A Semi-automatic Mapping Selection in the Ontology Alignment Process

Hafedh Nefzi[1], Mohamed Farah[1], Imed Riadh Farah[1,2] and Basel Solaiman[2]
[1]*RIADI Laboratory, University of Manouba, Manouba, Tunisia*
[2]*ITI Laboratory, TELECOM-Bretagne, Brest, France*

Keywords: Ontology, Remote Sensing, Similarity Models and Measures, Alignment, Enrichment, UTA.

Abstract: Ontologies are considered as one of the most powerful tools for knowledge representation and reasoning. Thus, they are considered as a fundamental support for image annotation, indexing and retrieval. In order to build a remote sensing satellite image ontology that models the geographic objects that we find in a scene, their characteristics as well as their relationships, we propose to reuse existing geographic ontologies to enrich an ontological core. Reusing high quality resources (called source ontologies) helps ensuring a good quality for the extracted knowledge, and alleviating the conceptualization stage, i.e. avoiding building a new ontology from scratch. Ontology alignment is an important phase within the enrichment process. It is a process that allows discovering mappings between core and source ontologies, where each mapping is a couple of entities brought from each ontology and linked together either by an equivalence or a subsumption relationship. Such relationships are based on various similarity measures. In this paper, we first present a brief literature review of existing theoretical frameworks for similarity measures, then we describe a new alignment approach based on a semi-automatic mapping selection process that needs little human intervention. First experiments show the benefit from using the proposed approach.

1 INTRODUCTION

Remote sensing plays a very important role in the geographic information acquisition and interpretation, especially in order to study major facts affecting earth, such as the urbanism expansion or the vegetation and water resources evolution. They are also considered as a fundamental support to detect and monitor natural phenomena such as erosion, flooding, deforestation, desertification, etc.

The continuous technological progress in remote sensing has led to a phenomenal increase of multi-resolution, multi-spectral and multi-temporal satellite images. Thus experts are no more able to deal with such big data. Researchers as well as practitioners definitely agree that processing such big data can only be dealt with (semi)automatically.

Nowadays, ontologies are considered as one of the most powerful tools for knowledge representation and reasoning. An ontology that models the geographic objects, their characteristics and relationships can be a good support for automatic annotation, indexing and retrieval of satellite images content.

In the literature, geographic, spatial as well as remote sensing satellite image ontologies belongs to two main categories: thematic geographic ontologies (agriculture, hydrology, planning, surveying, etc.) and spatial ontologies modelling spatial concepts for cartographic representation purposes (Parent et al., 1998) and (Gesbert, 2005).

One of the most popular spatial ontologies is that of Durand (Durand et al., 2007) which is constructed to model and interpret urban vegetation in satellite images. Each concept of this ontology is characterized by contextual attributes as well as a priori values of spatial and spectral attributes (spectral signatures, indices, etc.). In (Hudelot et al., 2006), we find a spatial ontology that models topological, distance and directional relations between objects within a scene. *SatellitesSceneOntology* developed within the DAFOE platform (Charlet et al., 2010) is also a spatial ontology that contains concepts coming mainly from the *Corine Land Cover database* and related to agricultural areas, land use, buildings, vegetation, water areas, etc. In addition, it handles topological, distance and directional relations.

Concerning thematic geographic ontologies, we mention the hydrographic ontology, named *Hydrol-*

Nefzi H., Farah M., Riadh Farah I. and Solaiman B..
A Semi-automatic Mapping Selection in the Ontology Alignment Process.
DOI: 10.5220/0005162204590466
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2014), pages 459-466
ISBN: 978-989-758-049-9
Copyright © 2014 SCITEPRESS (Science and Technology Publications, Lda.)

ogy, which covers topographical features involved in the retention and transport of inner and surface water such as slopes, roads, coastlines, floodplains, hills, artificial or natural bodies of water, etc. The *FusionTopoCarto2* ontology takes its concepts from *BDTopo* and *BDCarto* databases.The concepts of this ontology are divided into two main categories: artificial topographic concepts and natural geographic concepts. It offers a wide number of concepts related to hydrography, topography and agriculture. The *OntoGeo* ontology covers concepts related to various geographical objects that may exist in a satellite scene, as well as events that may occur and change the status of these objects.

Studying and analysing these geographic ontologies and many others let us to conclude that none of these ontologies entirely covers satellite image features, but at the same time they are somehow complementary. In (Nefzi et al., 2013), authors propose to build a new remote sensing image ontology reusing the available ontological resources such as the aforementioned ones. This enrichment process has proven to guarantee a high quality satellite image ontology which encompasses the knowledge related to geographic objects, their relationships and their spatial, spectral and contextual characteristics.

Reusing of existing ontological resources helps to alleviate the conceptualization stage by avoiding intensive text processing and datamining activities on textual corpora for the extraction of relevant keywords and their grouping into concepts. In addition, it ensures a high quality of the extracted knowledge since these ontological resources are usually constructed and validated by domain experts.

Ontology alignment is an important phase within the enrichment process. It is a process that allows discovering mappings between core and source ontologies, where each mapping is a couple of entities brought from each ontology and linked together either by an equivalence or a subsumption relationship. Such relationships are based on various similarity measures.

In this paper, we propose a new alignment approach based on a semi-automatic mapping selection process that needs little human intervention. In fact, given a reduced set of mappings having sufficiently high similarity scores, the expert is asked to rank them by decreasing order of relevance. We then infer an utility-based aggregation model of the similarity measures which is as close as possible to the expert feedback. This model can therefore be used to rank all the remaining mappings which do not appear in the reduced set.

The rest of the paper is organized as follows: in section 2, we present a brief literature review of existing theoretical frameworks for similarity measures. Section 3 describes the semi-automatic mapping selection approach that we propose for the alignment of different ontologies for the enrichment of a remote sensing core ontology. In Section 4 we present results of first experimentations. Finally, we conclude in Section 5.

2 SIMILARITY MODELS

Similarity measures have a long tradition in many fields such as information retrieval, artificial intelligence, and cognitive science. They have also become popular in semantic geospatial web (Egenhofer, 2002) and they are being applied to compare concepts, to improve searching and browsing through ontologies, as well as for matching and aligning ontologies (Shvaiko and Euzenat, 2008).

Given two sets E_i and E_j, a similarity measure between two elements or objects e_i and e_j from E_i and E_j respectively, is defined as a real valued function $Sim : (E_i \times E_j) \rightarrow [0,1]$ where 0 means that both objects are dissimilar and 1 tells that they are identical.

2.1 Geometric Models

In geometric models, objects are represented as vectors in a multi-dimensional vector space, where each dimension is a property or an attribute of the object. The range of each dimension represents all possible values of the property. In these models, the definition of the similarity between two objects is obtained by their internal contents. The similarity measure is seen as inversely proportional to the spatial distance. The most commonly used similarity measure is based on the Minkowski dissimilarity metric :

$$Dissim_{Minkowsky}(e_i, e_j) = \left[\sum_{k=1}^{n} \mid e_i^k - e_j^k \mid^r \right]^{1/r}$$

where n is the number of dimensions, and r is a parameter used to indicate the distance kind, such as the Manhattan distance *(r = 1)* or the euclidean distance *(r = 2)* (Schlicker et al., 2006).

There are many other similarity measures in the geometric model which are reported in the literature such as the Jaccard distance, the cosine distance, the Dice distance, or the Gardenfors distance (Gardenfors, 2000; Schwering and Raubal, 2005).

According to (Tversky, 1977), the main limitations of similarity measures from the geometric models are that they have difficulties in describing the objects that have a large number of features and that the similarity is only based on common characteristics.

2.2 Feature Models

As the geometric model, feature model uses properties, also called features, to describe the objects to be compared. The features correspond to properties, either concrete or abstract which can represent nominal, ordinal, interval and ratio scaled variables. This model is based on sets theory and takes into account features that are common to both the objects to be compared as well as the differentiating features which are specific to each one. Thus this model differs from the previous one as the properties are boolean rather than fuzzy, i.e. features either hold or do not hold for an object. For instance, the *forest* concept can be represented by the *shape*, *area*, *vegetation*, *occupation* and *relief* features. The vegetation component can have as values *low vegetation*, *mediumvegetation* or *high vegetation*. Similarly, the occupation score ranges from 0 to 100. Consequently, for an *evergreen forest* concept having as values for both features *vegetation* and *occupation high vegetation* and *80* in the geometric model, will have both features set to true in the feature model.

There are many similarity measures of the feature model, many of which are based on the Tversky measure:

$$Sim_{Tversky}\left(e_i,e_j\right) = F\left(S_i \cap S_j, S_i \setminus S_j, S_j \setminus S_i\right)$$

where e_i and e_j are the objects to be compared, S_i and S_j their corresponding sets of features, and F is a linear function. Depending of the definition of F, we find different similarity measures such as the contrast measure as described by the following formula:

$$Sim_C\left(e_i,e_j\right) = \theta \cdot f\left(A \cap B\right) + \alpha \cdot f\left(A \setminus B\right) + \beta \cdot f\left(B \setminus A\right)$$

where θ, α and β are positive constants and f captures the saliency of a feature set. For instance, the Lesak measure is a specific case of the contrast measure.

We have also the ratio measure which is given by:

$$Sim_R\left(e_i,e_j\right) = \frac{f\left(A \cap B\right)}{f\left(A \cap B\right) + \alpha \cdot f\left(A \setminus B\right) + \beta \cdot f\left(B \setminus A\right)}$$

The Jaccard measure can be seen as a specific case of the ratio measure. Another feature model calculates similarity by taking the ratio of common to distinctive features (Sjoberg, 1972). Another well known similarity measure of the feature model is the Hamming distance (Hamming, 1950).

2.3 Network Models

Network models are based on graph theory where knowledge is modelled using semantic networks where nodes represent units of knowledge, such as objects, concepts or properties, and edges connect nodes according to specific relations such as synonymy, antinomy, or subsumption. Most similarity distances are computed using standard graph-theoretic algorithms such as the shortest path. Network models are divided into three categories.

The first category consists of graph-theoretic measures which are mainly based on computing the number of edges that separate two nodes in a taxonomy. The most commonly used measures in the literature are those of Rada, Wu & Palmer , Leacock & Chodorow and Hirst & St-Onge.

The second category consists of information theoretic measures, called entropies. Each measure of this category is based on the computation of the so called Information Content *IC* of a concept *e*. For instance, the Resnik measure computes the similarity between two concepts e_i and e_j by computing the Information Content of their Least Common Subsumer *LCS* in the semantic graph (Resnik, 1995; Resnik, 1999). The Lin's Measure is a normalized version of Resnik measure by dividing $IC(LCS(e_i,e_j))$ by $IC(e_i)+IC(e_j)$ (Lin, 1998). We have many other variations of Lin's measure such as the Schlicker measure (Schlicker et al., 2006) and the GraSM measure (Couto and Coutinho, 2005).

The third category, also called hybrid approaches, combine graph based and information theoretic measures. In this category, we find the Jiang and Conrath measure (Jiang, 1997) as well as the Leacock and Chodorow measure (Leacock, 1998).

2.4 Alignment based Models

Measures of these models are well suited for comparing objects that are richly structured rather than just being a collection of unstructured features. Thus, comparing objects involves not only matching features, but also determining which elements correspond to, or align with, one another. Matching features are aligned to the extent that they play similar roles within their objects.'Matching features influence similarity more if they belong to parts that are placed in correspondence, and parts tend to be placed in correspondence if they have many features in common and if they are consistent with other emerging correspondences' (Goldstone, 1994; Goldstone, 1999; Markman, 1993). Alignment based measures are deeply inspired from work on analogical reasoning (Gentner, 1989; Holyoak and Thagard, 1989).

2.5 Transformational Models

In previous models, similarity measures are based on features and/or relations describing the concepts to be compared. In transformational models, a similarity measure of two objects e_i and e_j is rather viewed as the number of transformations required to alter one concept into another (Imai, 1977; Wiener Ehrlish et al., 1980).

Transformations can be viewed as terminological operations needed to modify concepts, such as mirroring, reversing and adding, as well as geometric operations such as rotating, reflecting, translating and dilating. In (Goldstone, 2005; Hahn et al., 2009), the similarity between two concepts corresponds to the complexity of the algorithm that transforms the representation of a specific concept to another.

One of the most well used similarity measures of this model, we mention the Levenshtein distance also called the Edit distance as well as the Jaro measure.

3 PROPOSED APPROACH

In this section, we give details on the alignment approach that we propose which is based on a semi-automatic mapping selection process that needs little human intervention.

Remember that we are focusing on a specific but critical process to build remote sensing image ontology by reusing a set of already existing remote sensing domain related ontologies. We choose one of them as a core or target ontology to be enriched with the remaining ones that we call source ontologies. Thus, the target ontology can be extended with new concepts modelling different objects that may exist in a satellite scene. Candidate ontologies, even though they do not take into account the spatial and spectral characteristics of objects in a standard remote sensing image, they model with different levels of granularity a variety of useful geographical objects covering several areas such as topography, hydrography, urban planning, etc.

The alignment process is based on selecting couples of concepts (also called mappings) $(c_i, c_j) \in O_i \times O_j$ where O_i is the core ontology and O_j is a source ontology. If c_i and c_j are sufficiently similar to each other, they can be considered in a further process in the enrichment process.

It is obvious that the quality of the resulting ontology depends on the quality of the selected mappings. Actually, there are two ways to evaluate the quality of a mapping: either automatically or manually. Automatic based approaches for the quality assessment of a mapping use a real valued function f that is supposed to capture the overall similarity between the concepts c_i and c_j to be aligned, i.e. $f(c_i, c_j)$ is an aggregation of a set of similarity measures presented in the previous section. The advantage of using automatic approaches is their scalability, i.e. their ability to handle large ontologies but they suffer from a disadvantage which is how to aggregate the various similarity measures.

Manual based approaches for the quality assessment of a mapping need the intervention of an expert. This way ensures the quality of the mapping since it is human based but cannot be considered in real life due to the size of the ontologies to be aligned.

We propose a semi-automatic mapping selection approach which combines the preceding ones. More specifically, we propose to select a reduced set of mappings and ask an expert to build a ranking of them in decreasing order of quality. Next, we automatically build an aggregation procedure of a set a similarity measures that tries to reproduce as much as possible the ranking of the expert. The last step consists of using the aggregation procedure to automatically rank all the remaining candidate mappings.

The proposed method finds its roots in the UTA method (Jacquet-Lagrèze and Siskos, 1982; Jacquet-Lagrèze et al., 1987) for ordinal regression and which uses both the information given by expert judgements, in a form of a ranking, on a reduced set of pairs of concepts to be aligned, as well as the evaluation of similarities between the involved concepts, to infer the parameters of a ranking model that it is as consistent as possible with the expert judgements.

Our method assumes that there exists a non-decreasing *marginal utility function* u_i corresponding to each similarity measure sim_i as well as an *additive utility function* U (Keeney and Raiffa, 1976) that encompasses the ranking model.

More formally, let $\widetilde{R} = \{a_1, a_2, \ldots, a_n\}$ denotes the top-n alignments having the best similarity scores according to a family $F = \{sim_1, sim_2, \ldots, sim_p\}$ of p similarity measures. F is supposed to satisfy consistency conditions (Roy, 1991), i.e. completeness (all relevant similarity measures are considered), monotonicity (increasing the evaluation of an alignment on some similarity measure leads to increasing its relevance to be considered), and non-redundancy (no superfluous similarity measures are considered). Let $sim(a_j) = [sim_1(a_j), \ldots, sim_p(a_j)]$ be the multicriteria evaluation vector of alignment a_j. We assume, without loss of generality, that the greater is $sim_i(a_j)$, the better is alignment a_j on similarity measure sim_i.

The ranking model U can therefore be written as follows:

$$U(a_j) = \sum_{i=1}^{p} u_i(sim_i(a_j))$$

Let $sim_{i*} = \min_j\{sim_i(a_j)\}$, $sim_i^* = \max_j\{sim_i(a_j)\}$ be respectively the worst and the best (finite) evaluations on sim_i. $E_i = [sim_{i*}, sim_i^*]$ is the scale of similarity measure sim_i, i.e. the range in which the values of similarity measure sim_j are found. Consequently, the evaluation space is $E = \prod_{sim_i \in F} E_i$ and $sim(a_j) \in E$ is a profile in such space E.

Expert judgements are supposed to be given in the form of a ranking that can be modelled using 2 global preference relations: an indifference relation I and a strict preference relation P. Therefore, the following generally holds for U:

$$\begin{aligned} a_j\, P\, a_k &\Leftrightarrow U(a_j) > U(a_k) \\ &\Leftrightarrow \sum_{i=1}^{p} u_i(sim_i(a_j)) > \sum_{i=1}^{p} u_i(sim_i(a_k)) \\ a_j\, I\, a_k &\Leftrightarrow U(a_j) = U(a_k) \\ &\Leftrightarrow \sum_{i=1}^{p} u_i(sim_i(a_j)) = \sum_{i=1}^{p} u_i(sim_i(a_k)) \end{aligned}$$

The subjective ranking is therefore a *complete preorder* $S = (P, I)$ on a reduced subset $\widetilde{R} \subset R$ of alignments with multicriteria evaluations on E. I and P are respectively the symmetric and asymmetric parts of this preorder that we call hereafter the *reference preorder*.

As in the UTA method, we consider that for each similarity measure sim_i, the corresponding marginal utility function u_i is estimated by a piecewise linear function. Thus, the range E_i is divided into $\alpha_i \geq 1$ equal sub-intervals $[sim_i^k, sim_i^{k+1}], \forall k = 1, \ldots, (\alpha_i - 1)$ where α_i is a parameter. If E_i is discrete, α_i can be set to the number of grades of the interval E_i or a subset of these grades. Therefore, each end point sim_i^k is given by the following formula:

$$sim_i^k = sim_{i*} + \frac{k-1}{\alpha_i - 1}(sim_i^* - sim_{i*})$$

Estimating the u_i functions is equivalent to estimating the variables $u_i(sim_i^k) = u_i^k$. Therefore, the marginal utility of an alignment a_j w.r.t. similarity measure sim_i, is approximated by a linear interpolation as shown in Figure 1. Thus, for $sim_i(a_j) \in [sim_i^k, sim_i^{k+1}]$, we have:

$$u_i(sim_i(a_j)) = u_i^k + \frac{sim_i(a_j) - sim_i^k}{sim_i^{k+1} - sim_i^k}(u_i^{(k+1)} - u_i^k)$$

To find the variables u_i^k, we need to resolve the following linear program LP:

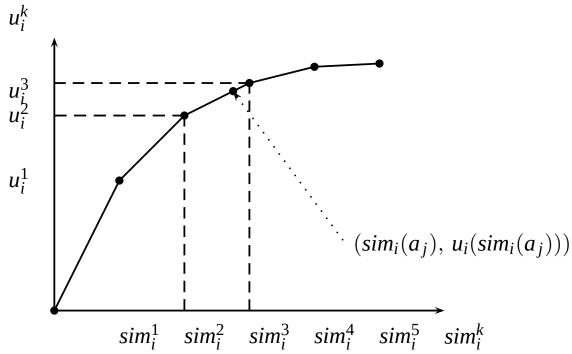

Figure 1: Computation of the marginal utility of an alignment a_j w.r.t. similarity measure sim_i (In the above figure, $u_i(sim_i(a_j)) = u_i^2 + \frac{sim_i(a_j) - sim_i^2}{sim_i^3 - sim_i^2}(u_i^3 - u_i^2)$).

$$\begin{aligned} &Min \quad Z = \sum_{a_j \in \widetilde{R}} \sigma(a_j) \\ &s.t \\ &\quad \sum_{i=1}^{p}(u_i(sim_i(a_j)) - u_i(sim_i(a_k))) + \sigma(a_j) \\ &\quad -\sigma(a_k) \geq \delta; \forall (a_j, a_k) \in \widetilde{R}^2 : a_j P a_k \\ &\quad \sum_{i=1}^{p}(u_i(sim_i(a_j)) - u_i(sim_i(a_k))) + \sigma(a_j) \\ &\quad -\sigma(a_k) = 0; \forall (a_j, a_k) \in \widetilde{R}^2 : a_j I a_k \\ &\quad u_i^{(k+1)} - u_i^k \geq s_i; \forall i = 1, \ldots, p; k = 1, \ldots, (\alpha_i - 1) \\ &\quad \sum_{i=1}^{p} u_i(sim_i^*) = 1 \\ &\quad u_i(sim_{i*}) = 0; \forall i = 1, \ldots, p \\ &\quad u_i^k \geq 0, \forall i = 1, \ldots, p; k = 1, \ldots, (\alpha_i - 1) \\ &\quad \sigma(a_j) \geq 0; \forall a_j \in \widetilde{R} \end{aligned}$$

In the preceding linear program LP, the first two family of constraints model the reference preorder S. Using transformations of equation (3), they only involve the principle variables u_i^k. The third family of constraints are set since u_i are supposed to be non-decreasing marginal utility functions. The 4th constraint as well as the 5th family of constraints are set for normalization purposes: alignments scores will range from 0 to 1. The last two family of constraints specify that the principle variables u_i^k as well as the auxiliary variables $\sigma(a_j)$ are non-negative. Moreover, auxiliary variables $\sigma(a_j)$ model errors, δ is an arbitrary small positive value parameter so as to significantly discriminate two successive equivalence classes of $\widetilde{R}$, and s_i is an indifference threshold parameter defined on similarity measures sim_i to model imprecision.

The linear program LP can be resolved using the Simplex algorithm. Besides, the structure of the preceding LP is such that it is more useful to solve the dual in order to save time and memory.

If the optimal solution is $Z^* = 0$, then there exists a least one utility function U compatible with the reference preorder S. When the optimal value $Z^* > 0$, then there is no value function U compatible with the reference preorder S. In such circumstances, we can pursue one of the following strategies:

- increase the number α_i of breakpoints sim_i^k for one or several marginal utility functions u_i,
- ask the user to revise the reference preorder on $\widetilde{R}$, or
- search over the relaxed domain $Z \leq Z^* + \varepsilon$ an additive value function U giving a preorder $\widehat{S}$ on $\widetilde{R}$ which is sufficiently close to the reference preorder S, in the sense of Kendall-tau distance or Spearman-footrule distance. Branch and bound methods could be used here.

The resulting solution of the preceding LP program is therefore used to compute the similarity score of all the remaining alignments of R using formula of equations (3) and (3), and rank them accordingly. This guaranties that the resulting ranking is coherent with the expert judgements.

Our model differs from common linear combination methods w.r.t. the following features. First, our model incorporates a semi-supervised learning phase which allows expert intervention. Moreover, linear combination methods consider that performances on each similarity measure sim_i increases linearly all along the range E_i, which is not necessary in our model as shown in Figure 2.

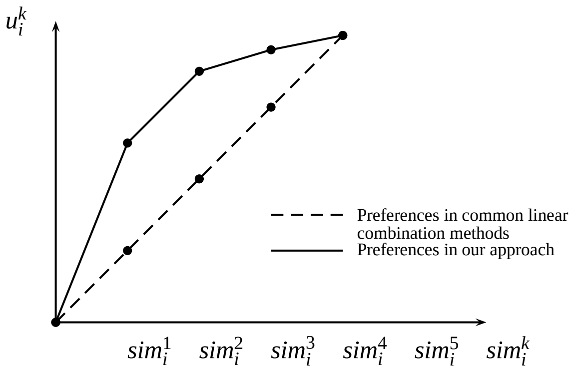

Figure 2: Encoding of performances.

4 EXPERIMENTATION

In this section, we report results of first experimentations that show that the proposed approach is promising.

First, we select the SatellitesSceneOntology ontology of DAFOE platform as a target ontology and the FTT ontology as a source ontology. For the comparison of concepts, we apply 3 similarity measures: Hamming, Maedche & Staab, and Wu & Palmer.

Table 1 summarizes some mappings as well as their similarities with respect to Hamming (H), Maedche(M) and Wu & Palmer (WP) measures. Their ranking (R) is established according the sum of the values of the different similarity measures (global score: GS).

Table 1: Similarity values and their rankings.

Mapping	H	M	WP	GS	R
(sea, seas)	1	0.43	1	2.43	1
(forest, rain forests)	0.66	0.67	0.53	1.86	2
(Building, Mansions)	0.25	0.52	0.88	1.65	3
(construction, building)	0.21	0.095	0.93	1.235	4
(forest, vidaucts)	0.08	0.55	0.25	0.88	5

Asking experts to rank these mappings, we find that they propose two different rankings (cf. Table 2) depending to the fact that we compare concepts with respect to the equivalence relation (*eq*) or the subsumption relation (*sub*).

Table 2: Rankings of the experts.

Mapping	Ranking (*eq*)	Ranking (*sub*)
(sea, seas)	1	3
(forest, rain forests)	2	1
(building, mansions)	4	2
(construction, building)	3	4
(forest, viaducts)	5	5

This table shows that in the equivalence context, the ranking of mappings 3 and 4 is reversed which is easy to explain since the mansion is a sub-concept of building whereas building and construction are synonyms, therefore more ‘similar’ in the equivalence context.

The application of the UTA method are summarized in tables 3.

As we can see, for each similarity measure, the first column indicates the similarity value, the second one indicates the utility score in the equivalence context, and the third column indicates the utility score in the subsumption context. Let remark that for the first mapping (forest, forest), the overall utility score in the equivalence context is remarkably higher than the one found in the subsumption context, which means that both concepts are more equivalent than subsumed. Also for the mapping (forest, rain forests), the overall utility score in the equivalence context is remarkably lower than the one found in the subsumption context, which means that both concepts are more subsumed than equivalent.

The score of any mapping a in the equivalence context is given by the following formula:

Table 3: Rankings of the experts.

Mapping	Hamming		
(sea, seas)	1	0.374	0.199
(forest, rain forests)	0.66	0.340	0.199
(Building, Mansions)	0.25	0.121	0.092
(forest, vidaucts)	0.08	0	0
Mapping	**Maedche**		
(sea, seas)	0.43	0	0
(forest, rain forests)	0.67	0.158	0.467
(Building, Mansions)	0.52	0	0.167
(forest, vidaucts)	0.55	0	0.167
Mapping	**Wu & Palmer**		
(sea, seas)	1	0.467	0.333
(forest, rain forests)	0.53	0.292	0.303
(Building, Mansions)	0.88	0.383	0.333
(forest, vidaucts)	0.25	0	0

Mapping	Scoring (*eq*)	Scoring (*sub*)
(sea, seas)	0.84	0.53
(forest, rain forests)	0.79	0.97
(Building, Mansions)	0.51	0.58
(forest, vidaucts)	0	0.16

$$\begin{aligned} U(a) &= 0.374 * u_1(Sim_{Hamming}) \\ &+ 0.158 * u_2(Sim_{Maedche}) \\ &+ +0.46 * u_3(Sim_{WUP}) \end{aligned}$$

whereas its score in the subsumption context is given by the following formula:

$$\begin{aligned} U(a) &= 0.199 * u'_1(Sim_{Hamming}) \\ &+ 0.467 * u'2(Sim_{Maedche}) \\ &+ 0.333 * u'3(Sim_{WUP}) \end{aligned}$$

These formula can be now used to rank all the candidate mappings coming from SatellitesSceneOntology and FTT ontologies.

5 CONCLUSION

In this paper, after presenting a literature review of the main similarity models used to map or align ontology entities, we propose a semi-automatic mapping selection process in order to build a satellite images ontology by reusing geographical object ontologies. The main advantage of our work is that it needs little human intervention to monitor the mapping process. First experimentations show that our approach is promising.

REFERENCES

Charlet, j., Szulman, S., Aussenac-Gilles, N., Nazarenko, Hernandez, N., Nadah, N., Sardet, E., Delahousse, J., Valry Tguiak, H., and Baneyx, A. (2010). Dafoe: une plateforme pour construire des ontologies partir de textes et de thesaurus. In *10ime Confrence Internationale Francophone sur l'Extraction et la Gestion des Connaissances*. EGC.

Couto, F. M.and Silva, M. J. and Coutinho, P. M. (2005). Semantic similarity over the gene ontology: family correlation and selecting disjunctive ancestors. In ACM, editor, *the 14th ACM International Conference on Information and Knowledge Management*, pages 343–344.

Durand, N., Derivaux, S., Forestier, G., Wemmert, C., Gancarski, P., Boussaid, O., and Puissant, A. (2007). Ontology–based object recognition for remote sensing image interpretation. In *IEEE International Conference on Tools with Artificial Intelligence*, pages 472–479, Greece.

Egenhofer, M. (2002). Toward the semantic geospatial web. In *10th ACM International Symposium on Advances in Geographic Information Systems*, 10.1145/585147.585148, pages 1–4. ACM.

Gardenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. Massachusetts Institute of technology, Cambridge, 2004 edition.

Gentner, D. (1989). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7:155–170.

Gesbert, N. (2005). *Etude de la formalisation des spcifications de bases de donnes gographiques en vue de leur intgration*. PhD thesis, Universit de Marne la Valle et IGN.

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:3–28.

Goldstone, R. L. (1999). Similarity. In *The MIT encyclopedia of the cognitive sciences*, pages 763–765.

Goldstone, R. L. (2005). Similarity. In *Cambridge handbook of thinking and reasonning*, pages 13–36.

Hahn, U., Close, J., and Graf, M. (2009). Transformation direction influences shape similarity judgements. *Psychological science*, pages 447–454.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, pages 147–160.

Holyoak, K. J. and Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13:295–355.

Hudelot, C., Atif, J., and Bloch, I. (2006). Ontologie de relations spatiales floues pour l'interprtation d'images. In *Rencontres francophones sur la Logique Floue et ses Applications*, Toulouse, France. LFA 2006.

Imai, S. (1977). Pattern similarity and cognitive transformations. *Acta Psychologica*, 41(6):433–447.

Jacquet-Lagrèze, E., Meziani, R., and Slowinski, R. (1987). Molp with an interactive assessment of a piecewise utility function. *Eur. J. Oper. Res*, 31(3):350–357.

Jacquet-Lagrèze, E. and Siskos, Y. (1982). Assessing a set of additive utility functions for multicriteria decision making: the UTA method. *European Journal of Operational Research*, 10:151–164.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

Jiang, J.and Conrath, D. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *International Conference on Research in Computational Linguistics*, Taiwan.

Keeney, R. and Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. J. Wiley, New York.

Leacock, C.and Chodorow, M. (1998). Combining local context and wordnet similarity for word sense identification. In *MIT Press*, pages 265–283.

Lin, D. (1998). An information–theoretic definition of similarity. In Madison, M.-K., editor, *the fifteenth International Conference on Machine Learning*, pages 296–304.

Markman, A. B., . G. D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25:431–467.

Nefzi, H., Messaoudi, W., Farah, M., and Farah, I. R. (2013). Vers une ontologie riche ddie l'imagerie satellitaire par rutilisation de ressources existantes. In *TAIMA'2013*, pages 35–46.

Parent, C., Spaccapietra, S., and Zimanyi, E. (1998). modle conceptuel spatio–temporel. *Revue internationale de gomantique*, (7):317–352.

Resnik, P. (1995). Using information content to evaluate semantic similarity in taxonomy. In *14th International Joint Conference on Artificial Intelligence*, Montreal.

Resnik, P. (1999). Semantic similarity in a taxonomy: An information based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*.

Roy, B. (1991). The outranking approach and the foundations of ELECTRE methods. *Theory and Decision*, 31:49–73.

Schlicker, A., Domingues, F. S., Rahnenf, J., and Lengauer, T. (2006). A new measure for functional similarity of gene products based on gene ontology. *BMC Bioinformatics*, 7(302).

Schwering, A. and Raubal, M. (2005). Measuring semantic similarity between geospatial conceptual regions. In *the First International Conference on GeoSpatial Semantics*, Mexico City,Mexico.

Shvaiko, P. and Euzenat, J. (2008). Ten challenges for ontology matching. In in Computer Science, L. N., editor, *OTM'08: Proceedings of the OTM 2008 Confederated International*, volume 5332, pages 1164–1182. OTM, Springer.

Sjoberg, L. (1972). A cognitive theory of similarity. *Goteborg Psychological Reports*, 2.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4):327–352.

Wiener Ehrlish, W., Bart, W., and Millward, R. (1980). An analysis of generative representation systems. *mathematical psychology*, pages 219–246.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS