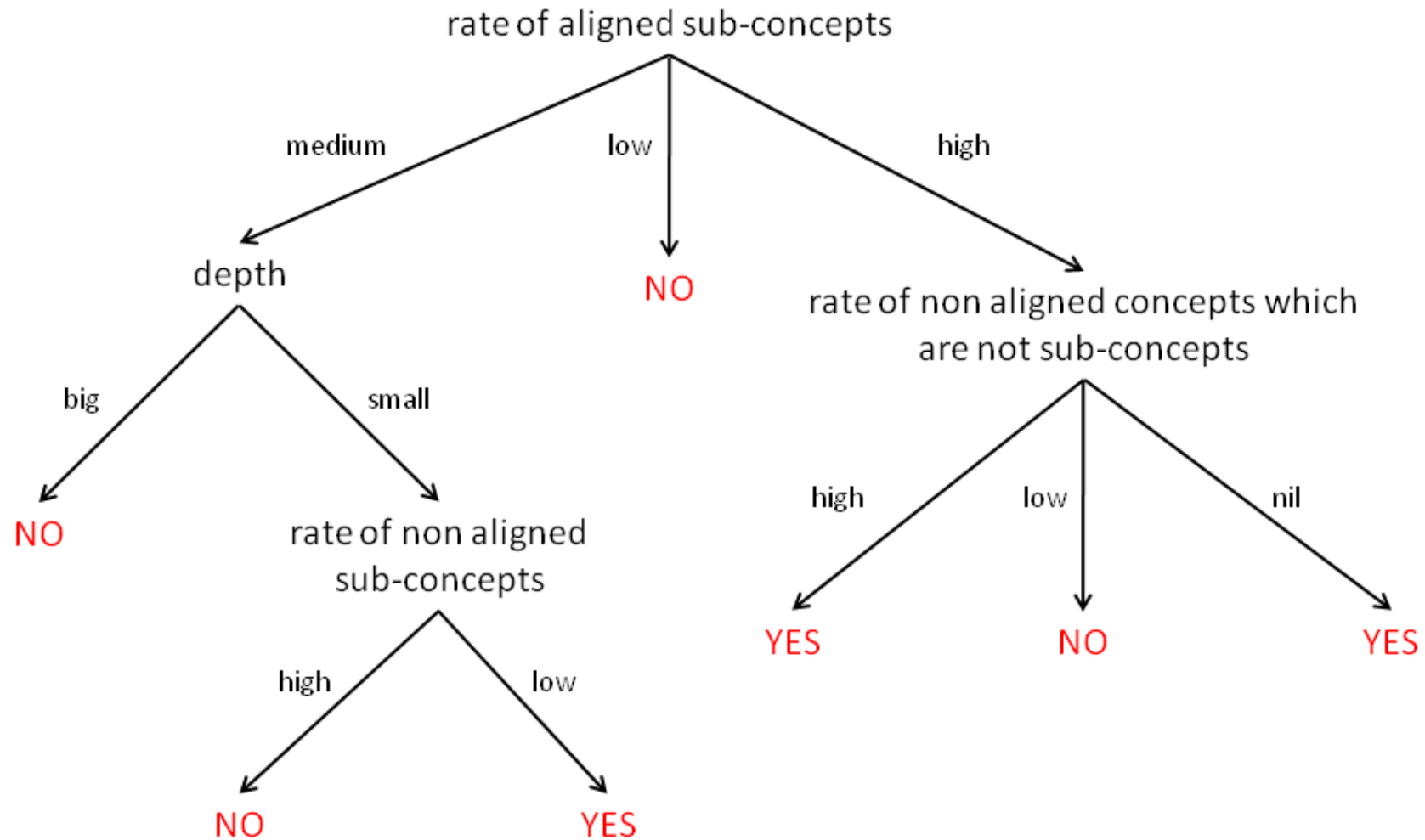
## Calcul des sommets importants

- N'importe quel alignement (Supposition : alignement symétrique)

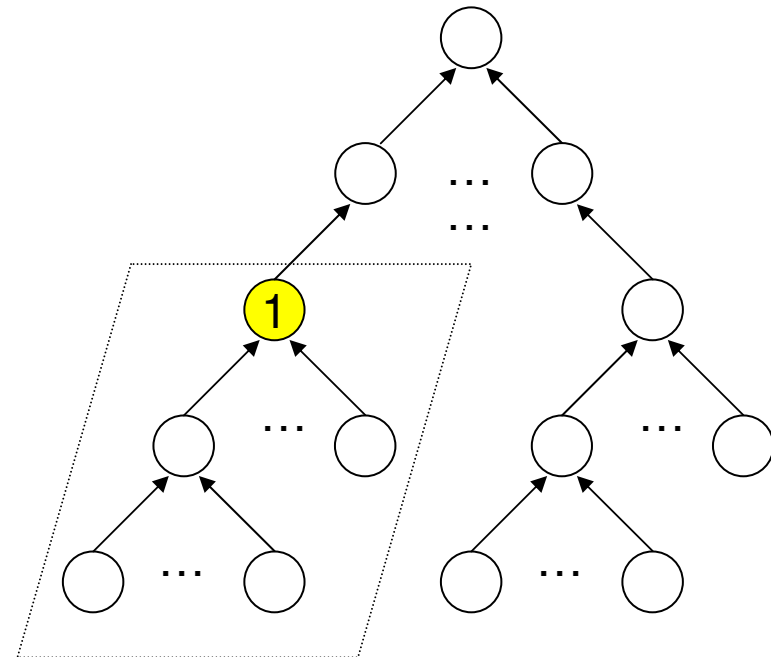
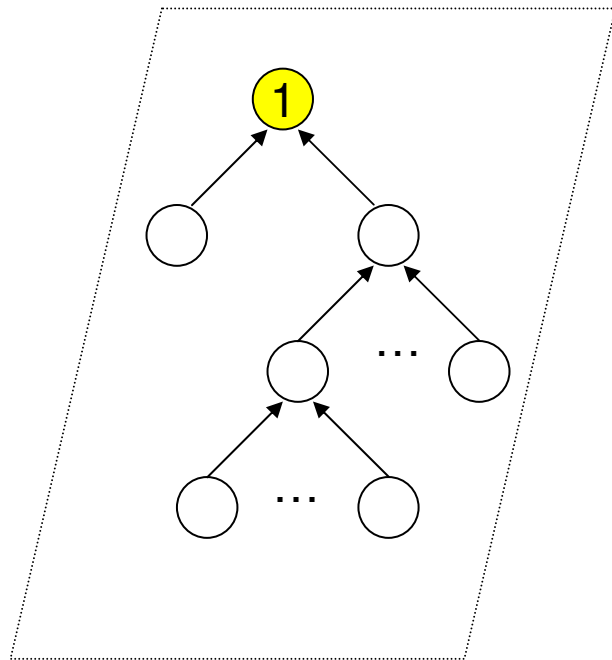
$\langle \textit{Concept}_{cible}, \textit{Concept}_{source}, \textit{Score} \rangle$

- Utilisation d'un arbre de décision

# Détermination des sommets importants



# Détermination des sommets importants



# Méthode basée sur le modèle vectoriel

## Principe :

Calculer la distance entre deux ontologies comme on calcule la distance entre deux documents textuels en recherche d'information.



# Distance entre parties obtenues (MV)

-  $\vec{d}_j (w_{1,j}, w_{2,j}, \dots, w_{n,j})$

-  $w_{i,j} = tf_{i,j} * idf_i$

$$tf_{i,j} = \frac{freq_{i,j}}{MAX_l freq_{i,j}} \quad idf_i = \log \frac{N}{n_i}$$

-  $similarité(d_i, d_j) = \cos(\vec{d}_i, \vec{d}_j) = \frac{\vec{d}_i \cdot \vec{d}_j}{\sqrt{d_i \cdot d_i * d_j \cdot d_j}}$

# Distance entre parties obtenues

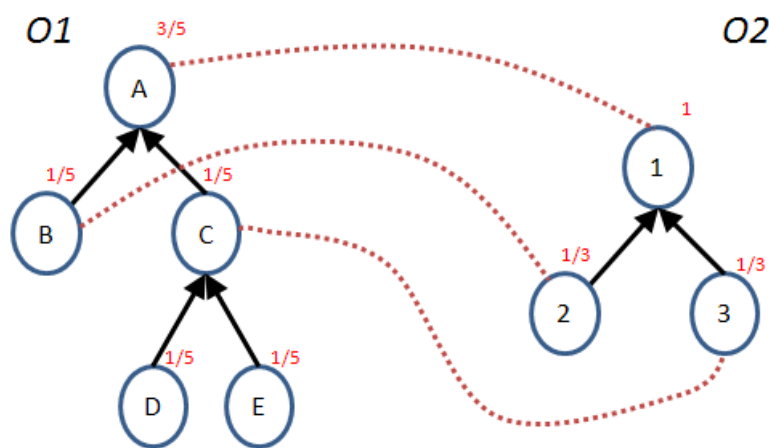
Adaptation du modèle vectoriel pour le calcul de distance entre ontologies : poids des termes

$$tf_{C,P} = \frac{|D \in P \wedge D \subseteq C|}{|P|} \quad idf_C = \begin{cases} \frac{\sum_{\substack{A \subseteq C \\ B \subseteq D}} s_{A,B} \wedge \langle A, B, s_{A,B} \rangle \in Al}{|A \subseteq C|} \\ 1 \text{ si } C \text{ n'est pas aligné} \end{cases}$$

$$w_{C,O} = tf_{C,O} * idf_C$$

# Distance entre parties obtenues

## Adaptation du modèle vectoriel pour le calcul de distance entre ontologies : exemple illustratif



*Alignement* = {<A, 1, 1>, <B, 2, 1>, <C, 3, 1>}

*Vocabulaire* = {A (ou 1), B (ou 2), C (ou 3), D, E}

Taille des vecteurs = 5 + 3 - 3 = 5

$V1 = (3/5, 1/5, 1/5, 1/5, 1/5)$  ;

$V2 = (1, 1/3, 1/3, 0, 0)$ .

Similarité(V1, V2) = 0.92

# Tests sur des ontologies réelles

## Ontologies en question :

- Ontologie IGN : décrit les entités topographiques présentes dans les bases de données de l'Institut Géographique National
- Extrait d'une ontologie de l'hydrographie développée à l'Ordnance Survey

→ le nombre total de mappings trouvé entre les deux ontologies est de 54.

# Tests sur des ontologies réelles

Table 1. Résultats obtenus après la détermination des sommets importants.

Ontologie	Sommet important	Taille	Similarité
Ontologie IGN	infrastructure_de_transport	181	15.83 %
	élément_hydrographique_terrestre	47	
Ordnance Survey	TopographicObject	118	61.02 %

The screenshot displays three panels from an ontology editor:

- Left Panel (SUBCLASS EXPLORER):** Shows the asserted hierarchy for the 'Fusion' project. The class 'infrastructure\_de\_transport' is selected, showing its subclasses: 'canalisation' (with subclasses 'aqueduc', 'conduite\_de\_matière\_première', 'conduite\_forcée'), 'cours\_d\_eau\_artificiel', 'galerie\_d\_amenée\_d\_eau', 'gazoduc', and 'oléoduc'.
- Middle Panel (SUBCLASS EXPLORER):** Shows the asserted hierarchy for 'élément\_hydrographique\_terrestre', with subclasses: 'berge', 'bras' (with 'bras\_de\_décharge' and 'bras\_secondaire'), 'cascade' (with 'chute\_d\_eau'), 'cours\_d\_eau' (with 'fleuve', 'nant', 'rivière'), and 'ruisseau'.
- Right Panel (Class Explorer):** Shows the asserted class hierarchy for 'TopographicObject'. It is a subclass of 'Thing'. Its subclasses include: 'Aber', 'Backwater', 'Bed', 'Bifurcation', 'BoatingLake', 'BraidedRiver', 'BraidedRiverStretch', 'Canal', 'CanalFeeder', 'CanalStretch', 'Castle', 'Channel', 'Ditch', 'Drain' (with 'AgriculturalDrain' and 'Rhyne' as subclasses).

# Limites du MV

**Question** : la distance sémantique calculée avec le MV mesure-t-elle la différence de la structure ou du niveau de détail ?

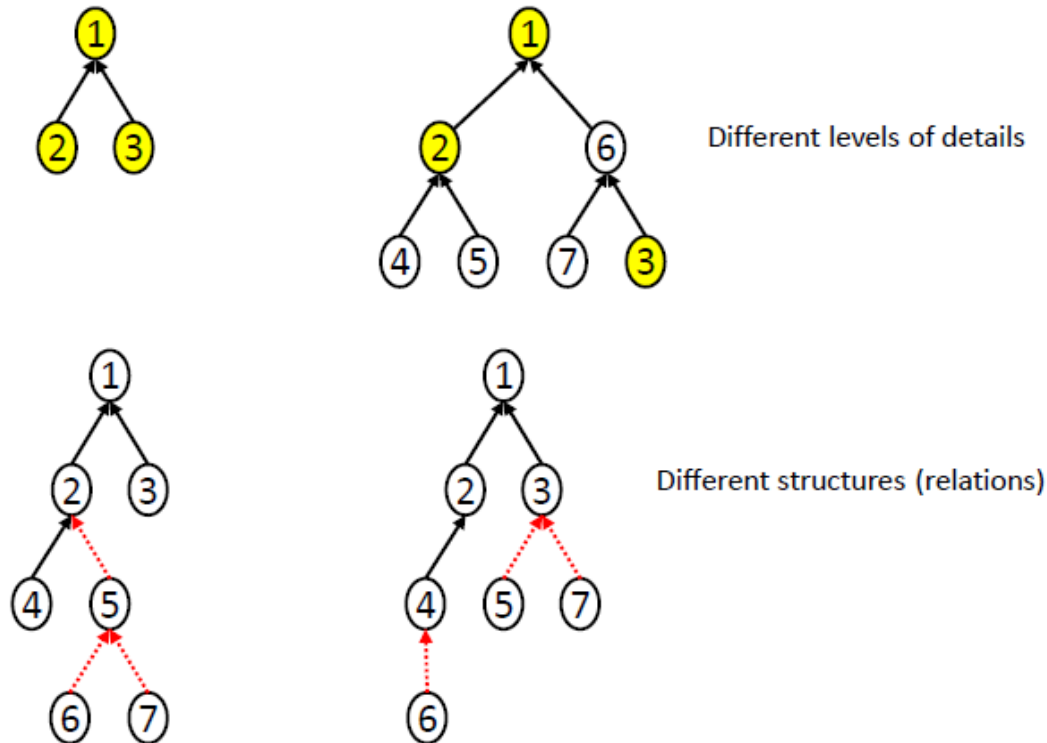
**Réponse** : on ne sait pas vraiment...

→ Nécessité d'une nouvelle méthode

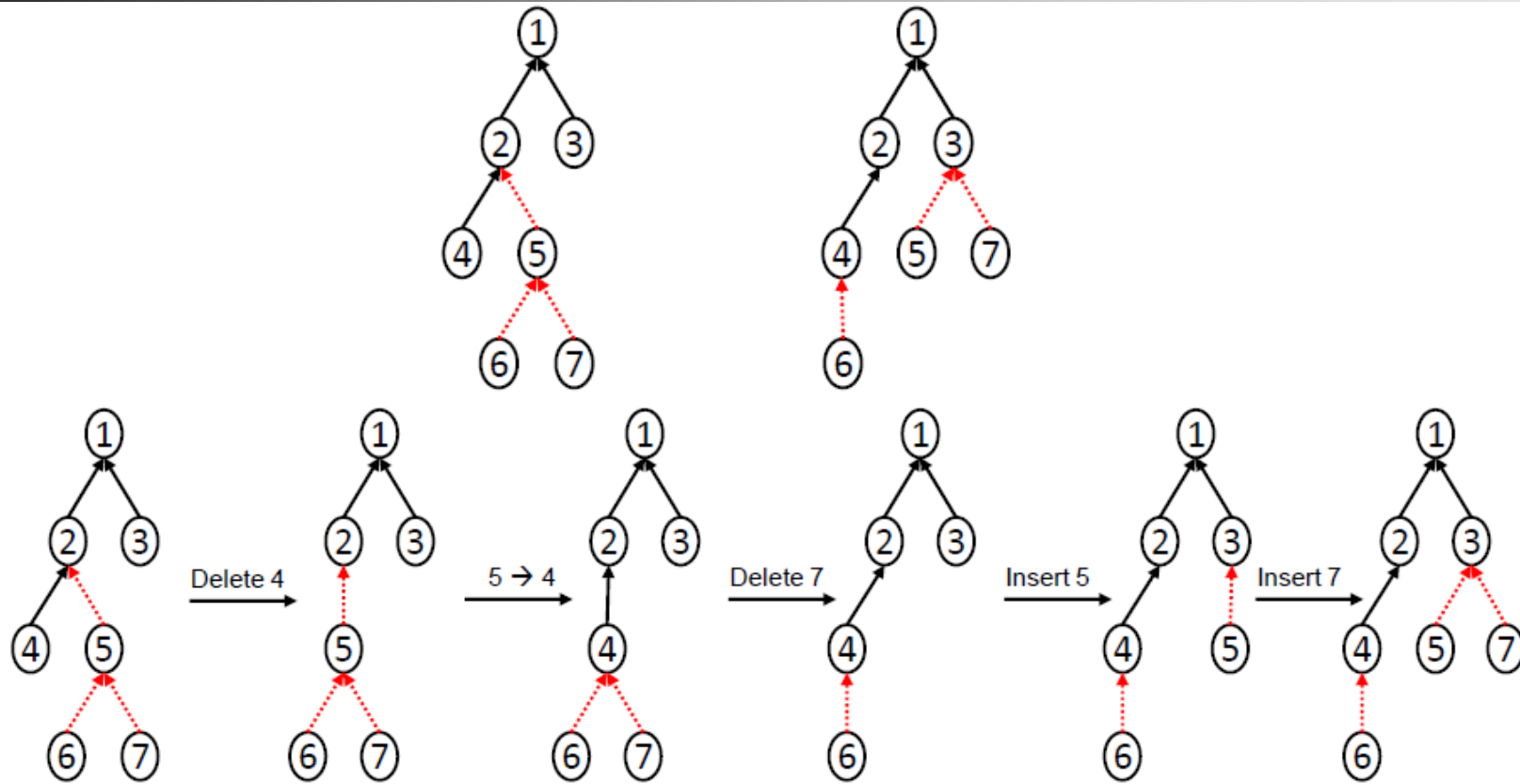
# Nouvelle méthode

## Principe :

- Comparer la différence des structures en utilisant la distance d'édition entre arbres ordonnés
- Calculer le niveau de détail de chaque ontologie



# Distance sémantique entre les structures



Coût de transformation = 5 (coût de suppression, ajout et renommage = 1)

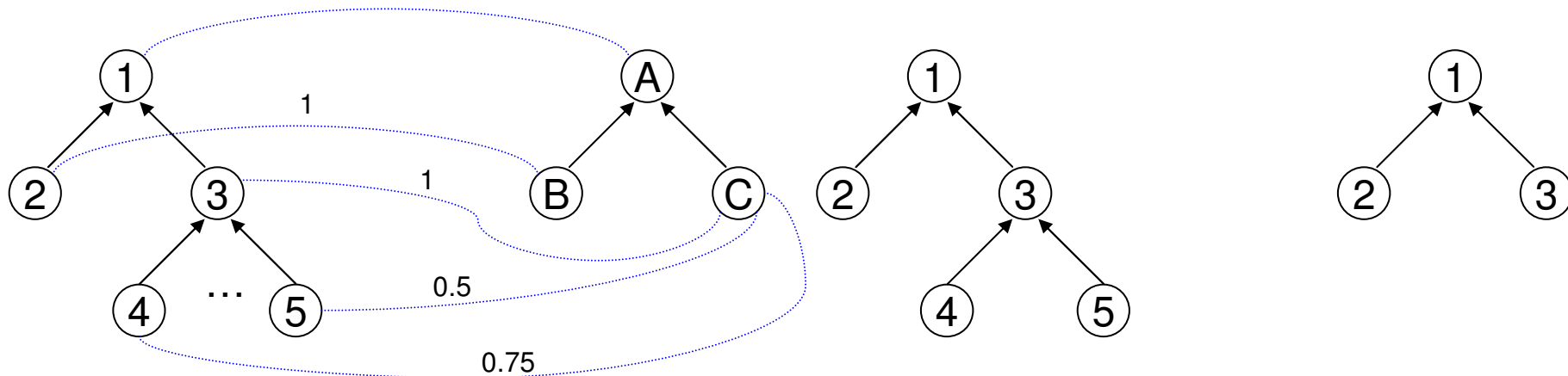
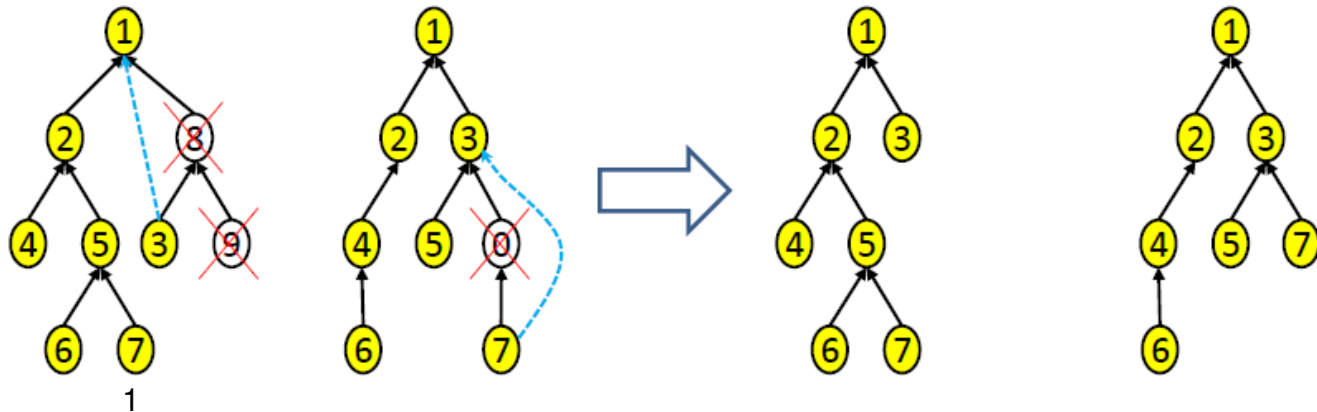
Normalisation / tailles des deux arbres  $\rightarrow$  distance =  $5/14 = 0,36$



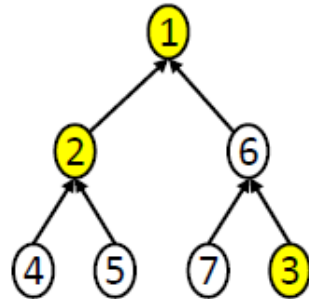
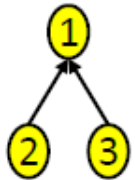
# Distance sémantique entre les structures

## Adaptation aux ontologies

- Garder uniquement les sommets alignés
- Transformer en arbres ordonnés



# Niveau de détail



## Alignement

- <1,1, 0.90>
- <2, 2, 0.80>
- <3, 3, 0.75>

## Niveau de détail (horizontal)

- Ontology 1 :  $(3 + 1 + 1) / 3 = 1.66$
- Ontology 2 :  $(7 + 3 + 1) / 3 = 3.66$
- Indice de confiance =  $(0.90 + 0.80 + 0.75) / 3 = 0.82$

## Niveau de détail (vertical)

- Ontology 1 :  $(1 + 1) / 2 = 1$
- Ontology 2 :  $(1 + 2) / 2 = 1.5$
- Indice de confiance =  $((0.90 + 0.80)/2 + (0.90 + 0.75)/2) / 2 = 0.84$

# Tests (pas encore finalisés)

Ontologies	Size of Ontology	Size of alignment	Main Concepts	Size of MC	Depth of MC	% of aligned sub-concepts
Buildings And Places	681	117	TopographicObject	347	1/6	47,6
Transportation	445		Vehicle	30	1/6	40,8
			TransportationDevice	158		
Buildings And Places	681	45	TopographicObject	347		
EarthRealm	488		OWL:Thing	334		
			TopographicRegion	158		
Buildings And Places	681	89	TopographicObject	347		
			OWL:Thing	334		
Hydrology	153		TopographicObject	119		
			OWL:Thing	34		
EarthRealm	488	49	TopographicRegion	158		
Hydrology	153		OWL:Thing	330		
			TopographicObject	119		
Buildings And Places	681		Place			
Fusion	766		Artificial Topographic Feature	158		
EarthRealm	488		TopographicRegion	158		
			OWL:Thing	330		
Fusion	766		Natural Topographic Entity			
			Relief Feature			
Fusion	766		Inland Hydrographic Feature			
			Transport Infrastructure			
Hydrology	153		TopographicObject	119		
Fusion	766		OWL:Thing			
			Transport Infrastructure			
Transportation	445		OWL:Thing	445		

# Tests (edit distance)

Ontology 1	Ontology 2	Main Concepts Ontology 1	Main Concepts Ontology 2	edit distance
Buildings And Places	Transportation	TopographicObject	TransportationDevice	0,18
		Vehicle	TransportationDevice	0,12
Buildings And Places	EarthRealm	TopographicObject	TopographicRegion	0,26
		OWL:Thing	OWL:Thing	0,17
Buildings And Places	Hydrology	TopographicObject	TopographicObject	0,13
		OWL:Thing	OWL:Thing	0,24
		OWL:Thing	TopographicObject	0,18
		TopographicObject	OWL:Thing	0,3
EarthRealm	Hydrology	TopographicRegion	TopographicObject	0,05
		OWL:Thing	TopographicObject	0,31
Buildings And Places	Fusion	Artificial Topographic Feature	Place	0,3
Fusion	EarthRealm	Natural Topographic Entity	OWL:Thing	0,13
		Natural Topographic Entity	TopographicRegion	0,17
		Relief Feature	TopographicRegion	0,25
		Relief Feature	OWL:Thing	0,11
Fusion	Hydrology	Inland Hydrographic Feature	TopographicObject	0,29
		Transport Infrastructure	TopographicObject	0,26
Fusion	Transportation	OWL:Thing	OWL:Thing	0,35
		Transport Infrastructure	OWL:Thing	0,28

# Tests (ND horizontal)

Ontology 1	Ontology 2	Main Concepts Ontology 1	Main Concepts Ontology 2	LD1	LD2	Confidence
Buildings And Places	Transportation	TopographicObject	TransportationDevice	1,55	4,55	0,69
		Vehicle	TransportationDevice	2,26	3,29	0,8
Buildings And Places	EarthRealm	TopographicObject	TopographicRegion	1,44	1,44	0,82
		OWL:Thing	OWL:Thing	2,14	3	0,71
Buildings And Places	Hydrology	TopographicObject	TopographicObject	14,94	5,17	0,8
		OWL:Thing	OWL:Thing	1,41	1,54	0,78
		OWL:Thing	TopographicObject	1,46	2	0,76
		TopographicObject	OWL:Thing	1,61	1,55	0,75
EarthRealm	Hydrology	TopographicRegion	TopographicObject	1,91	3,26	0,82
		OWL:Thing	TopographicObject	2,5	2,64	0,77
Buildings And Places	Fusion	Artificial Topographic Feature	Place	2,51	2,61	0,79
Fusion	EarthRealm	Natural Topographic Entity	OWL:Thing	1,31	3,25	0,75
		Natural Topographic Entity	TopographicRegion	1,71	1,55	0,89
		Relief Feature	TopographicRegion	1,58	1,29	0,85
		Relief Feature	OWL:Thing	2,16	2,58	0,8
Fusion	Hydrology	Inland Hydrographic Feature	TopographicObject	1,98	2,45	0,86
		Transport Infrastructure	TopographicObject	5,18	3,41	0,84
Fusion	Transportation	OWL:Thing	OWL:Thing	2,34	1,5	0,67
		Transport Infrastructure	OWL:Thing	6,14	3,64	0,74

# Tests (ND Vertical)

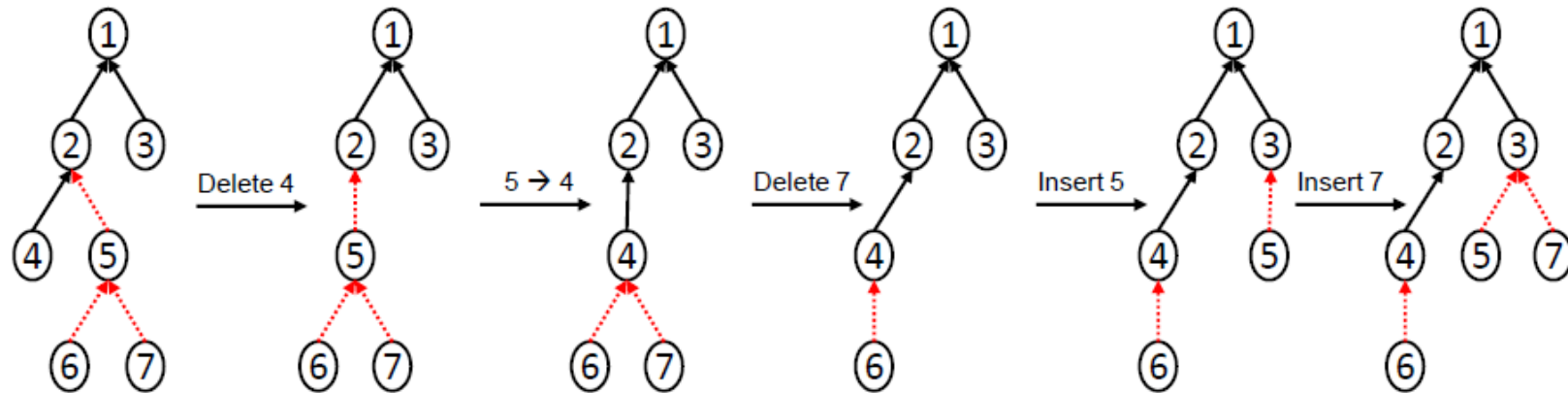
Ontology 1	Ontology 2	Main Concepts Ontology 1	Main Concepts Ontology 2	LD1	LD2	Confidence
Buildings And Places	Transportation	TopographicObject	TransportationDevice	3,33	4,66	
		Vehicle	TransportationDevice	2,42	4,5	
Buildings And Places	EarthRealm	TopographicObject	TopographicRegion	1,42	2,95	
		OWL:Thing	OWL:Thing	2,66	3,88	
Buildings And Places	Hydrology	TopographicObject	TopographicObject	1,54	1,22	
		OWL:Thing	OWL:Thing	1,36	2,6	
		OWL:Thing	TopographicObject	1,77	1,48	
		TopographicObject	OWL:Thing	3,05	1,05	
EarthRealm	Hydrology	TopographicRegion	TopographicObject	3,78	2,00	
		OWL:Thing	TopographicObject	5,04	2,52	
Buildings And Places	Fusion	Artificial Topographic Feature	Place	4,47	2,10	
Fusion	EarthRealm	Natural Topographic Entity	OWL:Thing	2,41	2,66	
		Natural Topographic Entity	TopographicRegion	2,52	2,82	
		Relief Feature	TopographicRegion	3,56	3,18	
		Relief Feature	OWL:Thing	3,64	3,77	
Fusion	Hydrology	Inland Hydrographic Feature	TopographicObject	2,69	2,33	
		Transport Infrastructure	TopographicObject	3,50	2,38	
Fusion	Transportation	OWL:Thing	OWL:Thing	3,84	3,42	
		Transport Infrastructure	OWL:Thing	3,57	3,19	



Merci de votre attention



# Distance sémantique entre les structures



## Coût de transformation

- coût =  $(1 - \text{score}) / \text{profondeur}$  pour chaque opération
- Normalisation / tailles des ontologies uniquement avec les sommets alignés



# Détermination des sommets importants

