# Ontology Alignment based on Instances using Hybrid Genetic Algorithm

Alex Alves, Kate Revoredo, Fernanda Baião

Research and Practice Group in Information Technology (NP2Tec)
Department of Applied Informatics – Federal University of the State of Rio de Janeiro (UNIRIO)
{alex.alves,katerevoredo,fernanda.baiao}@uniriotec.br

**Abstract.** The popularity of Ontology favored the appearance of several Ontologies to the same domain, thereby increasing the need of alignment techniques. In scenarios where ontologies comprising instances, the knowledge embedded in these instances can be useful to improve alignment. This paper extends a hybrid evolutionary approach, which applies a local optimization method, by taking instances into account in order to reduce premature convergence and, consequently, improve the quality of the resulting ontology alignment.

## 1    Introduction

Ontology is an explicit formal specification of the concepts in a domain and relations among them. In the existence of two or more ontologies for the same domain, there is the need of finding the correspondences between them. This task is known as *ontology alignment* [Shvaiko and Euzenat 2007] and is accomplished by evaluating the elements of the two ontologies, trying to find the best pair of corresponding elements. Finding this correspondence is not an easy task, especially in domains with many concepts and relations, where scalable approaches maybe necessary. In [Acampora et. al. 2012], ontology alignment has been formulated as an optimization problem applying a genetic algorithm with local search heuristics.

On the other hand, in scenarios where ontologies contain instances, the knowledge embedded in these instances can be useful to improve the alignments. Therefore, the ontology elements that may be considered for the alignment comprise its concepts, relations, or instances [Shvaiko and Euzenat 2007]. The alignment approach proposed in [Acampora et. al. 2012] considered only the first two elements. In this paper, we delineate ideas towards extending their approach by also considering instances.

## 2    Ontology Alignment based on Instances and through Genetic Algorithms

Genetic algorithms try to solve an optimization problem (or search) by manipulating a population of potential solutions that reproduces the process of natural evolution. Specifically, they operate on encoded representations of the solutions, called chromosomes, an equivalent representation of an individual feature in nature. The evolution algorithm starts from a population of individuals randomly generated and creates successive generations. At each generation, a natural selection process takes place, providing a mechanism for selecting the best solution to survive. Each solution is evaluated by means of a fitness function and compared to other solutions in the population. The higher a fitness value of an individual is the greater will be its chances of surviving. When creating a new generation, the recombination of genetic material among individuals of a generation applies two operators: crossover (which exchanges portions between two randomly selected chromosomes) and mutation (which causes random alteration of the genes of one chromosome). The evolution algorithm terminates when some specified conditions are reached [Acampora et. al. 2012].

In our approach, as in [Acampora et. al. 2012], a genetic algorithm is applied to solve the ontology-alignment problem. A chromosome corresponds to a potential alignment between two given ontologies ($O^1$ and $O^2$), and is represented by an integer vector $A = (e_0, \ldots, e_{n-1})$ such that, when the i-th position of A has a value of $e_j$, this means the alignment between the $e_i$ element from $O^1$ and the $e_j$ element from $O^2$, that is, the correspondence ($e_i$, $e_j$). The length of A is given by the number of the elements of $O^1$. Figure 1 illustrates an example of a fictitious alignment between the elements from two ontologies in a car domain (where each double-ended arrow connects a pair of corresponding elements) and the chromosome representing this alignment.
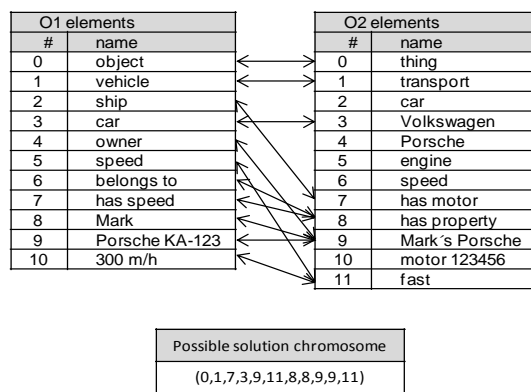
| O1 elements | | | O2 elements | |
| --- | --- | --- | --- | --- |
| # | name | | # | name |
| 0 | object | | 0 | thing |
| 1 | vehicle | | 1 | transport |
| 2 | ship | | 2 | car |
| 3 | car | | 3 | Volkswagen |
| 4 | owner | | 4 | Porsche |
| 5 | speed | | 5 | engine |
| 6 | belongs to | | 6 | speed |
| 7 | has speed | | 7 | has motor |
| 8 | Mark | | 8 | has property |
| 9 | Porsche KA-123 | | 9 | Mark´s Porsche |
| 10 | 300 m/h | | 10 | motor 123456 |
| | | | 11 | fast |

Possible solution chromosome

(0,1,7,3,9,11,8,8,9,9,11)

**Fig. 1.** A possible alignment between O$_1$ and O$_2$ ontologies
and its chromosome representation (adapted from [Acampora et al. 2012])

In order to take instances into account during the alignment problem, we propose an additional function. This function applies the concept of upPropagation [Massmann et al. 2011], in which the similarities between instances are propagated to their concepts when evaluating a possible solution. Moreover, our approach will initially adopt specific values for the genetic parameters, following the work of Souza [2012]: a selection rate of 50%, crossover probability of 80%, mutation probability of 10%, a 30% rate for reinsertion of best individuals, a 10% rate for reinsertion of the worst individuals, mortality of 5 generations, a local search frequency of every 100 generations and, finally, a 25% insertion neighborhood. By adopting this parameter value set, avoid solutions that persist for many generations, like super individuals or solutions very bad, is applied the concept of mortality in the population. Individuals that reach a certain age m are dropped from the new generation. Finally, we assume the existence of a reference alignment and predefined thresholds for precision, recall and F-measures as our stopping criteria.

## 3    Conclusion

Typically, ontologies are used by people, artificial agents and distributed applications that need to share domain information about a specific subject or area of knowledge. However, the creation of these ontologies is commonly performed in accordance with local needs and often without concern for reuse. In an ever-increasing frequent scenario where various ontologies for the same domain exist, alignment of them is a must, but still remains as a challenging problem. In many of these scenarios, instances may potentially bring extra information helping the alignment process, but are currently under-exploited in the literature, especially when combined with other approaches. In this paper, we propose to use instances to improve the alignment of ontologies through the use of a genetic algorithm combined with a local search heuristic to reduce premature convergence. Experiments are being performed to evaluate our proposal.

## References

Acampora, G., Loia, V., Salerno, S. and Vitiello, A. (2012), "A hybrid evolutionary approach for solving the ontology alignment problem", In: International Journal of Intelligent Systems, 27:189–216. doi:10.1002/int.20517.

Euzenat, J. and Shvaiko, P. (2007), "Ontology Matching", Springer-Verlag, Berlin Heidelberg. 2007, X, 334 p. 67 illus. ISBN 978-3-540-49611-3.

Massmann, S.; Raunich, Salvatore; Aumueller, David; Arnold, Patrick; Rahm e Erhard. (2011)" Evolution of the COMA Match System", OM-2011 (The Sixth International Workshop on Ontology Matching, October 24th, 2011, Bonn, Germany).

Souza, Jairo Francisco. (2012) "A heuristic approach single-objective for calibration in meta-ontology aligners". Rio de Janeiro, 105p. PhD Thesis, Department of Computer Science, Pontifícia Universidade Católica do Rio de Janeiro.