# OntoAlign++: a Combined Strategy for Improving Ontologies Alignment

**Miguel Gabriel Prazeres Carvalho**[1], **Maria Luiza Machado Campos**[1],
**Linair Maria Campos**[2], **Maria Cláudia Cavalcanti**[3]

Programa de Pós Graduação em Informática– UFRJ[1], Rio de Janeiro,RJ,Brasil
Programa de Pós Graduação em Ciência da Informação – UFF[2], Niterói, RJ, Brasil
Programa de Pós Graduação em Sistemas e Computação-IME[3], Rio de Janeiro,RJ,Brasil

{miguelgabriel[1],mluiza[1]}@ufrj.br, linair@cisi.coppe.ufrj.br[2], yoko@ime.eb.br[3]

***Abstract.*** *Ontology reuse is very important nowadays but, more specifically, ontology alignment still represents a challenge, despite the proposal of a great number of techniques and tools that implement it. This paper presents an approach that builds upon two already existent techniques. It considers both the enrichment of the ontologies with implicit terms and relationships contained on the ontologies terms definitions and on associating concepts of the ontologies to categories of foundational ontologies. Besides confirming the improvement on alignment results when using each of these approaches, our experiments showed even better results when these techniques were applied together.*

## 1. Introduction

In recent years, the use of ontologies has greatly increased in different areas, from serving as a basis for conceptual modeling, formally defining an abstraction of a given perspective of reality, to supporting resource interoperability and knowledge discovery from multiple sources.

However, due to an increasing demand, many ontologies are built in an ad hoc manner, lacking a systematic approach for their development. This contributes to several problems when using those ontologies, mainly compatibility and interoperability between them (Kohler et al. 2006). Also, inconsistencies in ontologies structure can lead to errors in the alignment process, mistakenly associating non similar terms (Silva et al. 2011; Kohler et al. 2006; Smith, Kohler, Kumar 2004). Several studies attempt to address these problems (Ehrig 2007; Lambrix and Tan 2006; Kalfoglou and Schorlemmer, 2003). More recently, our research group has conducted two studies in this area (Silva et al. 2011; Carvalho et al. 2011) considering strategies for complementing the ontologies explicit knowledge, by applying some previous treatment on selected ontologies before the alignment process, providing in both cases a significant improvement in the results. This paper aims to merge these two approaches creating a third one that is analyzed to collect evidences that it is possible to further improve the alignment process.

The remaining of this paper is organized as follows. Section 2 presents an overview of ontologies alignment strategies. Section 3 gives a brief summary of the approaches of Silva et al. (2011) and Carvalho et al. (2011). Section 4 presents the experimental analysis conducted on the biomedical ontologies scenario and discusses the results obtained. Finally, Section 5 presents concluding remarks and future work.

## 2. Ontology Alignment Techniques

In the context of ontologies reuse, the alignment process constitutes an important instrument for the combination of the information contained in multiple but related ontologies, identifying similarities between their individual elements. It is considered the process of establishing one-to-one equality relations between the terms of two ontologies from the same domain (Ehrig 2007).

There are many available alignment tools that implement a combination of alignment techniques proposed on different approaches throughout the years. These tools consider similarity as a measure associated to elements from the ontologies being aligned, that corresponds to a numeric value indicating how similar or different the elements are. Most of the tools calculate similarity based on a combination of alignment techniques (Euzenat and Shvaiko 2007). For this paper, we have focused on techniques that complement the existing terms and structure with concepts and relations already available in definitions or other ontology elements, as well as techniques that use top-level ontologies (Guizzardi 2009) to express the ontological commitment of the ontology conceptualization.

## 3. ONTOALIGN++ and approaches from Silva et al. and Carvalho et al.

In Silva et al. (2011), before the alignment itself, a preparation step associates terms from the top three levels of the domain ontology to terms from the foundational ontology used – BFO (2012). This association helps to prevent incorrect similarity assumptions in the alignment process, restricting the indication of equivalent terms to those derived from the same meta-category, i.e. those having the same conceptual nature. As an additional customization, it also takes into account previous alignments, which serve as a reference to validate correct alignments, and also to discard incorrect ones, avoiding that these are repeatedly presented to user validation afterwards. After this, other preliminary steps are also contemplated, such as fragment extraction and cleaning. In the ontology alignment step, after source and target ontologies are prepared, the alignment is then applied, based on the NOM (Naive Ontology Mapping) approach used by the FOAM tool (Ehrig and Sure 2005), but customized with selected measures, foundational ontologies and previous alignments.

The work of Carvalho et al. (2011) explores implicit information contained in ontologies (especially those contained in the definition field) and how this information can be extracted aiming at the improvement of various processes, including the alignment. This approach uses data mining techniques in order to extract new terms and relationships in ontologies, to allow for their semantic improvement, by complementing the ontologies with these elements. It uses linguistic tools, as GATE (Bontcheva et al. 2003) and NLTK (Bird et al. 2009), and is implemented through the EI-ONTO tool, which provides support for all the steps of the approach. The approach includes two macro-steps. The first macro-step has the goal of studying the corpus and is divided into three steps: (i) transform the corpus; (ii) treat the corpus; and (iii) categorize the corpus. The second macro-step is to find corpus patterns. It uses a machine learning strategy and aims at finding patterns in the definition and in the nomenclature of terms. After these steps, the extracted terms and relationships are temporarily added to the ontology, enriching the domain knowledge already represented, and improving the alignment results, as shown in Carvalho et al. (2011).

The ONTOALIGN++ approach takes advantage of the increased expressiveness derived from both approaches described previously. First, an existing ontology can be enriched by complementing it with further terms and relationships that are “implicitly” represented in the terms definitions. Secondly, applying Silva’s approach, a complementary semantic layer can be added to this enriched ontology, extending the ontology with a more precise representation of existing concepts. Using concepts from the foundation ontology, terms from the enriched ontology have their ontological commitment made explicit.

## 4. Experimentation and Results Analysis

Aiming to explore the chosen strategies and verify that their combined use enables real gain in the alignment process, we conducted an empirical study. Moreover, we added as a secondary objective of this study the verification of the efficiency of the individual use of each of these approaches when considering an improved version of the original ontologies (they had been enhanced by OBO since the last experiments conducted by the authors). The goal of this verification is to check if the use of Silva´s et al. (2011) and Carvalho´s et al. (2011) approaches still provide an effective improvement in the alignment process, even with new improved versions of original OBO ontologies used.

Adopting an empirical approach, we have verified the efficiency of the approaches both used individually and combined. For this evaluation we have chosen two ontologies to be aligned, the Biological Process Ontology (BPO) and INOH Event Ontology. We executed four experiments and producing four corresponding results: (i) Ontologies aligned without any additional information; (ii) Ontologies aligned using Carvalho´s et al. approach (2011); (iii) Ontologies aligned using Silva´s et al. approach (2010); (iv) Ontologies aligned using Ontoalign++ approach. The first experiment was carried out without additional information. For the second experiment, we have applied Carvalho´s approach (2011) on each ontology, identifying 198 relationships in the BPO and 59 relationships in the INOH. These relationships were manually validated, resulting on 187 BPO relationships selected as valid against 54 relationships in INOH. For the third experiment we applied the approach of Silva et al. (2011) to the original ontologies, using the strategy described in that work. The foundational ontology chosen was again the BFO, for its adequacy to the biomedical area. Terms from the first three levels (as defined in the approach) of the original ontologies where linked to BFO terms, resulting on two extended ontologies. For the last experiment we have combined both the enrichment and extension to the original ontologies. We first applied the approach of Carvalho et al. (2011), in fact, using the same enriched ontologies of the second experiment. After that, we associated these enriched ontologies to the terms of BFO, as in Silva et al. (2011).

Having prepared the ontologies for each experiment, we used the FOAM tool for executing the alignments, with the following parameters: alignment–fully Automatic; number of iterations - 10; cutoff value: - 0.97; strategy – Decision Tree (Decision Tree). After the alignments, the resulting matches were independently validated by two biologists with expertise in the area of genome sequencing.

### 4.1 Results analysis

The results were tabulated and are described in Table 1, where the alignments numbers correspond to the experiments as described previously. From these results, there are evidences that the combined use of the two approaches enhances the alignment process. In addition to the increase of pairs correctly aligned, there is also a decrease in the occurrence of pairs aligned with weak relations. Another important result is the improvement on alignments 2 and 3 when compared to alignment 1, confirming that Carvalho's et al. (2011) and Silva's et al. (2011) approaches, even when used individually, are important in order to increase the number of correctly aligned pairs. In this scenario, our evidences suggest that both approaches contribute to enhance the precision of the ontologies alignment process, and, more importantly, they can be combined to reach even better results. In fact, they are complementary to each other.

**Table 1: Alignment Results**

| Classification | Results | | | |
|---|---|---|---|---|
| Degree | Alignment 1 | Alignment 2 | Alignment 3 | Alignment 4 |
| 5 – correct | 37 | 45 | 43 | 49 |
| 4 – strong relation | 5 | 7 | 8 | 8 |
| 3 – medium relation | 5 | 5 | 5 | 5 |
| 2 – weak relation | 7 | 1 | 3 | 1 |
| 1 – incorrect | 1 | 1 | 1 | 1 |
| Total | 55 | 59 | 60 | 64 |

Observing the alignments, we noticed that some of the errors derived from imprecisions on the original ontologies, as some *is_a* relationships were mistakenly represented as *part-of* relationships, and vice-versa. Also, there are gaps in the specialization hierarchies, which induce errors in the alignment process. In this last case, we have evidences that some of these gaps could be removed by refining our enrichment strategy so that more intermediary *is_a* relationships could be extracted from the definitions.

## 5. Conclusion

Even with some quality improvement incorporated more recently on existing ontologies, their reuse still present considerable challenges. Most often, when trying to reconcile overlapping domain ontologies it is not trivial to solve ambiguities and to identify similarities as main commitments that underline an ontology conceptualization which has not been properly externalized. Besides reevaluating two successful approaches used to improve the alignment of ontologies, this work also aimed at showing evidences of the advantage of combining them. The executed experiments yielded not only an increase in the number of pairs aligned as well as a decrease in the number of false alignments.  As future work, other possibilities could be explored, such as exploring associating terms from other levels of the domain ontology as well as exploring other extraction strategies and trying semi-automatic mechanisms for associating to the top-level ontology.

## References

BFO (2012) Basic Formal Ontology. Available at: http://www.ifomis.org/bfo.

Bird, S., Klein, E., Loper, E. (2009) Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit. Sebastopol, CA: O'Reilly.

Bontcheva, K., Kiryakov, A., Cunninghan, H., Popov, B., Dimitrov, M. (2003) Semantic web enabled, open source language technology. In EACL workshop on Language Technology and the Semantic Web: NLP and XML, Hungary.

Carvalho, M.G.P., Campos, L.M., Braganholo, V.P., Campos, M.L.M., Campos, M.L.A. (2011) Extracting New Relations to Improve Ontology Reuse. Journal of Information and Data Management, v. 2, p. 541-556.

Ehrig, M. (2007) Ontology Alignment: Bridging the Semantic Gap (Semantic Web and Beyond), Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Ehrig, M. and Sure, Y. (2005) FOAM - Framework for Ontology Alignment and Mapping Results of the Ontology Alignment Evaluation Initiative. In: Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies, Canada.

Euzenat, J. and Shvaiko, P. (2007) Ontology matching, Springer Verlag, Berlin, Germany.

Guizzardi, G. (2009) Ontology-Driven Conceptual Modeling - II Seminario de Pesquisa em Ontologia no Brasil. Available at ontobra.comp.ime.eb.br/apresentacoes/curso2.

Kalfoglou, Y. and Schorlemmer, M. (2003) Ontology mapping: the state of the art. Knowledge Engineering Review, v.18, n.1, p.1-31.

Köhler, J., Munn, K., Ruegg, A., Skusa, A., Smith, B. (2006) Quality Control for Terms and Definitions in Ontologies and Taxonomies. BMC Bioinformatics, v.7, n.212, p.1-12.

Lambrix, P. and TAN, H. (2006) SAMBO - A System for Aligning and Merging Biomedical Ontologies .Web Semantics: Science, Services and Agents on the World Wide Web, v.4, n.3, p.196-206.

Silva, V.S., Campos, M.L.M., Silva, J.C.P., Cavalcanti, M.C. (2011) An Approach for the Alignment of Biomedical Ontologies based on Foundational Ontologies. Journal of Information and Data Management, v. 2, p. 557-572.

Smith, B.; Köhler, J.; Kumar, A. (2004) On the Application of Formal Principles to Life Science Data: A Case Study in the Gene Ontology. In  Database Integration in the Life Sciences, p.1-17.