Laboratoire IBISC

Biologie Intégrative et Systèmes Complexes

ROMIE: a Resource-based Ontology Mapping

Amel Bouzeghoub[1] , Abdeltif Elbyed[1,2], Fariza Tahi[2]

[1]Institut Telecom SudParis, Evry, France

[2]IBISC University of Evry-Val d'Essonne, Genopole, France

RAPPORT DE RECHERCHE
IBISC-RR-2011-01
mars 2011

ROMIE: a Resource-based Ontology Mapping

Amel Bouzeghoub[a], Abdeltif Elbyed[a,b], Fariza Tahi[b]

[a] *Institut Telecom SudParis, Evry, France*
[b] *IBISC University of Evry-Val d'Essonne, Genopole, France*

Abstract

System interoperability is an important issue, widely recognised in information technology intensive organisations and in the research community of information systems. The wide adoption of the World Wide Web to access and distribute information further stresses the need for system interoperability. Initiatives solutions like the Semantic Web facilitate the localisation and the integration of the data in a more intelligent way via the use of ontologies. The Semantic Web offers a compelling vision, yet it raises a number of research challenges. One of the key challenges is to compare and map different ontologies, which evidently appears in integration tasks. The main goal of the work is to introduce a method for finding semantic correspondences among ontologies. The approach brings together syntactic, linguistics, structural and semantic (based on instance information) matching methods in order to provide a semiautomatic mapping. The approach consists of two phases: semantic enrichment phase and mapping phase. The enrichment phase is based on the analysis of the extension information (like web resources, data, documents, etc.) that are associated to the ontology concepts. At the end of enrichment phase, the ontology contains more semantic relations between its concepts which will be exploited in the mapping phase. A process of filtering enables us to automatically reduce the number of false relations. The validation of the correspondences is an interactive process (with an expert) which allows to improve the mapping process. The approach has been implemented in a prototype system called ROMIE (Resource based Ontology Mapping within and Interactive and Extensible environment). It was tested and evaluated on two applications: a biomedical application and technology enhanced learning (or e-learning) domain application.

Keywords:
Ontology mapping, instance-based mapping, ontology enrichment, similarity measures, semantic web.

1. Introduction

We assist these last years to the emergence of new applications that need to share information between various systems. The stake is to develop techniques that facilitate semantic interoperability between these information systems, which generally constitute autonomous and heterogeneous data sources.

Interoperability is an important question, largely identified in several domains such as Technology Enhanced Learning domain (TEL), Biology, Bioinformatics, and so on. The dependence and the share of information between organisations created a need for co-operation and co-ordination in order to make easier the exchange and the access to distant or local information. The large adoption of the Internet to reach and distribute information generates a crucial

Email address: Amel.Bouzeghoub@it-sudparis.eu, Abdeltif.Elbyed@it-sudparis.eu, fariza.tahi@ibisc.univ-evry.fr (Fariza Tahi)

need for the interoperability between systems. It is thus important to provide users methods and tools allowing a transparent access to the huge number of data and resources accessible via the Web. Ontologies are an important tool providing semantics to data and documents in several areas. They are used as basis for interoperability between systems and for data integration, by providing a common terminology over a domain. In most of considered domains, there exist an increasing number of ontologies, which are mostly complementary but could also contain important overlapping. In order to use the needed ontologies in an integrated way, "bridges", i.e. mappings, between ontologies must be built. Mapping two ontologies means to define semantic relations between their respective concepts. Mappings are often established manually by the domain experts, but because of the increasing number of ontologies as well as the increasing of their size, there is a need of automatic (or semi-automatic) mappings. Almost existing ontology mapping algorithms use and combine several matchers, mostly syntactic and linguistic. Some of them use information on the structure (topology) of considered ontologies. Very few use information on documents or resources (instances) connected to the ontologies.

In this paper, we propose an instance-based ontology mapping system called ROMIE (Resource-based Ontology Mapping). Most instance-based mapping systems rely on the assumption that there are individuals shared between the ontologies to be mapped. The overlap between instances of the classes belonging to different ontologies is then used to decide whether those classes have to be mapped. Often this precondition is not fulfilled. Therefore, we have developed in ROMIE a method where there is no need of instances shared between the different ontologies. We propose an original instance-based mapping where semantic relations between resources are propagated to ontology concepts. We propose also in this paper how to generate semantic relations between resources if they do not exit, using properties of the resources. We considered as a case study the application of TEL.

Our approach is composed of two principal phases: (i) semantic enrichment and (ii) mapping. The phase of enrichment consists to add new semantic relations between ontology concepts, by propagating relations existing (or generated by an analysis of the resources properties) between resources to the concepts linked to these resources. The mapping phase takes two ontologies (eventually enriched) and calculates similarities between concepts. The similarity calculation is based on the semantic distance and uses relationships generated in the ontology enrichment phase (if any). The mapping process uses in addition traditional methods of similarity measure such as syntactic, linguistic and structural methods. It presents also a filtering process which allows for automatically reduce the number of false generated mappings. Finally, ROMIE offers an interactive process of validation where validated mappings are continuously taken into account in the mapping process to improve the results.

The paper is organised as follows. In the first section we present a state of the art of the domain of ontology mapping and the motivation of our work. We present then the architecture and the different steps of the mapping process in ROMIE, before describing the ontology enrichment phase and the mapping process which is composed of different modules: syntactic and linguistic matchers, structural and semantic matchers, filtering module and validation module. We describe then how ROMIE was applied to the TEL domain and more precisely how the step of resources analysis for the enrichment step was achieved in this case study. Finally we describe the developed prototype and give results obtained in TEL domain and also in biomedical domain, before concluding.

2. State of the art and motivation of the work

A great number of ontology mapping approaches exist, as pointed out in [18, 27, 10]. All of them use several matchers, which are often of different types: syntactic, linguistic, structural (topological) and semantic. The semantic matchers are usually based on semantic relations between the concepts in each of considered ontologies. Some of existing approaches of ontology mapping consider the expressive ontology language for defining these relations. For example, in [11], authors use a subset of OWL Lite for this purpose but they mainly focus on the comparison of the structural aspects of ontologies.

There are several works on ontology mapping based on the instance-based (resource-based) approach [16, 20, 19, 4, 3, 31]. In all these works, to define a similarity measure between concepts, there is an explicit reference to the ontology model of OWL Lite and the similarity is defined among OWL objects (i.e., concepts) in terms of the number of common instances that characterise each concept. In [20, 35], a system called Automs is proposed. It creates a semantic mapping based on ontology metadata. The ontology model adopted in this approach refers also to the hierarchy relation. In [16], the proposed approach uses the Jaccard measure to calculate the statistics of common instances between two concepts. In [19], four matchers are defined in order to determine the instance-based similarity, using the number of instances that are associated or not associated to two compared concepts. In other words, the degree of similarity between two concepts takes into account the number of shared instances. In [31], the authors address the problem of migrating instances between ontologies; they exploit existing mappings between ontologies to reclassify a set of instances of one source ontology into related target ontology.

We can find in the literature ontology mapping methods for specific application domains. For instance Lambrix and Tan developed a system, called SAMBO, for aligning and merging biomedical ontologies, which uses and combines different matchers and mapping strategies [21, 22]. Kirsten and al. proposed in [19] an instance-based matching method for life science ontologies. The authors have suggested the use of data contained in the biological data source Ensembl [1] in order to find potential mappings between concepts sharing same data. Their method was tested on two biological ontologies, GO and OMIM.

Several other mapping systems have been proposed in the literature. Chimaera [23] is an environment of merging ontologies; it uses heuristics for finding ontology portions to reorganise. OntoMorph [6] is a system close to Chimaera; it uses almost available matching techniques. Prompt [28] is a semi-automatic system for aligning and merging ontologies, close to OntoMorph and Chimaera systems. It is perhaps the most well-known system, as it is a plugin in the well-known ontology editor, Protégé [26]. It searches for possible matching between concepts using linguistic similarity and ontology structure. GLUE [8] is an ontology mapping system which uses information on instances (name, size, ...) and on the words frequency; it takes into account the common sense, the domain constraints and the ontology structure (concepts neighbourhood). It implements three learning strategies: one for the names, one for the concepts and one for combining the two approaches (in a probabilistic manner). Finally, FOAM [9] is a framework for Ontology alignment and mapping based on a combination of a rule-based approach and a machine learning approach.

The goal of our work is to offer a system able to identify the semantic links between ontologies with high selectivity and sensitivity, and better interactivity with users, compared to existing systems. In this purpose, we developed methods that allow maximising the number of generated mappings while minimising the number of false mappings, using all information contained in the considered ontologies, information contained in the resources associated to the ontologies, as well as information provided by users.

- It is not clear to know in the existing tools if they have an interactive process neither how this interaction is realised. We consider that the interaction with users is a fundamental point that is important to integrate in the mapping tools. In our system, we offer different types of interactions with users: (i) validation or invalidation of the generated mappings; (ii) choice of the methods to use for the calculation of similarity between concepts and (iii) choice of parameters values (thresholds and confidence values for each method).

- One important goal of our system is to reduce the number of false positive generated mappings. For this purpose, we developed a method for filtering the generated mappings which go far away than methods based on thresholds traditionally used in ontology mapping systems. Our method exploits all possible links (hierarchical and semantic links) existing between concepts of each ontology in order to detect certain anomalies and contradictions between the obtained mappings.

- The step of mapping results validation is an important phase in the mapping process. In existing systems, the validation is done by the expert at the end, once all results are provided. In our system, there is a continuous interaction with the expert, where each generated mapping could be validated at any time; the validated mappings are then used continuously in the mapping process for improving the mapping.

- Existing systems provide a mapping between all concepts of two given ontologies. It is necessary to offer to users a mapping system able to identify the mappings between particular concepts. Our system allows to achieve a total or a partial mapping of the ontologies.

- Several existing works on mapping ontologies are concentrated on techniques calculating similarities between concepts and on how combining them. Existing algorithms success in the best case to propose 70% of correct mappings and to identify 80% of existing mappings. Besides, these values change according to the structure and the semantic richness of the ontologies. Therefore there still being an effort to do in order to obtain better reliability of the mapping. Our goal is to analyze and exploit the resources (or instances) linked to ontologies, in order to enrich their semantic and improve results of the mapping process.

  We suppose that the semantic richness of ontology concepts come from the interpretation of the resources and information annotated by these concepts. There is indeed no standardisation allowing that the information interpretations done when creating an ontology are the only ones possible conceptualisations in the real word. Consequently, the starting point of the comparison and the creation of the semantic correspondences between heterogeneous ontologies is to semantically enrich these ontologies before the mapping process. The semantic enrichment allows to make explicit the "hidden" semantic of the different ontology concepts. And more the semantic is explicitly specified, more the ontologies comparison becomes reliable and feasible.

In this work, we were interested by two domain applications, the domain of Technologies Enhanced Learning (TEL) and the biomedical domain.

3. General architecture of ROMIE

The mapping process is based on the measure of similarity between concepts of different ontologies. An important step used by all existing algorithms and systems for mapping ontologies is the use of syntactic and linguistic methods in order to measure the similarity between concepts in a terminological point a view. Nevertheless, these methods are not sufficient for a good and appropriate measure. That is why other types of methods, namely structural (topological) methods, are often used. These methods take into account the structure of the ontology, mostly information on concept neighbourhood, i.e. fathers and children of the concepts to map. Besides these relations of father and child between concepts of the ontology, we consider in our system another kind of relations, namely semantic relations, which are very important in order to add a semantic level to the similarity. By semantic relation, we mean any relation, different from those usually used (hierarchical relations, e.g. "is a" relations) which may exist between two concepts. Unfortunately, existing ontologies have rarely this type of relations. Therefore, an important step of our system is the enrichment of each of considered ontology by semantic relations between concepts, before performing mappings. These relations are generated using information on resources connected to the ontology concepts (this step is explained in the next section). The mapping process in our system is performed in several steps (see Figure 1):

- Similarity values calculation: syntactic and linguistic methods (matchers) are applied on cou-

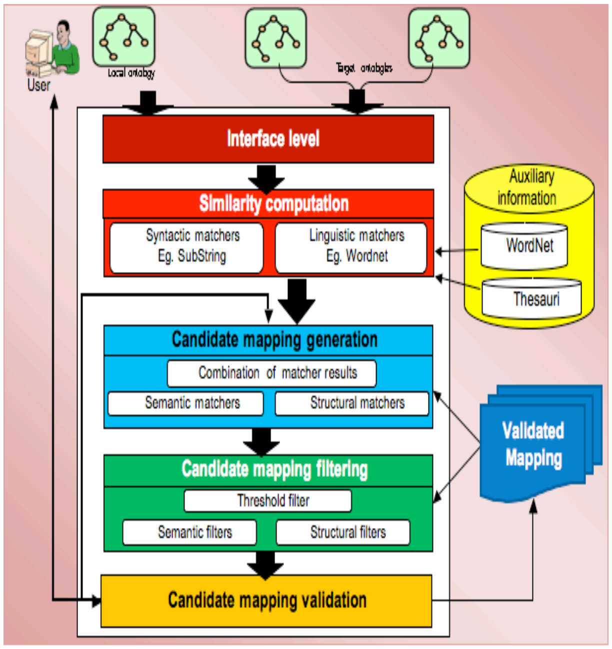

Figure 1: General architecture of the mapping process

ples of concepts, in order to measure their terminological similarity.

- Candidate mappings Generation: candidate mappings are generated in this step, with a similarity values resulting from the combination of the similarity values returned in the first step. In this step, structural (topological) and semantic matchers are also applied to improve the similarity values of candidate mappings and/or to generate new candidate mappings. They are based on relations between concepts, and on mappings already established and validated. Structural matchers use hierarchical relations between concepts (e.g. 'is-a' relation) when semantic matchers use other types of relations (eventually generated in the enrichment step).

- Candidate mappings filtering: in order to minimise the number of false candidate mappings, several methods are used for filtering the generated candidate mappings. Besides the conventional method of threshold, we use filtering methods based on structural and semantic relations.

- Candidate mapping validation: the last step of the mapping process is the validation of the generated mappings. The user can validate or invalidate each of the candidate mappings. When a mapping is validated, it is stored and then used by the system in the steps of candidate mappings generation and candidate mappings filtering.

  The different steps of the mapping process are described below in Section 5.

4. Ontology enrichment

An important contribution of this paper is the ontology enrichment, which plays a crucial role to improve mapping results. As presented in Section 3, we use hybrid approach of ontology mapping, which mixes linguistic, structural and semantic approaches. The semantic approach exploits different semantic characteristics of ontologies. The problem is that the number and/or the quality of existing semantic relations in ontologies are in general very low. The main objective of our instance-based approach is to exploit information about the resources connected to the ontology concepts in order to infer new semantic relations between the concepts.

The principle of our approach is to analyze the considered resources, in order to highlight relations between their instances, and then to propagate these relations to the ontology concepts linked to these instances (Figure 2).

In case there are no evident relations between resource instances, a first step of the ontology enrichment process is the generation of relations between resource instances. This step requires the analysis of the resources properties, which depends on the application domain. In this paper we illustrate this step on the TEL application domain (see Section 6).

Generation of semantic relations between concepts

For each relation *Rel* between two resources, an equivalent relation between associated concepts is generated. However, several resources could be linked

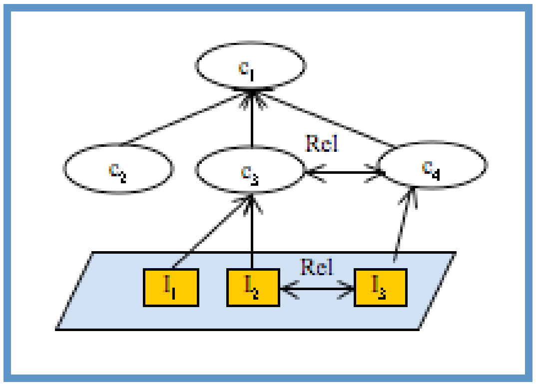

Figure 2: Propagation of relations between instances towards associated concepts.

to a concept. Therefore, we must consider the set of resources linked to each concept rather than an individual resource. Moreover, we can have different relations between two sets of resources, where each relation links a sub-set of the first set with a sub-set of the second set. Therefore, we have to distinguish several subsets of resources called equivalent resources, and several relations to propagate to a couple of concepts (see Figure 3).

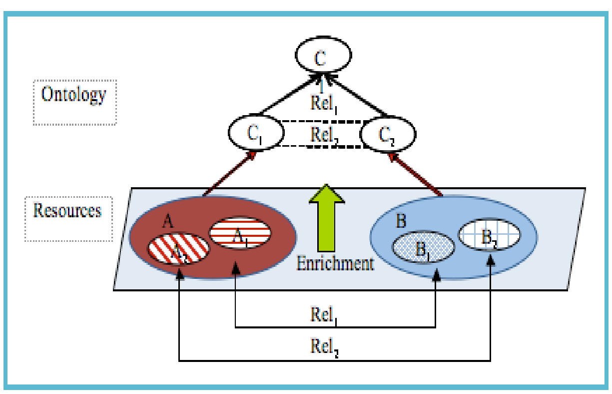

Figure 3: Propagation of relations between sets of resources towards associated concepts.

Given two concepts C_1 and C_2, each subset A_i of equivalent resources associated to C_1 may have a semantic relation Rel_i with another subset B_i associated to the concept C_2. The problem is that there are thus several Rel_i which could be propagated towards the corresponding concepts (in order to obtain a new semantic relation between the concepts C_1 and C_2) (see Figure 3). Our proposition is to propagate each relation Rel_i for which we associate a weight depending on the similarity between the corresponding subsets A_i and B_i. In this purpose, we use the well-known formula of Jaccard's similarity measure. The Jaccard measure is used to calculate the distance between two sets; it takes the lowest value 0 when they are disjoint, and the highest value 1 when they are equal.

Let $\Re = \{Rel_1, Rel_2, \ldots, Rel_K\}$ a set of semantic relations between two sets of resources A and B. We adopt for each Rel_i the following notation $A \leftarrow Rel_i \rightarrow B$. The weight of the relation Rel_i between the sets A_i and B_i, noted $\sigma_{Rel_i}(A_i, B_i)$, is calculated by the Jaccard measure formula:

$$\sigma_{Rel_i}(A_i, B_i) = \frac{|A \bigcap_{Rel_i} B|}{|A \bigcup B|}$$

with:

$$A \bigcap_{Rel_i} B = \{A_i / \forall a \in A_i, \exists b \in B_i / a \leftarrow Rel_i \rightarrow b\}$$

5. Mapping process

Ontology mapping is the task of relating the vocabulary of two ontologies in such a way that the mathematical structure of ontological signatures (vocabulary) is respected. Structure-preserving mappings between mathematical structures are called morphisms. We thus study the mapping of ontologies through the ontology morphism properties.

After defining the ontology morphism principle, we describe in this section the different steps of our mapping process and the different methods we developed.

5.1. Ontology morphism

The principle of ontology morphism is to consider that each relation (structural or semantic relation) between two concepts of an ontology is equivalent to the relation between the corresponding concepts (by mapping relations) in another ontology. This leads to the following definitions:

- *Definition 1*: An ontology O is defined by a tuple $O = (C, R, <, \sigma)$ consisting of (i) two disjoint sets C and R called concept identifiers and relation identifiers respectively, (ii) a partial order $<$ on C called concept hierarchy or taxonomy, (iii) a function σ: $R \rightarrow C \times C$ called signature that associates to relations couple of concepts.

- *Definition 2*: An ontology morphism between two ontologies $O = (C, R, <, \sigma)$ and $O' = (C', R', <', \sigma')$ is the couple of functions (F, G) such that: $F : C \rightarrow C'$ and $G : R \rightarrow R'$. Given c and d two elements (concepts) of C and r an element (relation) of R we note that:

  $F(c) = c'$ is the corresponding concept of c, $F(d) = d'$ is the corresponding concept of d and $G(r) = r'$ is an equivalent relation to r, particularly r' equal to r.

Using ontology morphism, we can deduce the following rules:

- *Rule 1*: If $c < d$ then $F(c) <' F(d)$; that means: "If c precedes d in ontology O, then c' precedes d' in ontology O'."

- *Rule 2*: If $\sigma(r) = (c, d)$ then $\sigma'(G(r)) = (F(c), F(d))$; that means: "If r is a relation between c and d in ontology O, then r' (eventually r) is a relation between c' and d'."

5.2. Syntactic and linguistic matchers

Syntactic and linguistic methods (matchers) are applied on couples of concepts in order to measure their terminological similarity. We use in our case several conventional and known syntactic methods (e.g. Hamming distance, subString distance, N-Gramme distance, Levenshtein distance), which calculate an edit distance between terms. And we use a linguistic method based on WordNet dictionary [24, 13], which is a lexical database for the English language providing detailed and precise descriptions of words. Each of these matchers returns a similarity value. Each matcher returns a five-tuple $< M, Rel, c, d, Conf, SV >$ where:

- M: is the matcher name (e.g. WordNet);

- Rel: is the type of relation between the concepts c and d generated by the matcher (include relation ($\supseteq$), overlap relation ($\subseteq$) or equal relation ($\equiv$));

- $Conf$: is a confidence level associated to the matcher;

- SV: is the similarity value (between 0 and 1) between c and d calculated by the matcher M.

The similarity values returned by the different syntactic and linguistic matchers are then combined. For each couple (c, d) of concepts, a candidate mapping of the form: $Hp =< Rel, c, d, Conf_{Hp}, SV_{Hp} >$ is generated, where:

$$Conf_{Hp} = \sum_i Conf_i$$

and

$$SV_{Hp} = \frac{\sum_i Conf_i \times SV_i}{\sum_i Conf_i}$$

and $Conf_i$ and SV_i are respectively the confidence level of the method M_i and the similarity value returned by the method M_i for the couple of concepts (c, d).

5.3. Structural and semantic matchers

Relations between ontology concepts are exploited in the mapping process by specific matchers called structural or semantic matchers. Structural matchers use structural or topological relations between concepts. Semantic matchers use semantic relations between concepts, which are either relations generated by the ontology enrichment step, or relations belonging to the original ontologies; in the first case they are considered as instance-based (or resource-based) semantic matchers.

These matchers allow (i) to improve (i.e. increase) the similarity value of the candidate mappings obtained in the precedent step and (ii) to generate new candidate mappings thanks to mappings already validated.

These matchers are based on the principle that: two concepts are more likely to be the same if their structural or semantic neighbourhood concepts are similar. Here are examples of rules:

- R1: Two concepts are similar if their "fathers" (i.e. super-concepts) are similar.

- R2: Two concepts are similar if their "children" (i.e. sub-concepts) are similar.

- R3: Two concepts are similar if their "neighbourhood" are similar.

The two first rules are used by the structural matchers and concern the hierarchical structure of the ontologies, where the third one is used by the semantic matcher. These rules are based on the notion of ontology morphism presented above (Section 5.1).

5.3.1. Structural matchers

We defined two structural matchers: the "Top-down" matcher (based on the rule R1) and the "Bottom-up" matcher (based on the rule R2).

- Top-down matcher: this method consists to use a validated mapping between two concepts for improving a candidate mapping generated between their "children" concepts (Figure 4, left):

  If $\exists$ a validated mapping $Map =< Rel, c, d, Conf, SV >$ between a concept c of $O1$ and a concept d of $O2$

  and

  If $\exists$ a candidate mapping $Hp =< Rel, c', d', Conf_{H_p}, SV_{H_p} >$ between a concept c' of $O1$ and a concept d' of $O2$ where:

  c (respectively d) is the "father" concept of c' (respectively d')

  Then modify Hp such as:

$$SV_{H_p} = SV_{H_p} + \left(\frac{SV}{min(NbrOfChild(c), NbrOfChild(d))}\right)$$

  and

$$Conf_{H_p} = Conf_{H_p} + Conf_{Top-Ddown}$$

  with $NbrOfChild(c)$ (respectively $NbrOfChild(d)$) the number of "children" concepts of c (respectively d) and $Conf_{Top-Ddown}$ the confidence level of the Top-down matcher.

- Bottom-up matcher: this method consists to use a validated mapping between two concepts for generating a new candidate mapping between their "father" concepts or for improving it if it already exists (Figure 4, middle):

  If $\exists$ a validated mapping $Map < Rel, c, d, Conf, SV >$ between a concept c of $O1$ and a concept d of $O2$, then:

  Let us consider the "father" concept c' of c (in $O1$) and the "father" concept d' of d (in $O2$)

  If $\exists$ a candidate mapping $Hp < Rel, c', d', Conf_{H_p}, SV_{H_p} >$ between c' and d',

  Then modify Hp such as:

$$SV_{H_p} = SV_{H_p} + \left(\frac{SV}{min(NbrOfChild(c'), NbrOfChild(d'))}\right)$$

  and

$$Conf_{H_p} = Conf_{H_p} + Conf_{Bottom-up}$$

  otherwise generate a new candidate mapping $Hp < Rel, c', d', Conf_{H_p}, SV_{H_p} >$ such as:

$$SV_{H_p} = \left(\frac{SV}{min(NbrOfChild(c'), NbrOfChild(d'))}\right)$$

  and

$$Conf_{H_p} = Conf_{Bottom-up}$$

  with $NbrOfChild(c')$ (respectively $NbrOfChild(d')$) the number of "children" concepts of c' (respectively (d') and $Conf_{Bottom-up}$ the confidence level of the Bottom-up matcher.

The transitive property of the hierarchical (structural) relations of ontologies allow to define direct and indirect children or fathers. The Top-down and Bottom-up matchers defined above consider direct "children" respectively "fathers" concepts. We also developed Top-down and Bottom-up matchers which consider indirect "children" and "fathers" concepts. They are called "Top-down-ind" and "Bottom-up-ind" matchers.

- Top-down-ind matcher: this method consists to use a validated mapping between two concepts for improving a candidate mapping generated between their indirect children concepts (Figure 4, right):

  If $\exists$ a validated mapping $Map =< Rel, c, d, Conf, SV >$ between a concept c of $O1$ and a concept d of $O2$

  and

  If $\exists$ a candidate mapping $Hp =< Rel, c', d', Conf_{H_p}, SV_{H_p} >$ between a concept c' of $O1$ and a concept d' of $O2$ where:

  c is the "father" concept of level i of c' and d is the "father" concept of level j of d',
  Then modify Hp such as:

$$SV_{H_p} = SV_{H_p} + \left(\frac{SV}{max(i,j) \times min(NbrOfChild_i(c), NbrOfChild_j(d))}\right)$$

and

$$Conf_{H_p} = Conf_{H_p} + \frac{Conf_{Top-Ddown}}{max(i,j)}$$

  where $NbrOfChild_i(c)$ is the number of "children" concepts of c at level i and $NbrOfChild_j(d)$ is the number of "children" concepts of d at level j; and $Conf_{Top-Ddown}$ is the confidence level of the Top-down matcher.

- Bottom-up-ind matcher: it uses the same method than in Bottom-up matcher, but considering indirect children with the same principle than in the Top-down-ind matcher described below, except that only the case of improvement of existing candidate mappings is considered (there is no generation of new candidate mappings).

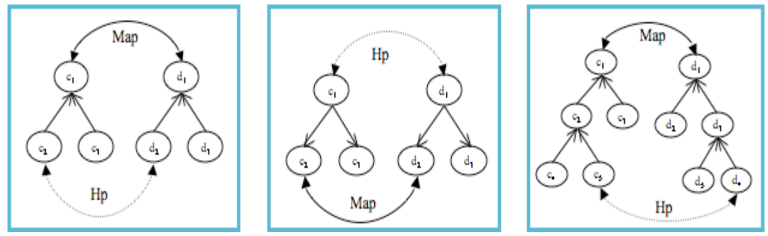

Figure 4: Examples of situations where mapping candidates could be improved using a Top-down matcher (left), a Bottom-up matcher (middle) and by a Top-down-ind matcher (right)

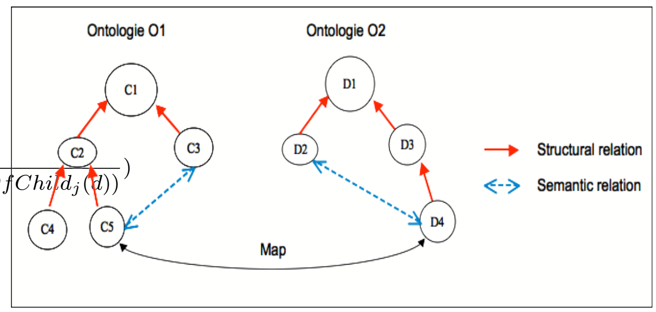

Figure 5: Example of ontologies with semantic relations allowing to generate new mapping candidates. Because we have a validated mapping Map between $C5$ and $D4$, a candidate mapping could be generated between $C3$ and $D2$ by the semantic matcher since there is a same semantic relation between $C5$ and $C3$ and between $D4$ and $D2$ (this relation could be a relation generated by the enrichment phase).

5.3.2. Semantic matchers

As stressed above, besides hierarchical (structural) relations, we use in our mapping process other type of relations existing between ontology concepts, the semantic relations, which could be generated by the enrichment step (see Section 4). In the same way than in structural matchers, semantic matchers allow to improve the mapping process, by generating new mapping candidates and/or improving already generated ones, thanks in this case to the rule R3 of "neighbourhood" given above (Figure 5).

As described in Section 4, a "weight" noted σ is associated to each semantic relation generated by the enrichment phase. In case of semantic relations not generated by the enrichment phase, i.e. semantic relations existing in the original ontology, this weight doesn't exist. It is therefore considered equal to 1 in the following algorithm:

If $\exists$ a validated mapping $Map < Rel, c, d, Conf, SV >$ between a concept c of $O1$ and a concept d of $O2$, then:

Let us consider a concept c' of $O1$ where c is in relation with c' by Rel with a weight σ_1 (we note

($c\ Rel(\sigma_1)\ c'$)) and a concept d' of $O2$ where d is in relation with d' by Rel with the weight σ_2 (we note ($d\ Rel(\sigma_2)\ d'$))

If $\exists$ a candidate mapping $Hp =< Rel, c', d', Conf_{H_p}, SV_{H_p} >$ between c' and d',

Then modify Hp such as:

$$Conf' = Conf' + \sigma \times Conf_{Semantic}$$

and

$$SV' = SV' + \sigma \times \frac{SV}{MinOfConc(c,d)}$$

otherwise generate a new candidate mapping $Hp =< Rel, c', d', Conf_{H_p}, SV_{H_p} >$ such as:

$$Conf' = \sigma \times Conf_{Semantic}$$

and

$$SV' = \sigma \times \frac{SV}{MinOfConc(c,d)}$$

where $\sigma = min(\sigma_1, \sigma_2)$, $Conf_{Semantic}$ is the confidence level of the Semantic matcher and

$$MinOfConc(c,d) = min(NbrOfConc(c, Rel(\sigma_1)), NbrOfConc(d, Rel(\sigma_2)))$$

with $NbrOfConc(c, Rel(\sigma_1))$ (respectively $NbrOfConc(d, Rel(\sigma_2))$) the number of concepts linked to the concept c (respectively d) by the relation $Rel(\sigma_1)$ (respectively $Rel(\sigma_2)$).

5.4. Filtering process

One of the characteristics of ROMIE is the ability to minimise the number of false candidate mappings. In almost existing mapping systems in the literature, the filtering step consists only in the use of a threshold for the similarity values. In ROMIE, besides a threshold filter, we developed several methods for filtering the generated candidate mappings. They are based on the structural and semantic relations (including the semantic relations generated by the ontology enrichment step) existing between ontology concepts. Each method aims to reduce the number of false candidate mappings by using specific comparison rules able to raise contradictions or anomalies. These rules can exploit the topological structure or the semantic relations linking the ontology concepts, respecting the rules of ontology morphism. We distinguish two types of comparison: (i) comparison between two candidate mappings and (ii) comparison between a candidate mapping with a validated mapping. We detect then candidate mappings which are contradictory with validated mappings or with other candidate mappings. In the following we present some of these filters. Let us consider two ontologies $O1$ and $O2$:

- Filter 1: this filter treats the case where a candidate mapping is contradictory with a validated mapping (Figure 6, left):

  If $\exists$ a candidate mapping $Hp < Rel, c, d', Conf_1, SV_1 >$ and $\exists$ a validated mapping $Map < Rel, c', d, Conf, SV >$ such as:

  c and c' are concepts of $O1$ and d and d' are concepts of $O2$ where c (respectively d) is in relation with c' (respectively d') by a structural relation or by a semantic relation,

  Then eliminate Hp.

  (respectively d) (respectively d')

- Filter 2: this method treats the case where a candidate mapping is contradictory with another candidate mapping (Figure 6, middle):

  If $\exists$ two candidate mappings $Hp_1 < Rel, c, d', Conf_1, SV_1 >$ and $HP_2 < Rel, c', d, Conf_2, SV_2 >$ such as:

  c and c' are concepts of $O1$ and d and d' are concepts of $O2$ where c (respectively d) is in relation with c' (respectively d') by a structural relation or by a semantic relation,

  Then if $SV_1 > SV_2$ then eliminate Hp_2 otherwise eliminate HP_1.

- Filter 3: in this filter, a validated mapping allows to eliminate a candidate mapping which is not consistent with this mapping and with another candidate mapping (Figure 6, right):

  If $\exists$ a candidate mapping $Hp_1 < Rel, c, d, Conf_1, SV_1 >$ and $\exists$ a validated mapping $Map < Rel, c', d', Conf, SV >$ such as:

c and c' are concepts of $O1$ and d and d' are concepts of $O2$ where c is in relation with c' and d is in relation with d' by a structural relation or semantic relation, then if $\exists$ a candidate mapping $Hp_2 < Rel, c, d'', Conf_2, SV_2 >$ with d'' a concept of $O2$, eliminate Hp_2.

- Filter 4: we defined a similarity threshold and a confidence threshold below which the candidate mappings are not considered. The value of these two thresholds could be modified by the user via a specific interface.

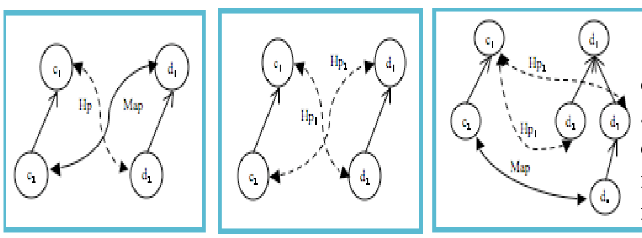

Figure 6: Examples of situations where candidate mappings could be filtered (eliminated) because of inconsistencies

5.5. Validation process

The validation step is an important phase in the mapping process. It involves the human in order to validate or invalidate the mappings generated by the system.

The validation process is characterised by its interactive aspect. It exploits the user feedback as well as its satisfaction with the proposed results. If the user is not satisfied, the mapping process could be restarted using other parameter values and/or other matchers, in order to improve the results.

In all cases, validated mappings are stored and continuously used in the process mapping for improving the results. As described above, the two steps of mappings generation and mappings filtering use validated mappings (if there are). Therefore, when there are new validated mappings, these two steps are restarted (see Figure 7).

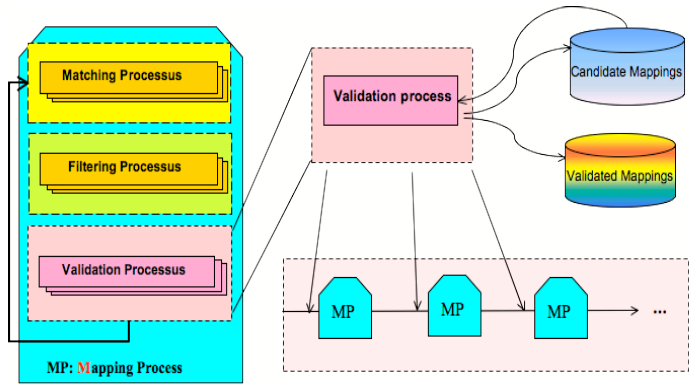

Figure 7: Validation process

Mapping results are validated with different levels of certainty and reliability. For each mapping is associated a parameter V which is incremented whenever a new user validates it. This parameter is incremented by a value comprised between 0 and 1, representing the expertise level of the user (set by the user itself). Thus, more a mapping has a high value V, more it has a high reliability.

6. ROMIE applied to TEL

6.1. Application domain TEL

We consider as application the context of Technology Enhanced Learning (TEL). Ontologies offer a great potential in higher education providing in particular the sharing and reusing of information across educational systems and enabling intelligent and personalised learner support. The increased functionality that ontologies imply will bring new opportunities to e-learning. Learners will be able to interact with distant educational systems easily and in a personalised way. An overview of ontologies for education field and an initial report on the development of an ontology-driven web portal O4E are presented in [7]. In addition, the number of ontologies in TEL domain is growing considerably. An urgent need in TEL is the discovery of suitable resources and the organisation of those resources to perform a learning task.

Ontology-driven TEL has concept-based representation of the specific subject domain and learning resources indexed by domain concepts.

6.2. TEL resource model

We have defined a resource model in order to give a semantic description of a resource. A pedagogic resource is a component that should be described by a set of metadatas when it is added to the system. This description allows to find, to compose and to adapt it. A component is associated to one or several concepts of the domain model. In our model describing pedagogic resources, we distinguish two types of metadatas: (i) the educative characteristics (author, language, media type) using LOM standard (DIRE CE QUE C'EST, DONNER REF), and (ii) the semantic associated to the resources. The description of this semantic is divided into three parts: the prerequisites, the content and the acquisition function (function allowing to evolve the model of the learner). Each of these parts references concepts of the domain model. A resource "SQL query" for instance has as prerequisite the concept "relational Algebra", as content the concept "SQL" and as acquisition function the modification of the learner model by associating to the concept "SQL" a "high" level.

The prerequisite of a resource is described by the triplet *(concept, role, level)*, where the *concept* belongs to the domain model, the *role* indicates for which aspects of the concept this resource is concerned (for instance "introduction", "definition", "description", "application") and the *level* specifies the level of the resource difficulty ("low", "medium", "high"). The content of a resource is described with the couple *(concept, role)*. The acquisition function indicates which triplet *(concept, role, level)* will be added to the learner model, expressing the satisfaction and validation state of the learner knowledge.

We consider that a resource is any digital object like a set of web pages, a file or a program (a simulator for example) in any format (e.g., text, video, audio). We just suppose that it is a unit accessible via an URI. Each resource is described by metadata (e.g., author, title, language) and is indexed with the concepts of the domain ontology. In other words, its content develops one or more concepts. In addition, each resource may have prerequisites (what is required by the resource) expressed by one or more concepts. For instance, in our application, a resource describing a course of C++ has as content the concept "C++ programming language" and as prerequisites the two concepts "C programming language" and "object-oriented methodology", which means that the C language and the object-oriented methodology must be learned before the C++ language.

6.3. Generation of relations between resources

The analysis of the resource model described above allows us to propose a set of semantic relations between resources.

Let us consider two resources R and R'. We note $Pre(R)$ (respectively $Pre(R')$) the prerequisite of R (respectively R') and $Cont(R)$ (respectively $Cont(R')$) the content of R (respectively R'). The analysis of resource properties allows us to propose a set of relations between resources (listed in Table 1). These relations can be deduced automatically and can have characteristics of symmetry and/or transitivity.

Resource properties	Relation names	Relation characteristics
R is weak-substitutable by R' if $Pre(R) \subset Pre(R')$	Weak-substitution	Transitive
R is equivalent to R' if $Pre(R) = Pre(R')$ and $Cont(R) = Cont(R')$	Equivalence	Symmetric and Transitive
R is part of R' if $Cont(R) \subset Cont(R')$ and $Pre(R) \subset Pre(R')$	Part of	Transitive
R strongly-precedes R' if $Cont(R) = Pre(R')$	Strong-precedence	
R weakly-precedes R' if $Cont(R) \subset Pre(R')$	Weak-precedence	
R is strong-crossed to R' if $Cont(R) = Pre(R')$ and $Cont(R') = Pre(R)$	Strong-Crossing	Symmetric
R is weak-crossed to R' if $Cont(R) \subset Pre(R')$ and $Cont(R') \subset Pre(R)$	Weak-Crossing	
R is more general than R' if $Pre(R) \supset Pre(R')$ and $Cont(R) = Cont(R')$	More general	Transitive
R is more specific than R' if $Pre(R) \subset Pre(R')$ and $Cont(R) = Cont(R')$	More specific	Transitive
R is mismatched to R' if there is no relation between their properties	Mismatch	

Table 1: Relations and properties between resources.

7. ROMIE prototype and results

7.1. Implementation

ROMIE was implemented using the technology of multi-agent system (MAS). We have thus a system completely modular and extensible, where the different components (agents) can run in parallel, communicate and cooperate, in order to achieve results. Each agent has its own behaviour and communicates with its environment (the other agents and the user) by sending information messages, queries or responses.

ROMIE is composed of five agent types: the ontology agents (OA), the matcher agents (MA), the candidate mapping generation agent (MGA), the mapping filtering agents (MFA) and the mapping validation agent (MVA).

The OA agents play the role of intermediate between the user, the data and the mapping system, when the other agents communicate and cooperate in order to achieve the different mapping tasks described above.

The prototype is implemented using the multi-agent platform JADE (Java Agent Development Framework) [17]. The different agents communicate using FIPA-ACL (Foundation for Intelligent Physical Agents) language [14], which is one of the most used languages in MAS.

We used OntoBroker system [30] to manage the ontologies. OntoBroker integrates various input formats of ontologies like RDF(S), F-Logic or OWL. The different methods we developed are implemented with logic rules which make our system easily extensible.

7.2. Results in TEL application

We used two educative repositories where learning objects are indexed with ontologies of the educative domain: SQL [5] and ACM [2]. The ACM/CCS ontology is a classification for computer science domain. It classifies nine main sub-domains organised into sections. In our tests, we considered a part of SQL (30 concepts annotating 120 resources) as local ontology and a part of ACM (two sections: Computer systems organisation section and software section, connected to 100 resources) as distant ontology (DONNER LE NOMBRE DE CONCEPTS DE ACM). The two considered ontologies SQL and ACM are in a first step automatically enriched by semantic relations thanks to the resources.

We performed several tests in order to evaluate the performance of ROMIE in generating mappings between ontologies:

- In a first step, we analysed the impact of the resource-based matchers (Figure 8, left) on the mapping results. In this purpose, we first applied only linguistic and syntactic matchers; then, in the second test, we used the structural matchers to generate more mappings; and in the last test, we used the semantic matchers. Each test is presented separately in order to thoroughly show the importance of each of the methods that ROMIE uses for generating mappings.

- In a second step, we analysed the impact of the resource-based filters (Figure 8, right) in improving the mapping results obtained in the first step. We firstly applied only the structural filters then we added the semantic filters, which use the semantic relations generated during ontology enrichment.

To evaluate the results obtained by ROMIE, we calculated three metrics for each result: percentage of true positive mappings (mappings correctly identified), percentage of false negative mappings (mappings not discovered) and percentage of false positive mappings (wrong mappings).

As shown in Figure 8, the mapping results are more and more improved thanks to the succession of methods of matching and filtering we use. We can see that when we use only linguistic and syntactic matchers, we find more than 50% of the mappings, but 50% of the mappings generated by our system are false. Applying in addition the structural matchers, we obtain more mappings (70%) but more false mappings are also generated. And thanks to the resource-based matchers (semantic matchers), we success to find almost all mappings (95%). Nevertheless, the number of wrong mappings still becomes high. But thanks to structural filters and to resource-based filters (semantic filters), the number of wrong mappings decrease consequently, achieving some 10% on the total number of obtained mappings.

7.3. Results in Biomedical application

ROMIE is a mapping system which could be used on any application domain. Even resources are not available for enriching the considered ontologies, our

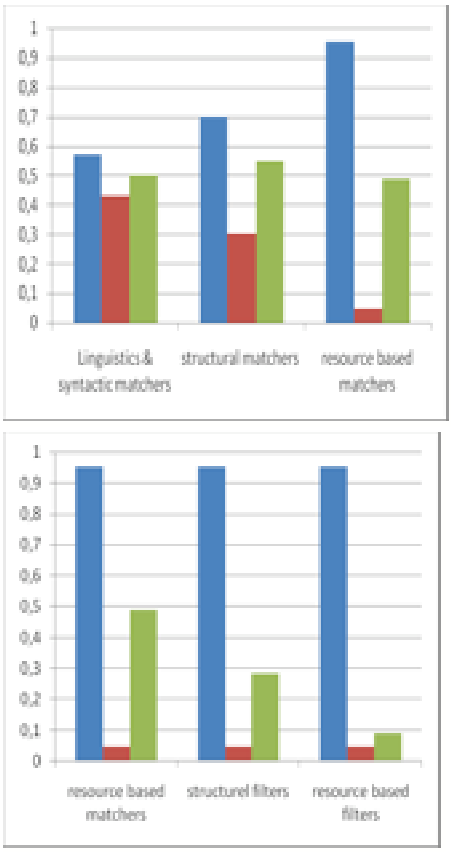

Figure 8: Mapping results obtained by ROMIE on ACM and SQL ontologies. Three metrics are calculated: percentage of mappings correctly identified (in blue), percentage of false negative mappings (in red) and percentage of false positive mappings (in green).

system is able to predict efficiently mappings between the ontologies.

We show in this section results obtained by ROMIE on an example of biomedical ontologies not linked to resources, in order to show the efficiency of ROMIE even when the considered ontologies have not been enriched (nor step of ontology enrichment).

Several biomedical ontologies are available in OBO web site [15]. We present here the results obtained with two ontologies, namely: MeSH (Medical Subject Headings) ontology [25] and MA (Adult Mouse Anatomy) ontology [12]. These ontologies cover a similar anatomy context and are developed independently. MeSH is a controlled vocabulary produced by the American National Library of Medicine and is used for indexing, cataloguing, and searching for biomedical and health-related information and documents. It consists of sets of terms or descriptors in a hierarchical structure and contains more than 1400 concepts. MA organises anatomical structures for the postnatal mouse spatially and functionally, using "is a" and "part of" relationships. The ontology is used to describe expression data for adult mouse and phenotype data pertinent to anatomy in standardised ways. MA ontology contains more than 2400 anatomical concepts. For our experimentation we focused on three categories developed by both ontologies, namely, nose (with 15 concepts in MeSH and 18 concepts in MA), ear (39 concepts in MeSH and 77 concepts in MA), and eye (45 concepts in MeSH and 112 concepts in MA).

All our evaluations are based on the metrics of recall and precision calculated considering mappings generated by the system and mappings identified manually by the expert:

$$Precision = \frac{N_{correct}}{N_{found}}$$

$$Recall = \frac{N_{correct}}{N_{expected}}$$

where $N_{correct}$ corresponds to the number of mappings correctly predicted, N_{found} to the number of mappings generated by ROMIE and $N_{expected}$ to the number of mappings manually identified by the expert (note that $N_{correct}$ is the intersection between the two sets N_{found} and $N_{expected}$).

In our tests, the expert provides 9 mappings between nose concepts, 27 mappings between ear concepts and 27 mappings between eye concepts.

Figure 9 illustrates the impact of each type of matcher and filter on the results.

We can see that the recall obtained by combining all matchers is higher but the precision is less important. The threshold filter allowed to select candidate mappings with confidence and similarity values greater than given thresholds. We considered in our tests a confidence threshold of 5 and a

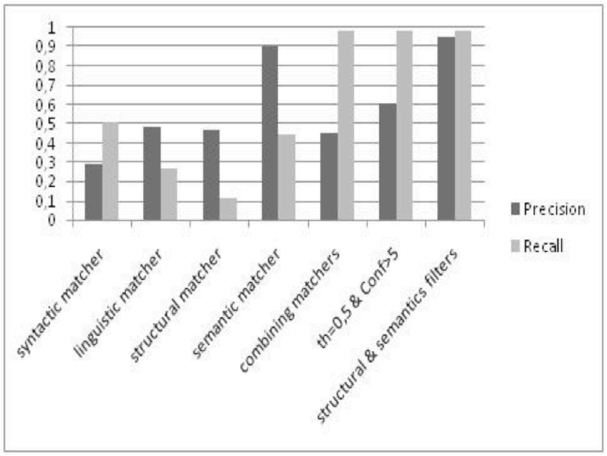

Figure 9: Results obtained by the different ROMIE matchers and filters on MA and MeSH ontologies.

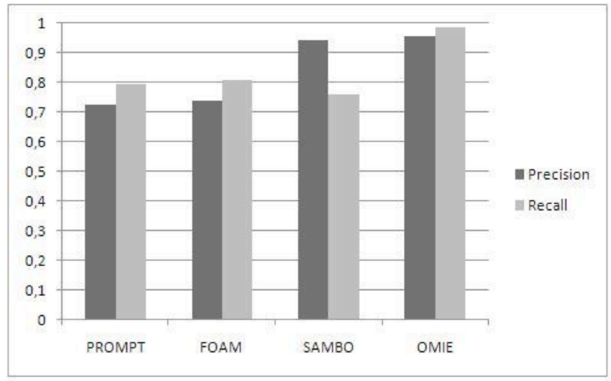

Figure 10: Comparison of mapping results obtained by ROMIE, PROMPT, FOAM and SAMBO on MA and MeSH ontologies

similarity threshold of 0.4. This filter improved slightly the precision of the results. As we can see on the Figure 9, the structural and semantic filters had a much more positive effect in the detection of wrong candidate mappings.

ROMIE was compared to three well known systems: PROMPT [28], FOAM [9] and SAMBO [21]. We compared the results obtained by ROMIE on MeSH and MA ontologies with the ones obtained by Lambrix and Tan in [21], where they evaluated their ontology mapping tool SAMBO with two well known and available tools: PROMPT [28] and FOAM [9]. Figure 10 shows the recall and the precision obtained by each of the four systems (PROMPT, FOAM, SAMBO and ROMIE).

We can see that the precision of ROMIE is higher than the ones of PROMPT and FOAM and is equal to the precision of SAMBO, which is equal to 95%. However, the recall of ROMIE is the higher one (close to 100%), thanks to our efficient filtering process. For instance, ROMIE successfully generated all mappings provided (manually) by the domain expert for the 'eye' concepts.

8. Conclusion

Thanks to the web, users could access to a huge amount of data and resources. In a large number of domains and particularly in TEL and biomedical domain, ontologies are more and more used to allow the exchange of these data and resources between different users. Unfortunately, it is hard to find one global ontology in a given domain. Several ontologies must be considered, that's why the task of mapping ontologies is very important. In this paper, we presented an original and efficient method of ontology mapping called ROMIE (Resource based Ontology Mapping within an Interactive and Extensible environment), which presents several characteristics:

- Extensibility: adding, removing or modifying the initial system configuration (matchers and thresholds) allows to fit to the features of the application domain. The system is relatively generic because its weak dependency in relation to the semantics of the used presentation language.
- Adaptability: mapping discovery is based on relations between resources indexed to ontologies (instances, documents, etc.). No need to have a rich ontology, it is enough to have enough indexed resources so that method can be applied.

- Flexibility: several mapping methods are used and combined (syntactic, linguistic, structural and semantic). This allows to produce mappings richer in terms of semantics than mapping methods based only on measurement of similarities.

- Interactivity: the process of mapping validation uses and exploits knowledge of the users.

We proposed a mapping algorithm based on resources which improves consequently the mapping process. The principle is to enrich, in a first step, the ontology by semantic relations between concepts, by propagating relations existing between resources relied to the concepts. Resources relationships could be generated by an analysis of the resources properties which depend on the domain application.

In this paper, our first objective was to show how semantic matchers based on resource relationships improve mapping results. We considered for this purpose the TEL application domain, where we defined a set of resource relationships which allowed to enrich considered ontologies. We have also shown the efficiency of our mapping system even resources are not available (so no enrichment of the considered ontologies), considering the biomedical application domain.

Our instance-based approach for ontology mapping is different than instance-based approaches existing in literature. Almost existing instance based approaches assume that the instance level is shared between ontologies. Therefore, the instance-based mapping between concepts of two ontologies is based on the overlap of their instance sets. But in many situations there will be no shared individuals. In ROMIE, this is not required, since it is only necessary to have semantic relations between the instances or resources associated to considered concepts.

Besides the enrichment step, ROMIE presents several properties which makes it efficient comparing to other existing systems. One of these properties is its filtering method, where wrong generated mappings are determined automatically. The filtering method uses several rules including incompatibility between mappings. In existing works, filtering process is generally based only on a threshold value. Another property is the validation step, which is interactive and continuous, influencing continuously the mapping process. The validated mappings are indeed continuously taken into account and used by the different steps and methods of the mapping process (matchers, filters) in order to improve mapping results.

There are of course several perspectives, extensions and improvements to our work:

- We presented in this paper a list of semantic relations between resources for the TEL application. We are currently working on defining other relations.

- Our ontology mapping system is also in use in bioinformatics and life sciences domain (particularly in Saphir ANR Project [33, 34]). In this paper we presented results obtained by using ROMIE without the resource-based matchers. We currently study the possibility of using resources, in order to enrich ontologies and then improve the mapping process.

- Other lexical dictionaries can be used, instead or besides WordNet dictionary. This is the case for life sciences ontologies for instance, where we plan to use UMLS (Unified Medical Language System) [29], which is a well-known metathesaurus for biomedical terminology.

- Our different matchers and filters use confidence and similarity thresholds which are set manually in our prototype and deduced thanks to several tests we achieved. Users can change these threshold values and set their own thresholds via the interface. One of our further works is to develop a method based on statistical learning in order to set automatically adequate values to these thresholds.

- To generalise to other applications, we propose to extract dependency relations between instances (like 'part of', 'summary', 'antithesis' or 'purpose') using for example Rhetorical Structure Theory [32].

References

[1] Ensembl data source. www.ensembl.org.

[2] ACM. Acm computing classification system: Advancing computing as a science and profession. http://www.acm.org/.

[3] R. Albertoni and M. De Martino. Semantic similarity of ontology instances tailored on the application context. In *ODBASE*, 2006.

[4] P. Bernstein, S. Melnik, and P. Mork. Interactive schema translation with instance-level mappings. In *VLDB Conference*, 2005.

[5] A. Bouzeghoub, B. Defude, F. Duitama, and C. Lecocq. A knowledge-based approach to describe and adapt learning objects. *International Journal on E-Learning. Special Issue: Learning Objects in Context*, 5(1):95–102, 2006.

[6] H. Chalupsky. OntoMorph: A translation system for symbolic knowledge. In *Principles of Knowledge Representation and Reasoning Proceedings of the Seventh International Conference*, pages 471–482, 2000.

[7] D. Dicheval, S. Sosnovsky, T. Gavrilova, and P. Brusilovsky. Ontological web portal for educational ontologies. In *International Workshop on Applications of Semantic Web technologies for E-Learning (SW-EL)*, 2005.

[8] AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Y. Halevy. Learning to map between ontologies on the semantic web. In *International World Wide Web conference*, pages 662–673, 2002.

[9] Marc Ehrig and York Sure. FOAM - framework for ontology alignment and mapping - results of the ontology alignment evaluation initiative. In *Workshop on Integrating Ontologies*, volume 156, pages 72–76, 2005.

[10] J. Euzenat, T. Le Bach, J. Barrasa, P. Bouquet, J. De Bo, R. Dieng-Kuntz, M. Ehrig, M. Hauswirth, M. Jarrar, R. Lara, D. Maynard, A. Napoli, G. Stamou, H. Stuckenschmidt, P. Shvaiko, S. Tessaris, S. Van Acker, and I. Zaihrayeu. State of the art on ontology. Deliverable 2.2.3, IST Knowledge web NoE, Knowledge web NoE, June 2004.

[11] J. Euzenat and P. Valtchev. Similarity-based ontology alignment in owl-lite. In *ECAI*, pages 333–337, 2004.

[12] T.F. Hayamizu, M. Mangan, J. Corradi, J. Kadin, and M. Ringwald. The adult mouse anatomical dictionary: a tool for annotating and integrating data. *Genome Biology*, 6(3), 2005.

[13] http://wordnet.princeton.edu. Wordnet a lexical database for the english language.

[14] http://www.fipa.org/specs/fipa00037. FIPA ACL communicative act library specification.

[15] http://www.obofoundry.org/. The open biological and biomedical ontologies, 2009.

[16] A. Isaac, van der Meij L., S. Schlobach, and S. Wang. An empirical study of instance-based ontology matching. In *ISWC+ASWC*, 2007.

[17] Jade. Java agent development framework. http://jade.tilab.com/.

[18] Y. Kalfoglou and M. Schorlemmer. Ontology mapping: the state of the art. *The Knowledge Engineering Review*, 18(1):1–31, 2003.

[19] Toralf Kirsten, Andreas Thor, and Erhard Rahm. Instance-based matching of large life science ontologies. In *DILS*, pages 172–187, 2007.

[20] K. Kotis, A. Valarakos, and George Vouros. Automs: Automating ontology mapping through synthesis of methods. In *OAEI (Ontology Alignment Evaluation Initiative) contest, Ontology Matching International Workshop*, 2006.

[21] P. Lambrix and H. Tan. SAMBO–a system for aligning and merging biomedical ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(3):196–206, 2006.

[22] P. Lambrix, H. Tan, and Q. Liu. SAMBO and SAMBOdtf results for the ontology alignment evaluation initiative 2008. In *OM*, 2008.

[23] Deborah L. McGuinness, Richard Fikes, James Rice, and Steve Wilder. The chimaera ontology environment. In *AAAI/IAAI*, pages 1123–1124, 2000.

[24] G. A. Miller. Wordnet: A lexical database for english). *Communications of the ACM*, 38(11):39–41, 1995.

[25] National Library of Medicine. Medical Subject Headings (MeSH). http://www.nlm.nih.gov/mesh.

[26] N. NOY, R. W. FERGERSON, and M. A. MUSEN. The knowledge model of protg2000 : combining interoperability and flexibility. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 17–32, 2004.

[27] N. F. Noy. Semantic integration: A survey of ontology-based approaches. *SIGMOD Record*, 33(4), 2004.

[28] Natalya Fridman Noy and Mark A. Musen. PROMPT: Algorithm and tool for automated ontology merging and alignment. In *AAAI/IAAI*, pages 450–455, 2000.

[29] National Library of Medicine. UMLS: Unified medical language system, 2009. http://www.nlm.nih.gov/research/umls/.

[30] Ontoprise. OntoBroker user guide, 2009. http://www.ontoprise.de.

[31] L. Serafini and A. Tamilin. Instance migration in heterogeneous ontology environments. In *ISWC+ASWC*, 2007.

[32] M. Taboada and W. C. Mann. Rhetorical structure theory: Looking back and moving ahead. *Discourse Studies*, 8(3):423–459, 2006.

[33] S. Thomas, E. Abdulhay, P. Baconnier, J. Fontecave, J.P. Francoise, F. Guillaud, P. Hannaert, A. Hernandez, V. Le Rolle, P. Maziere, F. Tahi, and F. Zehraoui. Saphir - a multi-scale, multi-resolution modeling environment targeting blood pressure regulation and fluid homeostasis. In *Conf Proc IEEE Eng Med Biol Society*, pages 6649–52, 2007.

[34] S.R. Thomas, P. Baconnier, J. Fontecave, J.-P. Franoise, F. Guillaud, P. Hannaert, A. Hernandez, V. Le Rolle, P. Maziere, F. Tahi, and R. White. SAPHIR: a physiome core model of body fluid homeostasis and blood pressure regulation. *Philosophical Transactions of the Royal Society A*, 13(366):3175–97, 2008.

[35] A. G. Valarakos, V. Spiliopoulos, and and-Vouros G. Kotis, K. Automs-f: A java framework for synthesizing ontology mapping methods. In *I-KNOW*, 2007.