InsMT+ Results for OAEI 2015 Instance Matching

Abderrahmane Khiat[1] , Moussa Benaissa[1]

LITIO Laboratory, University of Oran1 Ahmed Ben Bella, Oran, Algeria
abderrahmane_khiat@yahoo.com , moussabenaissa@yahoo.fr

Abstract. The InsMT+ is an improved version of InsMT system participated at OAEI 2014. The InsMT+ an automatic instance matching system which consists in identifying the instances that describe the same real-world objects. The InsMT+ applies different string-based matchers with a local filter. This is the second participation of our system and we have improved somehow the results obtained by the previous version.

Keywords: Terminological Techniques, String Based Similarity, Instance Mapping, Instance Matching, Linked Data, Web of Data, Semantic Interoperability, Semantic Web.

1 Presentation of the System

1.1 State, Purpose, General Statement

The *objective* of *Linked Data* with the emergence of *the Web of Data* is to *interlink semantically data together* in order to be *reused and processed automatically* by the *software agents*. These *data* described by *instances* are *heterogeneous* and *distributed*. The *Instance matching* is a very necessary task in Linked Data; it aims to identify the *instances* that *describe* the *same real-world objects*.

The *enormous volume* of data already available on the web and its continuity to increase, requires techniques and tools capable to identify the instances that describe the same real-world objects automatically.

In this paper, we describe InsMT+ an improved version of our InsMT system which participated in OAEI 2014. This second version consists to apply *different string-based matchers* with *a local filter*. The second version shows good results better than the previous one but still not very satisfiable. The details of each step of our system are described in the following section.

1.2 Specific Techniques Used

The process of our system consists in the following successive steps.

Step 1: Extraction and Normalization of Instances In this step, our system extracts the instances. Then, we have applied (1) case conversion (conversion of all words in same upper or lower case) and (2) stop word elimination to normalize the instance informations.

Step 2: Terminological Matchers In this step, our system calculates the similarities between instances, normalized in previous phase, using various string-based matching algorithms. More precisely the different string-based matching algorithms used are: levenshtein-distance, Jaro, SLIM-Winkler. The calculations of similarities by each string matching algorithm are represented in matrix.

Step 3: Local Filter In this step, our system applies a local filter on each matrix i.e. we choose for each string-based matching algorithm a threshold to realize a filter. We consider that: the similarities which are less than the threshold are set to 0. Our intuition behind this local filter is that the similarities which are less than the threshold can influence the strategy of the average aggregation.

Step 4: Aggregation of Similarities In this step, our system combines the similarities of each matrix (after we have applied a local filter) using the average aggregation method and the result of the aggregation is represented in a matrix.

Step 5: Global Filter and Identification of Alignment In this step, our system applies a second filter on the combined matrix (result of the previous step) in order to select the correspondences found using the maximum strategy with a threshold.

1.3 Adaptations Made for the Evaluation

We do not have made any specific adaptation for this first version of InsMT+, for OAEI 2015 evaluation campaign. All parameters are the same for instance matching track of OAEI 2015.

1.4 Link to the set of provided alignments (in align format)

The result of InsMT+ system can be downloaded from OAEI 2015 website `http://islab.di.unimi.it/im_oaei_2015/index.html`

2 Results

In this section, we present the results obtained by running InsMT+ on instance matching track of OAEI 2015 evaluation campaign.

2.1 Author Disambiguation Task

The goal of the author-dis task is to link OWL instances referring to the same person (i.e., author) based on their publications.

We present below the results obtained by running InsMT+ system on author disambiguation task (see Tab. 1).

Table 1: The results of InsMT+ on the Author Disambiguation Task of OAEI 2015.

Track	System	Expected mappings	Retrieved mappings	Precision	Recall	F-measure
Sandbox task	EXONA	854	854	0.941	0.941	0.941
Mainbox task	EXONA	8428	144827	0.0	0.0	NaN
Sandbox task	InsMT+	854	722	0.834	0.705	0.764
Mainbox task	InsMT+	8428	7372	0.76	0.665	0.709
Sandbox task	Lily	854	854	0.981	0.981	0.981
Mainbox task	Lily	8428	8428	0.964	0.964	0.964
Sandbox task	LogMap	854	779	0.994	0.906	0.948
Mainbox task	LogMap	8428	7030	0.996	0.831	0.906
Sandbox task	RiMOM	854	854	0.929	0.929	0.929
Mainbox task	RiMOM	8428	8428	0.911	0.911	0.911

* The results of InsMT+ are better compared to the first version participated in OAEI 2014, we can say that we have improved the results in terms of precision. However, the results are less better than other systems due to the simple techniques used in InsMT+. Since, InsMT+ is based only on String-based similarity.

2.2 Author Recognition Task

The goal of the author-rec task is to associate a person (i.e., author) with the corresponding publication report containing aggregated information about the publication activity of the person, such as number of publications, h-index, years of activity, number of citations.

We present below the results obtained by running InsMT+ system on author recognition task (see Tab. 2).

Table 2: The results of InsMT+ on the Author Recognition Task of OAEI 2015.

Track	System	Expected mappings	Retrieved mappings	Precision	Recall	F-measure
Sandbox task	EXONA	854	854	0.518	0.518	0.518
Mainbox task	EXONA	8428	8428	0.409	0.409	0.409
Sandbox task	InsMT+	854	90	0.556	0.059	0.106
Mainbox task	InsMT+	8428	961	0.246	0.028	0.05
Sandbox task	Lily	854	854	1.0	1.0	1.0
Mainbox task	Lily	8428	8424	0.999	0.998	0.999
Sandbox task	LogMap	854	854	1.0	1.0	1.0
Mainbox task	LogMap	8428	8436	0.999	1.0	0.999
Sandbox task	RiMOM	854	854	1.0	1.0	1.0
Mainbox task	RiMOM	8428	8428	0.999	0.999	0.999

* The results of InsMT+ on this track are not at all very satisfiable. However, we can remark that the number of retrieved mappings by our system is less 10 time than the mappings discovered by other systems, which explained the results obtained. We are trying to analyses the reason of these results in order to improve our system.

3 Conclusion

This is the second time that InsMT+ system has participated in SEAL platform and OAEI campaign. In this year, our system has participated only in two instance matching tracks of OAEI 2015 evaluation campaign. The InsMT+ system gives good results better than the InsMT system but these results still not statifaisable. As future Perspective, we attempt to improve more our system in order to get better results.

References

1. A. Doan, J. Madhavan, P. Domingos, and A. Halevy, Learning to map ontologies on the semantic web, in Proceedings of the International World Wide Web Conference (2003).
2. A. Maedche and V. Zacharias, Clustering ontologybased metadata in the semantic web, in Proceedings of the 13th ECML and 6th PKDD, (2002).
3. A. Khiat, M. Benaissa, InsMT / InsMTL results for OAEI 2014 instance matching. In Proceedings of the 9th International Workshop on Ontology Matching co-located with the 13th International Semantic Web Conference (ISWC 2014), October 20, pp. 120-125. CEURWS.org, Trentino, Italy, 2014.
4. A. Maedche, B. Motik, N. Silva and R. Volz ”Mafraa mappingframework for distributed ontologies”, Springer, Benjamins VR (eds) EKAW, Berlin, vol 2473, pp 235250, (2002).
5. K. Todorov, P. Geibel, KU. Kuhnberger ”Mining concept similarities for heterogeneous ontologies”, Springer, Berlin, ICDM, vol 6171. , pp 86100, (2010).
6. J. Euzenat and P. Valtchev, Similarity-based ontology alignment in owllite, in Proceedings of ECAI, (2004).
7. J. Euzenat and P. Shvaiko. OntologyMatching. Springer (2007).
8. M. Ehrig. Ontology Alignment Bridging the Semantic Gap. Springer (2007).
9. M. Jaro. Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida. Journal of America Statistical Association, 84(406):414-420, (1989).
10. A. Khiat et M. Benaissa: ”Nouvelle Approche dAlignement dOntologies base dInstances : trasferet des instances par linfrence”, In The Proceeding of International Conference On Artificial Intelligence and Information Technology, ICA2IT 2014, Ouargla, Algeria, (2014).
11. V. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady, 10:707-710, (1966).
12. W. Winkler. The state of record linkage and current research problems. Statistics of Income Division, Internal Revenue Service. Publication R99/04 (1999).
13. M. Ehrig and Y. Sure, Ontology mapping - an integrated approach, in Proceedings of the European Semantic Web Symposium ESWS, (2004).
14. B. Schopman, S. Wang, A. Isaac and S. Schlobach, Instance-Based Ontology Alignment by Instance Enrichment, Journal on Data Semantics, vol. 1, N 4, (2012).
15. E. Rahm Towards large-scale schema and ontology Alignment, ReCALL, (2011).
16. J. Li, J. Tang, Y. Li and Q. Luo, Rimom: a dynamic multistrategy ontology alignment framework, IEEE Trans Knowl, (2009).