

# An Asymmetric Approach to Discover the Complex Matching Between Ontologies

Fatma Kaabi<sup>1(✉)</sup> and Faiez Gargouri<sup>2</sup>

<sup>1</sup> Laboratory MIRACL, Faculty of Economic Sciences and Management, Sfax, Tunisia

kaabifatma@yahoo.fr

<sup>2</sup> Laboratory MIRACL, The Higher Institute of Computer Science and Multimedia of Sfax, BP 242, Sakiet Ezzit, 3021 Sfax, Tunisia

**Abstract.** This paper introduces an extensional and asymmetric alignment approach capable of identifying complex mappings between OWL ontologies. This approach employ the association rule to detect implicative and conjunctive mapping containing complex correspondences. Method for extracting the complex mappings is presented and results of experiments carried out on the large biomedical ontologies and the anatomy track available to Test library of Ontology Alignment Evaluation Initiative show the efficiency of the approach proposed.

## 1 Introduction

Ontology mapping is a well studied problem, several matching approaches have been proposed [10]. These methods aim at finding correspondences between the semantically related entities of those ontologies. From this approaches we can identify: the extensional approaches, and the intentional approaches. The majority of these methods finds only equivalence relations (CIDER-CL [6], YAM++ [7], LogMap [8]) and do not consider also the asymmetric relations like the subsumption. Most of the proposed approach are symmetrical and intentional. The only extensional and asymmetric method is the AROMA method [11]. Therefore, this method discovers only simple relationships.

Most existing matching approach concentrates on finding 1-1 mappings between two given ontologies. However, complex mappings are very useful in practice. Simple correspondences are not sufficient to express relationships that represent correspondences between entities since it (1) may be difficult to discover simple correspondences (or they do not exist) in certain cases, or (2) simple correspondences do not allow for expressing accurately relationships between entities.

As a motivating example, consider two ontologies  $\mathcal{O}_1$  and  $\mathcal{O}_2$  (Fig. 1) describing cell types.  $\mathcal{O}_1$  is a part of the ontology CL<sup>1</sup> and  $\mathcal{O}_2$  is an extract of the ontology BCGO<sup>2</sup>. The proposed methods [7, 8] can't find the most similar entity

<sup>1</sup> Cell Ontology (CL), <http://www.cellontology.org/>.

<sup>2</sup> Beta Cell Genomics Ontology, <https://github.com/obi-bcgo/bcgo>.

node in  $\mathcal{O}_2$  that maps to the entity node *phagocyte* in  $\mathcal{O}_1$ . But, the entity *phagocyte* can match to the intersection of the three entities *motile cell*, *native cell* et *stuff accumulating cell* dans  $\mathcal{O}_2$ . The terms describing the *phagocyte* concept are belonged in the dataset of the three concepts *motile cell*, *native cell* and *stuff accumulating cell*.

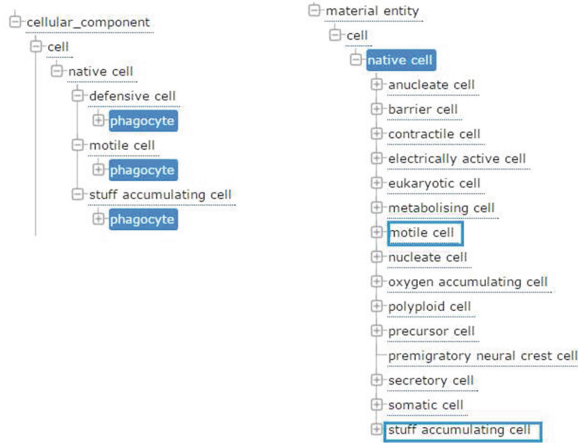


Fig. 1. Ontology O1 and ontology O2

The rest of this paper is organized as follows: First we review related work and we illustrate the limitations of existing complexes mapping approaches. Next we introduce the proposed method. Finally we present experimental results and concluding remarks.

## 2 Related Work

In order to find complex correspondences some approaches have been considered. We present in the following the most interesting ones. Doan and colleagues [12] developed a system **CGLUE** that uses machine learning techniques to semi-automatically generate semantic matching. This system finds disjunctions and equivalence relations between concepts. It finds complex matching between taxonomies. CGLUE is based on the notions of semantic similarity, expressed in terms of the joint probability distribution of the concepts involved. This system calculates the joint distribution of the concepts and use the joint distribution to compute any appropriate similarity measure.

CGLUE is based on the assumption that the children of any ontology entity are mutually exclusive and exhaustive. We note that the assumption maintains for many real ontology, in which the further specialization of an entity usually gives a partition of the instances of that entity. However, in many real ontologies,

very sibling entities share instances. Hence, for these domains this approximating assumption is not hold.

The two approaches [13,14] are based on the inductive logic programming, ILP, and attempts at creating alignments by using the learning theory. These approaches take complex correspondences into account and not only equivalence correspondences. But here it is not possible to create complex mappings without learning correspondences out of instances. Often ontologies do not contain any instances. Hence the learning theory cannot be applied in order to find complex correspondences in ontologies without instances.

**The pattern-based ontology matching approach** presented in [15] define patterns to discover automatically complex correspondences. A master alignment of ontologies is necessary. To detect these correspondences, a set of simple conditions must be satisfied for each model. These conditions are a combination of structural, linguistic techniques and types compatibility. The defined models are (notice that the notation  $i\#C$  is used to assign to an entity  $C$  from ontology  $O_i$ ):

1. *CAT* (Class by Attribute Type Pattern): this model detects correspondences as  $1\#A \equiv \exists 2\#R.2\#B$ ;
2. Class by Inverse Attribute Type Pattern ( $CAT^{-1}$ ): this model allows correspondences as which are written as  $1\#A \equiv 2\#B \cap 2\#R_1.T$ , to be detected;
3. *CAV* (Class by Attribute Value Pattern): this model detects correspondences as  $1\#A \equiv \exists 2\#R.\{\dots\}$ , (where  $\{\dots\}$  is a set of concrete data values)
4. *PC* (Property Chain Pattern): this model allows correspondences as  $1\#R \equiv 2\#P \circ 2\#Q$ .

This method can find a lot of complex correspondences. However, the used patterns cover only peculiar domains of ontologies.

After analysing these approaches, we note that the above mentioned methods only consider the equivalence relations between concepts and do not take into account the asymmetric relations such as the subsumption. To overcome these significant limitation, we have developed a new complex mapping methodology named ARCMA [4,5] (Association Rules Complex Matching Approach) which permit to map subsumption relations between entities.

### 3 A New Method for Complex Matching

The alignment method ARCMA [4], aims at finding complex correspondences between two OWL ontologies.

ARCMA follows three consecutive steps: (1) the term or data sets extraction (The pre-processing step), (2) the detection of association rules between entities of two ontologies and (3) the post-processing of results.

In the pre-processing step, a set of relevant terms embedded in the descriptions and entities instances is generated by using a natural language processing tools. We represent the entities (concepts and proprieties) by set of terms and data generated from their description and instances. We extract the name and

the terms contained in the annotations (labels, comments, etc.). We also add the local name, the annotations and the values of its instances [5].

In the second step, ARCMA detects the complex matching between two OWL ontologies using the association rule model and a statical measure, the implication intensity [2]. A valid association rule  $x \rightarrow y_1 \wedge \dots \wedge y_i \dots \wedge y_n$  means that the vocabulary associated to a source entity  $x$  aims to be included in the intersection of the relevant terms of set of entities  $y_i$ . For example, the valid rule *phagocyte*  $\rightarrow$  *motile cell*  $\wedge$  *native cell*  $\wedge$  *stuff accumulalating cell* could be interpret: The entity *phagocyte* corresponds to intersection of the three entities *motile cell*, *native cell* and *stuff accumulalating cell*. The post-processing eliminates the redundancies in matcher found.

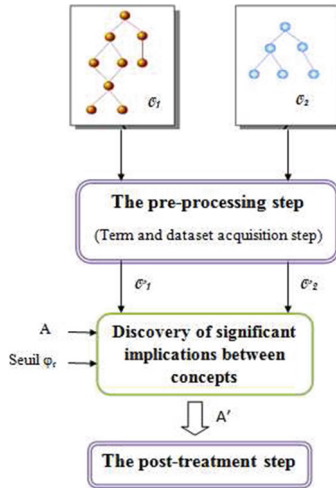


Fig. 2. The ARCMA process

Figure 2 illustrate the process of our method to discover the complex mappings between OWL ontologies. First, we use two OWL multiple inheritance ontologies. Then we apply a pretreatment process to define their relationship on a common extension. We also consider a reference alignment between these two ontologies. Next, we utilize the association rules to find complex correspondences type  $x \Rightarrow y_1 \wedge \dots \wedge y_i \dots \wedge y_n$ . Finally, we reduce the redundancy in the extracted rule set. A rule will be selected if none of its generative rules have a value of the implication intensity ( $\varphi$ ) greater than or equals to its  $\varphi$  value.

## 4 Evaluation

To estimate the performance of our approach, a prototype is realized in Java. Our system supports input two OWL ontologies and a reference alignment, then comparing the correspondence obtained by our tool and those by a manual mapping.

This evaluation is carried out by exploiting the two metrics alignment quality: precision and recall [16]. Precision measures the ratio of correctly found correspondences over the total number of returned correspondences. Recall compute the ratio of correctly found correspondences over the total number of expected correspondences.

The experiment is performed on the large biomedical ontologies and the anatomy track available to Test library of Ontology Alignment Evaluation Initiative OAEI<sup>3</sup>. The Large Biomedical track contains the mapping of FMA (78,989 classes), NCI Thesaurus (66,724 classes) and SNOMED CT (306,591 classes) and uses the UMLS Metathesaurus as the basis for the track’s reference mappings. The reference mappings only include subsumption and equivalence relations between classes. The track consists of three matching problems: FMA-NCI, FMA-SNOMED CT and SNOMED CT-NCI. The anatomy track includes the mapping of the two ontologies Adult Mouse Anatomy (AM) and part of the NCI thesaurus describing human anatomy. The reference mapping includes only equivalence correspondences between classes.

Our method ARCMA requires that the source ontology supports multiple inheritances. Among the Large Biomedical track and the anatomy track, there are only three ontologies containing multiple inheritances (SNOMED, AM and human). Hence, we will exploit these last ontologies and two references alignments: SNOMED CT-NCI and reference. The characteristics of these ontologies are shown in Table 1.

**Table 1.** Description of the ontologies used for the evaluation of ARCMA

Ontologies	Classes	Properties
Large SNOMED	122464 (40 % SNOMED)	55
Small SNOMED_fma	13412 (5 % SNOMED)	18
Small SNOMED_nci	51128 (17 % SNOMED)	63
Whole NCI	66724	190
Small NCI_fma	6488 (10 % NCI)	63
Small NCI_snomed	23958 (36 % NCI)	83
mouse	2744	3
human	3304	2

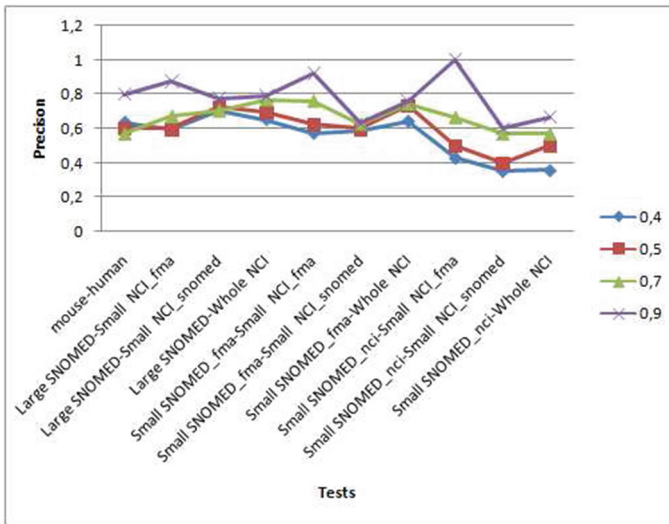
The Table 2 illustrates the results obtained by the alignment method ARCMA, with the rule selection threshold  $\varphi_r = 0,9$ .

In this table we note that in some tests such as *SmallSNOMED\_nci* – *SmallNCI\_fma*, the value of precision is 1, that means that the results of our method are the same given by an expert, and for the many other tests, the precision value is higher than 0.75, therefore, our system gives good results which are

<sup>3</sup> Ontology Alignment Evaluation Initiative Test library (2015), <http://oaei.ontologymatching.org/2015/seals-eval.html>.

**Table 2.** Performance measures of ARCMA

Tests	Precision	Recall
mouse-human	0,8	0,571
Large SNOMED-Small NCI_fma	0,844	0,776
Large SNOMED-Small NCI_snomed	0,813	0,765
Large SNOMED-Whole NCI	0,808	0,778
Small SNOMED_fma-Small NCI_fma	0,927	0,731
Small SNOMED_fma-Small NCI_snomed	0,729	0,714
Small SNOMED_fma-Whole NCI_whole	0,811	0,860
Small SNOMED_nci-Small NCI_fma	1	0,667
Small SNOMED_nci-Small NCI_snomed	0,6	0,429
Small SNOMED_nci-Whole NCI	0,667	0,286

**Fig. 3.** Values of precision as a function of the threshold value  $\varphi_r$ 

encouraging. For example, ARCMA discovered the following meaningful implications from SNOMED\_small\_overlapping\_nci to NCI\_small\_overlapping\_fma:

R1.CENTRAL\_NERVOUS\_SYSTEM\_TRACT\_STRUCTURE→CENTRAL\_NERVOUS\_SYSTEM\_PART AND NERVE

R2.DUODENAL\_PAPILLA\_STRUCTURE→BILIARY\_TRACT AND DUODENUM AND PANCREATIC\_DUCT

R3.COLONIC\_MUSCULARIS\_PROPRIA\_STRUCTURE→COLON AND MUSCULARIS\_PROPRIA

Figures 3 and 4 show the influence of rule selection threshold  $\varphi_r$  on the precision and recall of ARCMA. We note that the value of the precision increases

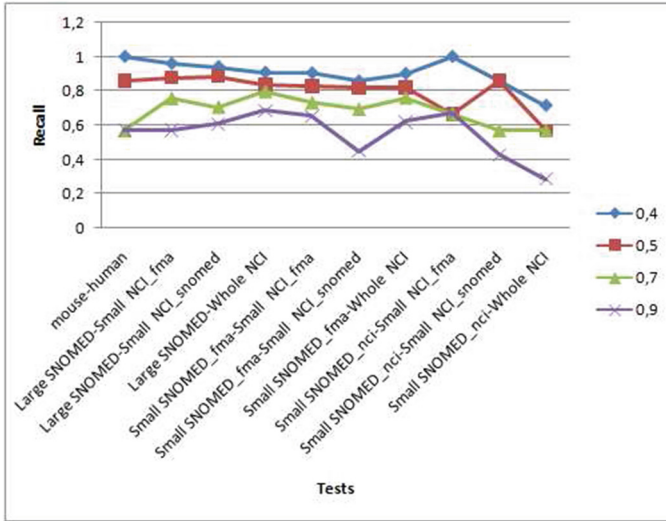


Fig. 4. Values of recall as a function of the threshold value  $\varphi_r$ .

with the higher level of the threshold. This phenomenon clearly shows a correlation between the deviation from independence situation and the relevance of rules. In general, we can conclude that ARCMA achieved a good precision/recall values. The high recall value can be explained by the fact that UMLS thesaurus contains definitions of highly technical medical terms.

## 5 Conclusion

In this paper, we proposed a new approach for discovering complex mappings between two OWL ontologies. We utilized the association rule and the statical measure, the implication intensity, to detect implicative and conjunctive mapping containing complex correspondences. We implemented the approach and experimentally evaluated it on the large biomedical ontologies and the anatomy track, which demonstrated the high precision of the discovered correspondences. The principal advantage of this approach is that it is simple. Besides, the use of the implication intensity measure permit to approve the validity of the complex correspondences and justifies the good precision values obtained by ARCMA.

## References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: The Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216 (1993)
2. Blanchard, J., Kuntz, P., Guillet, F., Gras, R.: Implication intensity: from the basic statistical definition to the entropic version, chap. 28, pp. 473–485. CRC Press (2003)

3. Do, H., Rahm, E.: A system for flexible combination of schema matching approaches. In: *The International Conference on Very Large Data Bases (VLDB 2002)*, pp. 610–621 (2002)
4. Kaâbi, F., Gargouri, F.: An approach to find complex matching between conceptual hierarchies. In: *Proceedings of the 2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2012)*, pp. 205–210. IEEE Computer Society, Washington, DC (2012)
5. Kaabi, F., Gargouri, F.: A new approach to discover the complex mappings between ontologies. *Int. J. Web Sci.* **1**(3) (2012)
6. Gracia, J., Asooja, K.: Monolingual, cross-lingual ontology matching with CIDER-CL: evaluation report for OAEI 2013. In: *Proceedings of the 8th International Conference on Ontology Matching*, vol. 1111, pp. 109–116. CEUR-WS.org (2013)
7. Nago, D.D., Bellehsene, Z.: YAM++: a multi-strategy based approach for ontology matching task. In: *Proceedings of the 8th International Conference on Ontology Matching, OM 2013*, vol. 1111, pp. 211–218. CEUR-WS.org (2013)
8. Jiménez-Ruiz, E., Grau, B.C., Solimando, A., Cross, V.V.: LogMap family results for OAEI 2015. In: *Proceedings of the 10th International Workshop on Ontology Matching collocated with the 14th International Semantic Web Conference (ISWC 2015)*, Bethlehem, PA, USA, 12 October 2015, pp. 171–175 (2015)
9. Kalfoglou, Y., Schorlemmer, M.: Ontology mapping: the state of the art. *Knowl. Eng. Rev.* **18**(1), 1–31 (2003)
10. Euzenat, J., Shvaiko, P.: *Ontology Matching*. Springer, Heidelberg (2007)
11. David, J., Guillet, F., Briand, H.: Association rule ontology matching approach. *Int. J. Semant. Web Inf. Syst.* **3**(2), 27–49 (2007)
12. Doan, A., Madhavan, J., Dhamankar, R., Domingos, P., Halevy, A.: Learning to match ontologies on the semantic web. *VLDB J.* **12**(4), 303–319 (2003)
13. Stuckenschmidt, H., Preu, L., Meilicke, C.: Learning complex ontology alignments a challenge for ILP research. In: *Proceedings of the 18th International Conference on Inductive Logic Programming* (2008)
14. Qin, H., Dou, D., LePendu, P.: Discovering executable semantic mappings between ontologies. In: Meersman, R., Tari, Z. (eds.) *OTM 2007, Part I. LNCS*, vol. 4803, pp. 832–849. Springer, Heidelberg (2007)
15. Ritze, D., Meilicke, C., Zamazal, O.S., Enschmidt, H.S.: A pattern-based ontology matching approach for detecting complex correspondences. In: *Proceedings of the ISWC 2009 Workshop on Ontology Matching* (2009)
16. Euzenat, J.: Semantic precision and recall for ontology alignment evaluation. In: *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad (IN), pp. 348–353. AAAI Press, Menlo Park (CAUS) (2007)