

# SERIMI – Resource Description Similarity, RDF Instance Matching and Interlinking

Samur Araujo<sup>1</sup>, Jan Hidders<sup>1</sup>, Daniel Schwabe<sup>2</sup>, Arjen P. de Vries<sup>1</sup>

<sup>1</sup> Delft University of Technology, PO Box 5031, 2600 GA Delft, the Netherlands  
{S.F.CardosodeAraujo, A.J.H.Hidders, A.P.deVries}@tudelft.nl

<sup>2</sup> Informatics Department, PUC-Rio Rua Marques de Sao Vicente, 225, Rio de Janeiro, Brazil  
dschwabe@inf.puc-rio.br

**Abstract.** This paper presents SERIMI, an automatic approach for solving the interlinking problem over RDF data. SERIMI matches instances between a source and a target datasets, without prior knowledge of the data, domain or schema of these datasets. Experiments conducted with benchmark collections demonstrate that our approach considerably outperforms state-of-the-art automatic approaches for solving the interlinking problem over RDF data.

## 1 Introduction

The interlinking of datasets published in the Linked Data Cloud (LDC) [1] is a challenging problem and a key factor for the success of the Semantic Web. Given the heterogeneity of the LDC, techniques aimed at supporting interlinking at instance level should ideally operate agnostic of a specific domain or schema.

SERIMI<sup>1</sup> focus in the instance-matching problem over RDF data, which refers to the process of determining whether two RDF resources refer to the same real-world entity in a given domain. We propose an unsupervised solution for this problem, which is composed of two phases: the selection phase and the disambiguation phase. SERIMI uses existing traditional information retrieval and string matching algorithms for solving the selection phase, and we propose an innovative function of similarity during the disambiguation phase. This function is designed to operate even when there is no direct ontology alignment between the source and target datasets being interlinked. Fig. 1 shows an overview of SERIMI's instance matching process.

## 2 Validation

We use the collection proposed in the DI track of the Ontology Alignment Evaluation Initiative (OAEI 2010) [2] for evaluating SERIMI. We focused our evaluation in the life science (LS) collection and in the Person-Restaurant (PR) collection proposed by this initiative. We used two baselines in our experiments: RiMOM [3] and

---

<sup>1</sup> <https://github.com/samuraraujo/SERIMI-RDF-Interlinking>

ObjectCoref [4]. These two systems are representative of the two main types of solution for the interlinking task and, more importantly, they have used the same set of datasets and reference alignment as our method, allowing a fair and direct comparison. On average, SERIMI outperforms RiMOM and ObjectCoref in 70% of the cases.

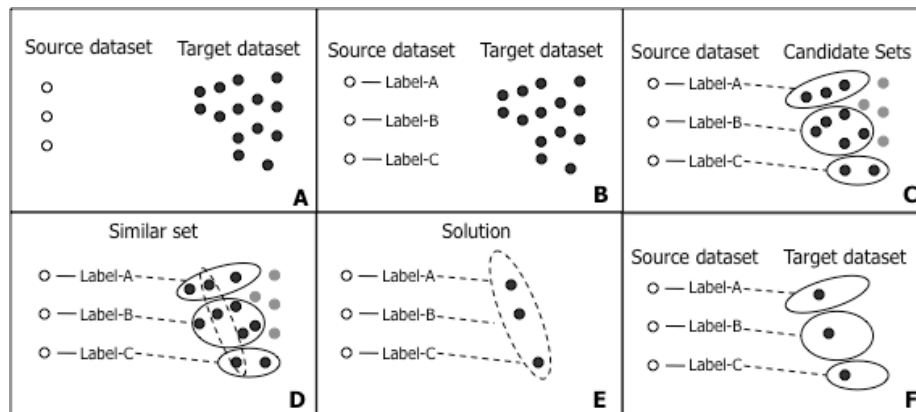


Fig. 1 – Overview of SERIMI interlinking process. (A) Given a source and target dataset and a set of source resources (instance of a class), (B) SERIMI obtains the label of these source resources and search for candidate resources in the target dataset that share a similar label. (C) For each source resource, SERIMI retrieves a set of candidate resources. (D) In order to disambiguate a set of candidate, SERIMI applies a novel function of similarity that selects the resources that are the most similar between all candidate sets (E). These selected resources are the solutions for the interlinking (F). The determination of this optimal cross section is a non-trivial process and it is done in a sophisticated way. This process assumes that the source resources belong to a homogeneous class of interest (e.g. musician, drugs, country, etc.)

### 3 Conclusion

SERIMI showed promising results for solving the task of RDF interlinking proposed by the OAEI 2010. As future work, we intend to evaluate this approach in different collections aiming to evaluate SERIMI as a domain specific solution and independent of domain solution for RDF interlinking.

### References

1. Bizer, C., Heath, T. and Berners-Lee, T. (2009) Linked Data - The Story So Far. *Int. J. Semantic Web Inf. Syst.*, 5 (3). pp. 1-22.
2. Ontology Alignment Evaluation Initiative 2010 Campaign - <http://oaei.ontologymatching.org/2010/>
3. Wang Z., Zhang X, Hou L., Zhao Y., Li J., Qi Y., Tang J. (2010). RiMOM Results for OAEI 2010. The Fifth International Workshop on Ontology Matching. Shanghai, China.
4. Hu W., Chen J., Cheng G., and Qu Y. (2010). ObjectCoref & Falcon-AO: Results for OAEI 2010. The Fifth International Workshop on Ontology Matching. Shanghai, China.