Alignment Evaluation of MaasMatch for the OAEI 2014 Campaign

Frederik C. Schadd, Nico Roos

Maastricht University, The Netherlands
{frederik.schadd, roos}@maastrichtuniversity.nl

Abstract. This paper summarizes the results of the fourth participation of the MaasMatch system in the Ontology Alignment Evaluation Initiative (OAEI) competition. We describe the performed changes to the MaasMatch system and evaluate the effect of these changes on the different datasets.

1 Presentation of the system

MaasMatch is a ontology mapping system with the initial focus of fully utilizing the information located in the concept names, labels and descriptions in order to produce a mapping between two ontologies [2,4]. This was achieved through the utilization of syntactic similarities and virtual documents, which can also be used as a disambiguation method for the improvement of lexical similarities [3,6].

1.1 Specific techniques used

The 2014 version of *MaasMatch* exhibits some notable changes compared to the 2013 version [5]. First, the system is now based on a de-centralized *configuration* system. For each presented mapping problem, the system queries its stored similarity measures whether the current problem is appropriate for that particular measure. Each measure independently evaluates whether the given ontologies contain a sufficient amount of exploitable input data and whether the ontologies have an appropriate size. The measures then report their results back to the system. As an example, the instance similarity would not consider itself appropriate if one of the given ontologies does not contain any instances. Additionally, each similarity also evaluates the size of the input ontologies, such that computationally expensive similarities are not executed on large-scale problems.

Using all similarities that have responded positively for the current problem, the system computes the similarity cube between the two ontologies. Here, all similarity measures are executed in *parallel* using a dynamic number of threads depending on the current hardware, such that the system can scale with the number of available computing cores. This facilitates the computation of alignments between large-scale ontologies through a more effective usage of all available computing power.

The resulting similarity cube is aggregated using the Dempster-Shafer theory, after which the result alignment is extracted. The entire mapping process of the *MaasMatch* system is visualized in Figure 1.

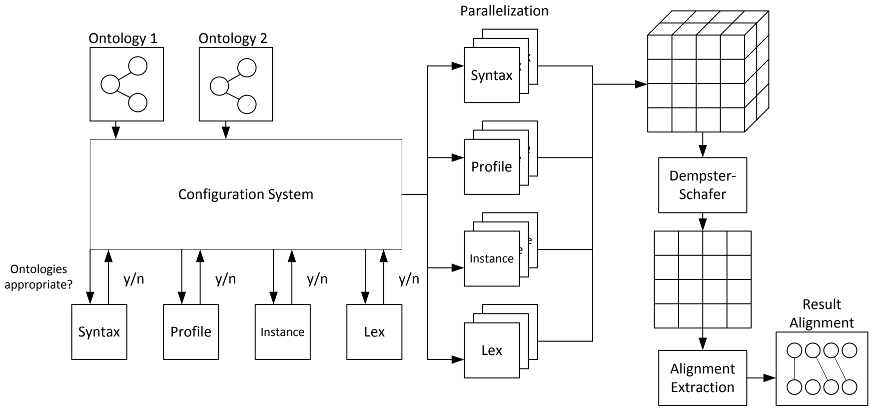

Fig. 1. Visualization of the MaasMatch architecture.

1.2 Adaptations made for the evaluation

While the system can provide correspondences with a wide range of confidence values, we have applied a hard threshold to the result extraction such that the evaluation on track which do not perform thresholding analysis better reflect the actual quality of the alignments. However, the applied threshold can easily be adjusted in the supplied configuration file.

1.3 Link to the system and parameters file

MaasMatch and its corresponding parameter file is available on the SEALS platform and can be downloaded at http://www.seals-project.eu/tool-services/browse-tools.

2 Results

This section presents the evaluation of the OAEI2014 results achieved by MaasMatch. When applicable, the performance of this year will be compared to the performance of the previous year [1]

2.1 Benchmark

The benchmark track consists of synthetic datasets, where an ontology is procedurally altered in various ways and to different extents, in order to see under what circumstances a system can still produce good results. Table 1 displays the results on the two evaluated datasets:

Compared to the results of the previous year [5], the performance of *MaasMatch* saw a shift towards the precision of the alignments. While the precisions of the previous year were in the range of 0.6, this year the precisions of the different benchmark

Test Set	Precision	F-Measure	Recall
biblio	0.97	0.56	0.37
cose	0.98	0.48	0.31
dog	0.92	0.55	0.39

Table 1. Harmonic means of the benchmark test sets.

ontologies ranged from 0.92 to 0.98. This came however at a slight cost of recall. The likely reason behind this is the re-introduction of a hard-threshold which is applied after the alignment extraction step.

Another interesting point of note is that, compared to the previous year [1], the testing procedure no longer caused issues in the execution of the system.

2.2 Anatomy

The anatomy dataset consists of a single matching task, which aligns a biomedical ontology describing the anatomy of a human to an ontology describing the anatomy of a mouse. Unique aspects about this ontology are their large sizes and the fact that they contains specialized vocabulary which is not often found in non-domain specific thesauri. Table 2 displays the results of this dataset.

Year	Precision	F-Measure	Recall	Runtime(s)
2013	0.359	0.409	0.476	8532
2014	0.914	0.803	0.716	49

Table 2. Results of the anatomy data set.

On the anatomy test track we can observe some significant improvements compared to last year's evaluation. First, we can see a significant improvement with regard to the alignment quality. Both the precision and recall have improved drastically compared to the previous year, with an absolute increase of 0.555 and 0.24 respectively. Additionally, the runtime for this dataset has been reduced drastically. Both the configuration system, which would not execute complex similarities (e.g. the lexical similarity), and the parallelized execution of all similarities contribute to this increase.

2.3 Conference

The confidence data set consists of numerous real-world ontologies describing the domain of organizing scientific conferences. The results of this track can be seen in Table 3.

Year	Precision	F-Measure	Recall
2013(ra1)	0.28	0.37	0.55
2014(ra1)	0.64	0.55	0.48
2013(ra2)	0.27	0.36	0.53
2014(ra2)	0.52	0.50	0.49

Table 3. Results of the conference data set.

Overall we can observe an improved performance on the conference dataset for both the *ra1* and *ra2* reference alignments. The likely reason behind this is the improved selection and aggregation of the similarity measures. The runtime of the entire evaluation was 68777 seconds. This is significantly higher than the runtime of the anatomy track, since the conference track consists of numerous small mapping tasks. The system analyses each task individually with regard to its complexity. For the anatomy track, the single task is evaluated as too large for time consuming similarity measures, such that these are dropped. However, any given mapping problem of the conference track is small enough such that the application of time consuming similarities is still feasible, resulting in the overall runtime being higher for this track than for the larger anatomy track.

2.4 Multifarm

The Multifarm data set is based on ontologies from the OntoFarm data set, that have been translated into a set of different languages in order to test the multi lingual capabilities of a specific system. The results of MaasMatch on this track can bee seen in Table 4.

Year	Precision	F-Measure	Recall
2013 (same ontology)	0.62	0.29	0.19
2014 (same ontology)	0.52	0.10	0.06
2013 (different ontology)	0.01	0.02	0.03
2014 (different ontology)	0.27	0.15	0.10

Table 4. Results of the multi-farm data set.

Despite the system not being designed for multi-lingual mapping, we saw an improvement in performance for the mapping tasks with different ontologies. For this part of the dataset, the precision was increased significantly while the recall saw a moderate increase.

For mapping tasks consisting of the same ontology being translated into different languages the overall performance was lower than the previous year. A likely reason for this is that the internal structures of the concepts are no longer taken into consideration compared to last year, such that a decreased performance for mapping problems with identical structures are to be expected.

2.5 Large BioMed

The Large Biomedical track consists of three mapping problems in which very large ontologies modelling the biomedical domain have to be mapped. The results of this track can be seen in table 5.

Year	Precision	F-Measure	Recall	Runtime(s)
2013 (FMA-NCI Task 1)	0.407	0.456	0.517	12,409
2014 (FMA-NCI Task 1)	0.808	0.824	0.840	1,460
2013 (FMA-SNOMED Task 1)	-	-	-	-
2014 (FMA-SNOMED Task 1)	0.655	0.664	0.674	4,605

Table 5. Results of the multi-farm data set.

We can observe some significant improvements compared to the results of the previous year. In the previous year, *MaasMatch* was unable to produce a result alignment within the set time limit for the FMA-SNOMED matching task. This year, the system did produce an alignment within the time limit with a F-measure of 0.664. The results for the FMA-NCI track have improve significantly. Both the precision and recall have improved over the previous year, resulting in an increase of F-Measure from 0.456 to 0.824. In addition, the required runtime for this task has been reduced by approximately 89%.

Some issues however remain for this dataset. Further improvements need to be made such that the system can tackle the largest task (NCI-SNOMED). This year, the system was unable to complete this task due to memory issues. The likely cause behind this is the current implementation of the profile similarity. To improve runtime, this similarity caches all concept profiles in memory such that these do not have to be re-created whenever a similarity computation is invoked on the same concept. However, due to the large size of the matching tasks this optimization is no longer a feasible solution due to memory constraints.

3 General comments

3.1 Comments on the results

Overall, we have seen improved results for all evaluation tracks, leading to competitive performances when compared to the other mapping systems. Furthermore, large-scale mapping problem were now solvable within a reasonable runtime for the first time. Some weaknesses still remain, for instance the result alignments being non-coherent, but ultimately the current iteration of the system has been largely successful.

3.2 Discussions on the way to improve the proposed system

The *MaasMatch* system saw some significant changes compared to last year's iteration, which is reflected in the different results for all tracks. Most observed changes were indeed positive. Some areas however remain were more improvements can be made. This year saw the introduction of a decentralized self-configuration system, where the logic of determining whether a similarity measure is appropriate is de-coupled to each particular metric. There current implementation however is only preliminary. We foresee an improved system which contains a set of testing problems, similar to the different tracks of OAEI, on which every similarity metric can be automatically evaluated with regard to its compatibility and run-time efficiency. These results could then be stored and consulted for any new mapping task.

Currently, multi-lingual problems are not supported. While we did investigate the possibility of multi-lingual adaptations, none of the available options were satisfactory. On-line solutions, e.g. *Google Translate* have the issue that these are typically commercial, such that there are no free options for research available, and limited with regard to the amount of queries on can issue per month, making the adoption for large-scale problems infeasible. *Off-line* options, such as *BabelNet* have the issue that these are much larger than the available storage per system on the SEALS platform (5.1GB as opposed to the 500MB limit). A solution would be to establish a private server on which *BabelNet* can be queried by the system, though this was not pursued due to time constraints.

4 Conclusion

In this paper we presented the results of the *MaasMatch* system for the 2014 OAEI campaign. The system has changed significantly compared to the previous year, which is reflected in the performance of the different tracks. Overall, most tracks have seen improvements with regard to alignment quality. The self-configuration system now made the mapping of large problem in a feasible time a possibility, as evidence in the runtime performance during the anatomy track.

References

1. Bernardo Cuenca Grau, Zlatan Dragisic, Kai Eckert, Jérôme Euzenat, Alfio Ferrara, Roger Granada, Valentina Ivanova, Ernesto Jiménez-Ruiz, Andreas Oskar Kempf, Patrick Lambrix, et al. Results of the ontology alignment evaluation initiative 2013. In *Proc. 8th ISWC workshop on ontology matching (OM)*, pages 61–100, 2013.
2. Frederik C. Schadd and Nico Roos. Maasmatch results for oaei 2011. In *Proceedings of The Sixth International Workshop on Ontology Matching (OM-2011) collocated with the 10th International Semantic Web Conference (ISWC-2011)*, pages 171–178, 2011.
3. Frederik C. Schadd and Nico Roos. Coupling of wordnet entries for ontology mapping using virtual documents. In *Proceedings of The Seventh International Workshop on Ontology Matching (OM-2012) collocated with the 11th International Semantic Web Conference (ISWC-2012)*, pages 25–36, 2012.

4. Frederik C. Schadd and Nico Roos. Maasmatch results for oaei 2012. In *Proceedings of The Seventh ISWC International Workshop on Ontology Matching*, pages 160–167, 2012.
5. Frederik C. Schadd and Nico Roos. Summary of the maasmatch participation in the oaei-2013 campaign. In *Proceedings of The Eighth International Workshop on Ontology Matching (OM-2013) collocated with the 12th International Semantic Web Conference (ISWC-2013)*, pages 139–145, 2013.
6. Frederik C Schadd and Nico Roos. Word-sense disambiguation for ontology mapping: Concept disambiguation using virtual documents and information retrieval techniques. *Journal on Data Semantics*, pages 1–20, 2014.