# Alignment-Based Measure of the Distance between Potentially Common Parts of Lightweight Ontologies

Ammar Mechouche, Nathalie Abadie and Sébastien Mustière,

Institut Géographique National, Laboratoire Cogit,
73 Avenue de Paris, 94160 Saint-Mandé, France
{Ammar.Mechouche, Nathalie-f.Abadie, Sebastien.Mustiere}@ign.fr

**Abstract.** We propose in this paper a method for measuring the distance between ontologies susceptible to describe a common domain and for assessing the feasibility of their integration. This method is in two steps: the first step determines the potentially common parts of two ontologies, based on a prior alignment carried out between them. The second step computes the distance between these parts with regards to both their levels of detail and their structures, by exploiting the mappings contained in the alignment and adapting the Tree Edit Distance method. We limit our study here to lightweight ontologies, i.e., taxonomies represented in OWL[1], the Ontology Web Language. This method was implemented and applied to real ontologies of the geographic domain. The results obtained so far seem significant.

**Keywords:** Distance between ontologies, Tree Edit Distance, Semantic Web.

## 1 Introduction

Measuring the distance between heterogeneous ontologies is useful for many applications: 1) retrieving ontologies on the web, e.g. finding an ontology to replace another [8], finding ontologies that can enrich other ones, finding people using same ontologies to create new collaborations, etc.; 2) ontology evolution, in order to know to what extent an ontology, especially its structure, has evolved; 3) ontology fusion and data integration, to know in advance if it may be possible to make joint studies on data described by heterogeneous ontologies.

Ontology matching [7], which is addressed by many works and which consists in computing alignments between ontologies, i.e. determining correspondences between semantically related entities from heterogeneous ontologies, is a key idea enabling interoperability in the semantic web. It has brought solutions to some problems like finding ontologies for query translation. However, for the applications cited above we need to compute global similarity measures between ontologies. Indeed, for example, to enrich an ontology from another one, we need to know if they have close structures and if the second ontology is more detailed than the first one. An alignment between two ontologies does not allow knowing if these latter are complementary or not.

[1] http://www.w3.org/TR/owl-features/

In the geographic domain, where data sources are annotated using heterogeneous ontologies [3][5][6][11], we mainly focus on assessing the similarity between ontologies based on three main criteria. The first similarity measure deals with the universes of discourse described by both ontologies: does the source ontology deal with the same domain as the target ontology or does it also provide additional knowledge about a related domain? For example, if we have a domain ontology describing landforms and vegetation, does the source ontology provide us, in addition to that, with knowledge about climate? The second similarity measure aims at comparing the taxonomic structures of both ontologies' common parts. The purpose is to assess whether they result from very different conceptualizations of the domain of interest or not: in other words, would it be difficult to communicate and exchange data with the community who produced this ontology? The third measure compares ontologies' levels of detail to assess whether the source ontology is more or less precise than the target ontology. This aims at automatically determining whether available geodata sources have the appropriate thematic level of detail or not for a specific task. If we are looking for data describing buildings, we may need to make sure that they also explicitly describe more specific buildings such as cabins or huts.

We propose here a new method for computing the distance between potentially common parts of aligned lightweight [17] ontologies. Some works have been dedicated to evaluating the global similarity between ontologies [2], [9], [14-16]. However, their similarity values are difficult to interpret, since they do not measure the difference with regards to particular criteria such as the structure or the level of detail. Our method, however, uses a pre-computed alignment between two ontologies and provides the user with measures with regards to both their structures and levels of detail. This allows to more efficiently assessing differences between ontologies. Moreover, it computes distance between potentially common parts of ontologies, since two ontologies may be similar on one common thematic but quite different on another one. In fact, when source ontologies have common parts with the target ontology, their taxonomic structures must also be compared in order to evaluate to what extent they result from similar conceptualization of the domain of interest. Moreover, the source ontologies' level of detail must be evaluated to assess whether they are more or less precise than the target ontology.

The remainder of this paper is structured as follows: section 2 presents our method for measuring the distance between sub-parts of ontologies. Section 3 presents the results obtained with our method, and section 4 gives some perspectives to our work.

## 2 Proposed Method for Assessing Differences between Ontologies

Our method, first, uses a simple decision tree and the results of an alignment carried out between the two ontologies to be compared, in order to determine the "important concepts" of each ontology, which encompass a large number of mapped concepts. Once the important concepts are determined, the sub-parts of ontologies whose they are roots are compared. The comparison of ontologies (or ontology parts) here consists in computing the distance with respect to their structures and their levels of detail. To do that, we propose a Tree Edit Distance based method to compute the distance between the structures of the compared ontology parts, and we exploit the results of the alignment performed between them to compute their levels of detail.

### 2.1 Determining the Important Concepts of two Aligned Ontologies

Our goal here is to determine the potentially common parts of two ontologies. To do that, we use existing tools to align the ontologies we want to compare and then we determine the parts of each ontology where the mappings are concentrated. Indeed, we believe that if two ontologies describe a common sub-domain then the concepts describing this sub-domain in both ontologies should generally be mapped together. We do not try here to improve the quality of the alignment, which is out of the scope of this paper. We suppose that the quality of the produced alignment is pretty good and the mappings it contains could be transformed into the following form: $< C_{O_1}, D_{O_2}, Score >$, which means that the concept *C* in the ontology *O1* is mapped with the concept *D* in the ontology *O2*, and this mapping has a score confidence *Score* (between 0 and 1).

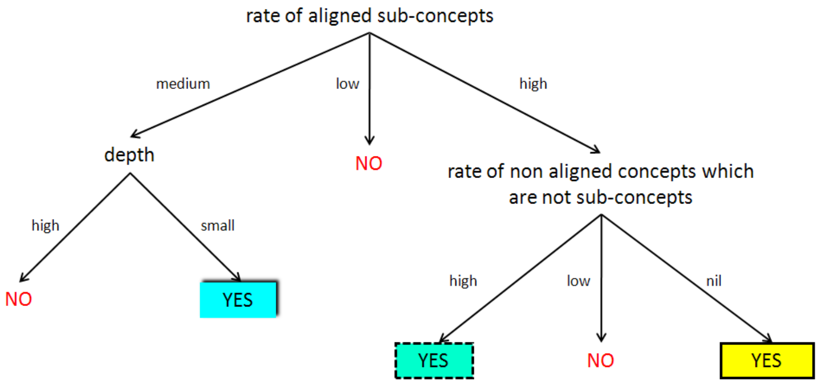


**Fig. 1.** The decision tree determining the important concepts of two aligned ontologies.

Before detailing our method, we consider that a concept is important if it encompasses a large number of mapped sub-concepts or if it encompasses a medium number of mapped sub-concepts but at the same time its depth in the ontology is small. In both cases, the important concepts define sub-parts of the ontology where the mappings are concentrated.

A first set of important concepts is determined thanks to the decision tree depicted in Fig. 1, which classifies each concept of the ontology in the class *YES* if it is an important concept or in the class *NO* if it is not . After that, additional rules allow to filter the first set of important concepts and to keep only the most significant ones**.**

The decision tree proposes three rules allowing determining important concepts:

The first rule (leading to the *YES* in the continuous and bold rectangle) detects the root of each ontology (the *owl:Thing* concept) as the important concept if the number of all mappings is high, since the root is the super class of any concept of the ontology, and it encompasses all mapped concepts. On the examples shown on Fig. 2 the important concepts deduced by this rule are in continuous and bold circles.

The second rule (leading to the *YES* in the dashed rectangle) allows determining the important concepts which encompass a large number of mapped concepts. This rule does not allow to a concept *C* to be important if there are too few remaining concepts in the ontology that are not mapped and not sub-concepts of *C*. In this case it is better to consider the root of the ontology as the important concept. For example, on Fig. 2-(a1) the concept *3* could be an important concept if we suppose that all mapped concepts are its sub-concepts. However, since there remains only one concept (concept *2*) in the ontology which is not its sub-concept it is better to consider only the root as an important concept in order to include this remaining concept. In Fig. 2(b1), however, there are two important concepts deduced by this rule (those with dashed lines). This is due to the fact that: 1) a large number of the mapped concepts are sub-classes of the concept *3* and consequently sub-classes of its parents ; 2) there remains a large number of concepts in the ontology which are not sub-concepts of *3* and which can constitute another part describing another thematic which is not common to the compared ontologies.

The third rule (leading to the *YES* in the shadow rectangle in the decision tree) allows determining the important concepts which encompass a medium number of mapped concepts and which are not deep in the ontology hierarchy. Indeed, in general, in an ontology, the distinction of the different described themes is made generally in its top level. In the example on Fig. 2(c1) the important concepts deduced by this rule are represented with shadow circles (*2*, *3* etc.). The important concepts are determined by this rule when the compared ontologies describe more than one common thematic.

Until now we determine a set of possible important concepts. For example, on Fig. 2(b1) and Fig. 2(c1) we have several important concepts which are deduced and we need to keep only the most significant among them, i.e. the more specific of them. To do this, we defined three additional rules allowing filtering these important concepts. They are tested in the following order, and only one of them is executed:

**Filtering Rule 1:** If we have only one important concept (the root), then keep it as the important concept of the ontology. For example, on Fig. 2(a2) the concept *1* will be the important concept of the ontology; in other terms, the source ontology will be compared as a whole to the determined sub-parts of the target ontology.

**Filtering Rule 2:** If we have important concepts deduced by the second rule of the decision tree (concepts in dashed circles), then we keep only the deepest among them as the important concept of the ontology. Indeed, in an ontology, the deeper the concepts are, the more they share common characteristics. In Fig. 2(b2), the important concept which will be kept is *3* since the Rule 1 will not be activated and the concept *3* is the deepest important concept deduced by the second rule of the decision tree. This way we obtain one important concept encompassing a large number of mapped concepts without including concepts describing other themes.

**Filtering Rule 3:** We keep all the deepest important concepts deduced by the third rule of the decision tree (concepts in shadow circles). For example, in figure Fig.

2(c2), we keep only the important concepts *2* and *3* since they are the deepest ones. These concepts should be roots of two different themes described by the ontology.

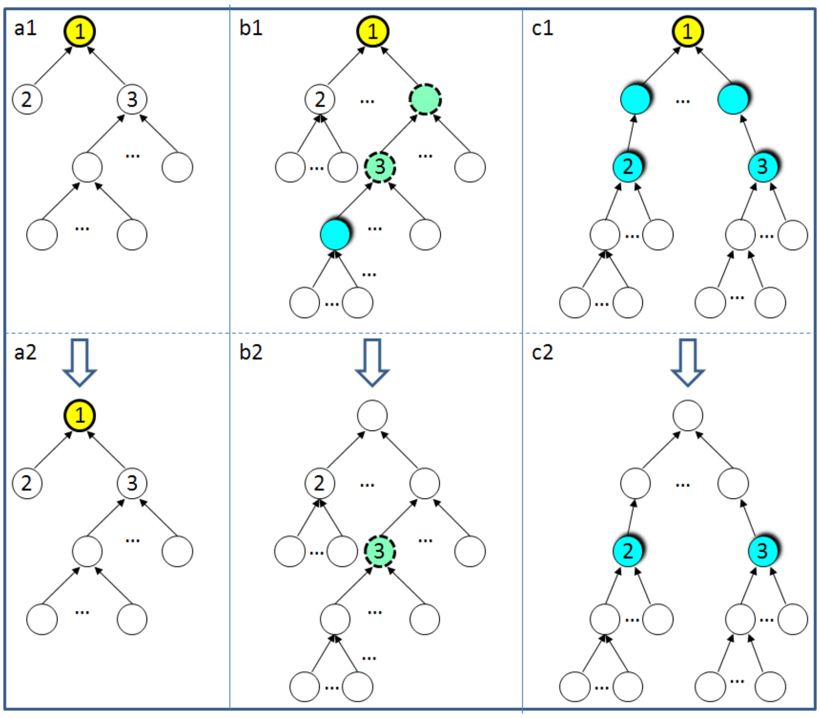


**Fig. 2.** Possible cases for determining the important concepts.

### 2.2 Computing the Distance between the Ontology Parts

We consider here that two aligned ontologies *O1* and *O2* have the same structure if the subsumption relations (or *is-a* relations) between all mapped concepts are preserved in both ontologies. In other terms, the ontologies structures are the same if we do not consider the non mapped concepts. If there is one *is-a* relation between two mapped concepts *C1* and *C2* in *O1*, and this relation does not exists in *O2* between the corresponding concept of *C1* and the corresponding concept of *C2*, then we consider that the structures of *O1* and *O2* are different.

On Fig. 3(1) we have two ontologies with the same structure, but with different levels of detail. In this case, it is interesting to indicate to the user that these ontologies have similar structures, but that one is more detailed than the other one. This allows the user to decide, for example, to enrich the first ontology from the second one or to know if their fusion would be costly or not. In figure Fig. 3(2), however, both ontologies offer the same vocabulary, but their structures are different. Typically, in this case it would be costly to fusion or to combine these ontologies.

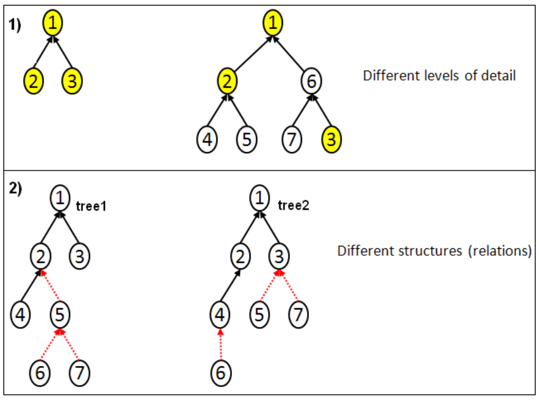


**Fig. 3.** Difference of levels of detail and structures between ontologies.

### 2.2.1 Comparing Ontologies' Structures

To compute the distance between the structures of two ontologies, we propose an adaptation of the Tree Edit Distance method [1], which is usually used to estimate the minimum effort which is necessary to transform an ordered tree into another one. We note that an ordered tree is a tree where the children of every node are ordered. The Tree Edit Distance method returns the minimum cost in terms of the number of operations (node insertion, node deletion and node renaming) which are necessary to transform one ordered tree into another one. Let us consider the example on Fig. 3(2). In order to transform *tree1* into *tree2* we need at least five operations (Fig. 4). So, transforming *tree1* into *tree2* costs 5.

In order to give a sense to the cost returned by the Tree Edit Distance method, we need to normalize it. To do this, we use the normalization formula proposed in [4] (formula (2)), where *NC* is the normalized cost considered as the distance between the two trees, *C* is the value returned by the Tree Edit Distance method, and |*tree 1*| and |*tree 2*| are the respective sizes (number of nodes) of the ordered trees *tree1* and *tree2*.

$$NC = \frac{C}{|tree_1| + |tree_2|} \qquad \textbf{(2)}$$

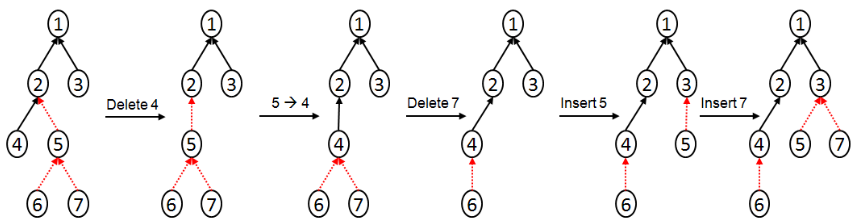


**Fig. 4.** Example of minimum operation number required to transform a tree into another one.

In the example on Fig. 4 the distance between *tree 1* and *tree 2* will be 5/(7+7) = 0.36, which means that similarity between *tree 1* and *tree 2* is 64%. In fact, the smaller the distance is, the closer the structures of ontologies are.

The adaptation of the Tree Edit Distance method is done as follows:

1. Every non mapped concept of each ontology is deleted and its mapped direct sub-concepts become direct sub-concepts of its closest mapped parent. Indeed, we are only interested in the structures formed by mapped concepts in each ontology. In Fig. 5, the concepts *8* and *9* in the ontology on the left are not mapped with concepts from the ontology on the right, so they are deleted and the concept *3* which is mapped becomes a direct child of the concept *1*. The same reasoning is applied to the concept *g*.

2. The next step consists in relabeling the concepts in the second ontology by the labels of their corresponding concepts in the first ontology. If a concept *C* from the second ontology is mapped with one concept *D* from the first ontology, then *C* is renamed to *D*. For example, on Fig. 5, the concept *a* is mapped with the concept *1*, then *a* is renamed to *1*. If a concept *C* of the second ontology is mapped with several concepts from the first ontology, then it takes the name of the corresponding concept having the highest score of similarity. On the Fig. 5, the concept *h* is mapped with the concepts *6* and *7* from the first ontology, however the score of the mapping with *7* is higher, then *h* will be renamed to *7*. If it had been mapped with several concepts from the first ontology having the same score of similarity, then the label of the most general subsuming concept of them would have been used for renaming. If no most general subsuming concept exists, then we would have chosen randomly the label of one of them for renaming. On the Fig. 5, the concept *f* is mapped with the concepts *5* and *6* from the first ontology with the same score, then *f* has to be renamed to *5*, since *5* is more general than *6*. We note that each concept from the first ontology is used at most one time to rename a concept in the second ontology.

3. The last step consists in ordering the obtained trees to allow using existing algorithms and tools for computing the Tree Edit Distance (which is our objective here). On the Fig. 5 we obtain two ordered trees whose distance was computed in the previous example (distance = 0.36).

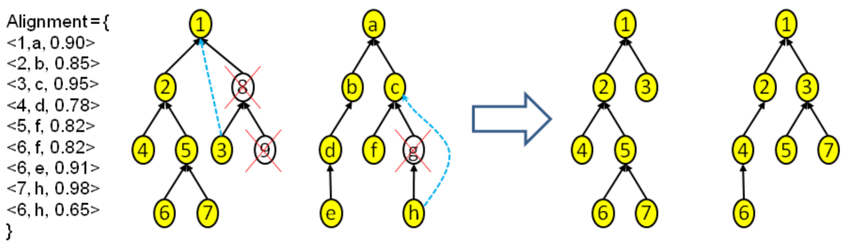


**Fig. 5.** Transforming two aligned ontologies into two ordered trees.

### 2.2.2 Computing the Levels of Detail of two Aligned Ontologies

Two lightweight ontologies having close structures may have different levels of detail. We define here the level of detail of a concept *C* as the number of its sub-classes. For example, on Fig. 3(1) the concept *2* is more detailed in the ontology on the right than in the ontology on the left, since it has two sub-concepts in the ontology on the right when it has no sub-concept in that on the left. Now, we need to have an overall indication allowing to know which of the two aligned ontologies is more detailed. We propose to do this by averaging, for each ontology, the levels of detail' values of its mapped concepts. The following example illustrates this more clearly.
**Illustrating example.** Let us consider the example on Fig. 3(1). To compute the level of detail of each ontology we create two vectors *V1*and *V2*, where *V1* contains the level of detail values of each mapped concept from *O1*, and *V2* contains the level of detail values of the corresponding concept in *O2* of each mapped concept from *O1*. Then, the average value of *V1* gives the level of detail of *O1*, and the average value of *V2* gives the level of detail of *O2*. Thus, we obtain the following vectors:

$$\begin{array}{ccc} \textit{Mapped Concepts} & V1 & V2 \\ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} & \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 6 \\ 2 \\ 0 \end{bmatrix} \\ & LD_{O_1} = 0.66 & LD_{O_2} = 2.66 \end{array}$$

As we can see it, *O2* is more detailed than *O1*.

## 3 Experiments and Results

The proposed methods were implemented and tested on real ontologies. The decision tree, determining the important concepts, was implemented in Java and using the Protégé OWL API[2]. The ontology structures comparison was also implemented as a Java program which uses the Protégé OWL API and reuses a Java implementation of the Tree Edit Distance method for ordered trees, available on the Web[3] and described in [1]. The goal here is to determine which parts of two aligned ontologies are complementary, i.e. which parts are related by a large number of mappings and which have close structures but different levels of detail. We chose to test our methods on five ontologies (ontologies' structures) describing geographic domains or domains close to geography, because our expertise in this domain allows us to better analyze the obtained results. The used ontologies are the followings:

- Building and Places ontology[4]: developed in United Kingdom, its purpose is to describe the building feature and place classes surveyed by Ordnance Survey.
- Transportation ontology[5]: this ontology describes transportation-related information in the CIA World Fact Book[6].

---

2 http://protege.stanford.edu/plugins/owl/api/

3 web.science.mq.edu.au/~swan/howtos/treedistance/

4 http://www.ordnancesurvey.co.uk/ontology/BuildingsAndPlaces/v1.1/BuildingsAndPlaces.owl

5 http://reliant.teknowledge.com/DAML/Transportation.owl

- Earth Realm ontology[7]: elements of this ontology include “atmosphere”, “ocean”, and “solid earth”, and associated subrealms (such as “ocean floor”)[8].
- Hydrology[9]: this ontology is developed by Ordnance Survey to describe in an unambiguous manner the inland hydrology feature classes.
- IGN ontology [12]: it is a bilingual ontology (French / English) which describes the topographic entities present in the geographic databases of the French Mapping Agency (IGN).

These ontologies are first pairwise aligned using the method proposed in the TaxoMap tool [10]. In order to determine the important concepts we defined the semantics of classification criteria used in the decision tree. So, we consider that the rate (percentage) of aligned concepts is high when it is superior to 80%. It is medium when it is comprised between 30% and 80%, and it is low when it is inferior to 30%. A concept of an ontology *O* is considered as deep if its depth in *O* is higher than a half of the depth of *O* itself. We note that, actually, these values are fixed intuitively, however we are working on in order determine them empirically. The obtained results with these values are summarized on Fig. 6.

| Ontologies | Ontology'size | Alignment'size | Main Concepts (MC) | Size of MC | Depth of MC | #Mappings |
|---|---|---|---|---|---|---|
| Buildings And Places | 692 | 117 | Vehice | 30 | 1/6 | 47 |
| Transportation | 445 | | TransportationDevice | 158 | 1/11 | 73 |
| Buildings And Places | 692 | 45 | TopographicObject | 347 | 1/6 | 25 |
| EarthRealm | 561 | | TopographicRegion | 158 | 3/8 | 14 |
| Buildings And Places | 692 | 89 | TopographicObject | 347 | 1/6 | 48 |
| Hydrology | 186 | | TopographicObject | 119 | 1/6 | 47 |
| EarthRealm | 561 | 49 | TopographicRegion | 158 | 3/8 | 20 |
| Hydrology | 186 | | TopographicObject | 119 | 1/6 | 43 |
| Buildings And Places | 692 | 288 | Place | 269 | 2/6 | 256 |
| IGN | 766 | | Artificial Topographic Feature | 609 | 1/8 | 288 |
| EarthRealm | 561 | 80 | TopographicRegion | 158 | 3/8 | 32 |
| IGN | 766 | | Relief Feature | 55 | 2/8 | 46 |
| IGN | 766 | 92 | Inland Hydrographic Feature | 48 | 2/8 | 32 |
| | | | Artificial Topographic Feature | 602 | 1/8 | 41 |
| Hydrology | 186 | | TopographicObject | 119 | 1/6 | 74 |
| IGN | 766 | 98 | Transport Infrastructure | 175 | 2/8 | 69 |
| Transportation | 445 | | OWL:Thing | 445 | 0/11 | 98 |

**Fig. 6.** Important concepts obtained on real geographic ontologies.

We observe on Fig. 6 that most of the used ontologies describe the topography; many important concepts are in relation to the topography, which is a good result since most of the used ontologies really describe the topographic objects. Also, if we look at the size of each determined partition (whose root is an important concept) and the number of its mappings, we deduce that the important concepts detection is pretty precise. For example, the number of mappings between the *Buildings and Places* ontology and the *IGN* ontology equals 288. From one side, there is only one important concept deduced for each ontology (respectively *Place* and *Artificial Topographic Feature*). From the other side, the number of mappings included in the part whose

[6] http://www.daml.org/ontologies/409

[7] http://sweet.jpl.nasa.gov/1.1/earthrealm.owl

[8] http://sweet.jpl.nasa.gov/guide.doc

[9] http://www.ordnancesurvey.co.uk/ontology/Hydrology/v2.0/Hydrology.owl

*Place* is the root represents 89% of the total number of existing mappings, and the part whose *Artificial Topographic Feature* is the root contains all the mappings. So, the mappings are concentrated in the parts determined by our method.

The next step is the comparison phase. In order to obtain significant results we consider only mappings with a score higher than 0.90. We first computed the distance between the ontology parts structures. The obtained results are shown on Fig. 8.

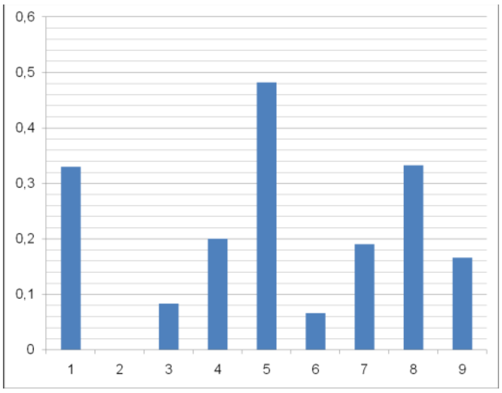


**Fig. 8.** Results of the comparison of the ontology parts structures. The vertical axis indicates the computed Tree Edit Distance measures, and the horizontal axis indicates the compared ontology parts. Indeed, the numbers 1...9 refer to pairs of ontology parts that are compared. See Fig. 9.

Fig. 8 shows that there are some ontology parts that have more similar structures than other ones. For example, the structure of the part of the *Buildings and Places* ontology whose the root is the important concept *Topographic Object* is very close to the structure of the part of the *Hydrology* ontology that has *Topographic Object* as a root. This is due to the fact that both ontologies are produced by the same institution, so with same conceptualization. The structure of the part of the *Buildings and Places* ontology whose the root is the important concept *Place* is, however, different from the structure of the part of the *IGN* ontology that has *Artificial Topographic Feature* as a root. In fact, in the metadata associated with *Building and Places* ontology, it is said that "…*The rationale behind the Buildings and Places module is to provide a minimal set of definitions to maximise the abiliuty to reuse. As a result it contains a* ***shallow hierarchy*** *and minimal property restrictions*". This explains the difference in structures between its parts and parts of the *IGN* ontology which is very structured.

Finally, we compared the levels of detail (LD) of our ontologies using the method presented above. The results obtained are shown on Fig. 9.

| **N° Pair** | **Ontology 1 (O1)** | **Ontology 2 (O2)** | **Important Concepts of O1** | **Important Concepts of O2** | **LD (O1)** | **LD (O2)** |
|---|---|---|---|---|---|---|
| 1 | Buildings And Places | Transportation | Vehice | TransportationDevice | 1 | 1,16 |
| 2 | Buildings And Places | EarthRealm | TopographicObject | TopographicRegion | 1,4 | 1,6 |
| 3 | Buildings And Places | Hydrology | TopographicObject | TopographicObject | 26,52 | 8,10 |
| 4 | EarthRealm | Hydrology | TopographicRegion | TopographicObject | 1,66 | 3,75 |
| 5 | IGN | Buildings And Places | Artificial Topographic Feature | Place | 2,48 | 4,35 |
| 6 | IGN | EarthRealm | Relief Feature | TopographicRegion | 1,71 | 1,21 |
| 7 | IGN | Hydrology | Inland Hydrographic Feature | TopographicObject | 1,93 | 3,12 |
| 8 | | | Artificial Topographic Feature | TopographicObject | 5,86 | 2,43 |
| 9 | IGN | Transportation | Transport Infrastructure | OWL:Thing | 1 | 1 |

**Fig. 9.** Result of the comparison of the ontology parts levels of detail (LD).

The results on the Fig. 9 are significant. For example, the *Hydrology* ontology is more detailed than the *IGN* ontology regarding the hydrographic features, and this can be explained as follows: from one side the metadata associated with the *Hydrology* ontology say that *the scope of this ontology includes permanent topographic features involved in the containment and transport of surface inland water of a size of* ***1*** *meter or greater including tidal water within rivers*. From the other side, we know from the IGN databases specifications that the IGN databases, from which is built the IGN ontology, include only water surfaces larger than **7.5** meter. Another notable difference of levels of detail is between the *Topographic Object* part of the *Hydrology* ontology and the *Topographic Object* part of the *Buildings and Places* ontology. This result combined with the previous one telling us that the structures of these parts are very similar shows that these ontology parts are complementary and may help the user or a program to decide for example to enrich (for example thanks to an importation operation) the *Topographic Object* part of the *Hydrology* ontology from the *Topographic Object* of the *Buildings and Places* ontology.

## 4 Conclusion and Perspectives

Means for evaluating distance between ontologies seems to us important for decision making systems in the context of data integration, ontology fusion, ontology evolution and ontology retrieval on the web.

We have presented in this paper a new method for measuring the distance between lightweight ontologies. Our method differs from existing methods in several ways: 1) it exploits alignments between ontologies rather than assuming that ontologies share exactly the same vocabulary; 2) it does not compare whole ontologies but only the potentially common parts of them determined by our decision tree, in order to more efficiently assess the differences between ontologies; 3) finally, our method provides indications about the level of detail of each ontology and computes a distance between the ontologies structures by adapting the Tree Edit Distance method, which was not used in the past in this context to the best of our knowledge. The proposed method is implemented and tested on several real geographic ontologies, and the results obtained so far seem significant.

In the future, we plan to improve our method with respect to several aspects: 1) the ontology parts are determined with our decision tree which consists in detecting ontology parts where the mappings are concentrated, it would be interesting to compare and combine our decision tree with existing methods for ontology partitioning [13], in order to obtain better partitions. Moreover, our decision tree may be learnt using existing algorithms like the ID3[10] algorithm, particularly in order to automatically determine the semantics associated with each classification criteria; 2) our comparison method is actually restricted to lightweight ontologies, we plan to extend it to heavyweight ontologies in order include in our comparison procedures more complex constraints and relations between concepts modeled in domain ontologies; 3) another perspective of this work is to compare our method to similar ones [18] and to study the influence of different matching techniques on our distance measure; 4) finally, we plan to integrate to our method other information to better

[10] http://en.wikipedia.org/wiki/ID3_algorithm

understand differences between ontologies, like metadata associated with ontologies, ontology utilization purposes, etc.

## Acknowledgement

This work is partly founded by the French Research Agency through the GeOnto project ANR-O7-MDCO-005 on “creation, alignment, comparison and use of geographic ontologies”.

## References

1. Zhang K., Shasha D.: Fast algorithms for the unit cost editing distance between trees. Journal of Algorithms, 11(4), pp. 581-621, (1990)
2. Maedche A., Staab S.: Measuring Similarity between Ontologies. EKAW, 251-63, (2002)
3. Rodriguez M.A, Egenhofer M.J.: Determining Semantic Similarity Among Entity Classes from Different Ontologies. IEEE TKDE, 15(2), p. 442-56. (2003)
4. Sager T., Bernstein A., Pinzger M., Kiefer C.: Detecting similar Java classes using tree algorithms. International Workshop on Mining Software Repositories, 65-71, (2006)
5. Paul, M., Ghosh S.K.: An Approach for Service Oriented Discovery and Retrieval of Spatial Data, Procedings International Workshop on Service Oriented Software Engineering (IW-SOSE’06), Shangay, China, pp 84—94, (2006)
6. Nambiar U., Ludäscher B., Lin K. and Baru, C.: The GEON portal: accelerating knowledge discovery in the geosciences, 8th ACM International Workshop on Web Information and Data Management, Aarlington, pp 83—90, (2006)
7. Euzenat J. and Shvaiko P.: Ontology Matching. Springer Verlag, p. 333, (2007)
8. Jérôme David, Jérôme Euzenat: Comparison between Ontology Distances (Preliminary Results). International Semantic Web Conference, 245-260,(2008)
9. Wang J.Z., Ali F. and Srimani P.K: An Efficient Method to Measure the Semantic Similarity of Ontologies. Advances in Grid and Pervasive Computing: p. 447-458, (2008)
10. Hamdi F., Zargavouna H., Safar, B. and Reynaud C.: TaxoMap in the OAEI 2008 alignment contest. Ontology Matching, p. 206-213, (2008)
11. Klien E.M.: Semantic Annotation of Geographic Information. Phd thesis, Institute for Geoinformatics, University of Muenster. Muenster, Germany, (2008)
12. Abadie N. and Mustière S.: Constitution d’une taxonomie géographique à partir des spécifications de bases de données. Actes de SAGEO, Montpellier, (2008)
13. Stuckenschmidt H., Parent C. and Stefano Spaccapietra (Eds.): Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization. LNCS Vol. 5445 Springer, ISBN 978-3-642-01906-7, (2009)
14. Euzenat J., Allocca C., David J., d'Aquin M., Le Duc C. and Sváb-Zamazal, O.: Ontology distances for contextualisation. Report 3.3.4, IST NeOn IP, NeOn, (2009)
15. Ngan L., Soong A. and Hung L.: Comparing two ontologies. International Journal on Web Engineering and Technologies, 5(1): p. 48-68, (2009)
16. Mechouche A., Abadie N., Mustière S. : Mesure de la distance sémantique entre parties potentiellement communes à deux taxonomies, *IC*, France, (2010)
17. Óscar Corcho: Ontology based document annotation: trends and open research problems. IJMSO 1(1): 47-57 (2006)
18. P. Wang, B. Xu. Lily: Ontology Alignment Results for OAEI 2009. OM (2009)