Summary of the MaasMatch participation in the OAEI-2013 campaign

Frederik C. Schadd, Nico Roos

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

Abstract. This paper summarizes the results of the third participation of the MaasMatch system in the Ontology Alignment Evaluation Initiative (OAEI) competition. Several additions were made to the MaasMatch system with the intent of rectifying its limitations, as observed during the previous OAEI campaign. The extent of the additions and their effect on the individual dataset will be elaborated.

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. 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 [3,4]. The results of the benchmark track in the OAEI 2012 competition [1] substantiated the evident conclusion that when the naming and annotation features of an ontology are not present or distorted, the system produces unsatisfactory mappings. Several additions have been made to rectify this issue, the details of which are presented in the next subsection.

1.1 Specific techniques used

The current version of MaasMatch utilizes a wider spectrum of similarity techniques than past versions. The overall setup in which these are used can be seen in Figure 1.

When given two input ontologies, these are parsed into an OWL format to allow further processing. For each configured similarity measure the pairwise similarities between the ontology concepts are computed, which are then combined into a similarity cube. The different similarity values are then aggregated, such that these can be used as initial vertex weights for the similarity flooding procedure [2]. The vertex weights are propagated until they converge, with a limit of 10 iteration configured to deal with situation where the values do not converge. However, in our own preliminary evaluations we found that on average only 4 iterations were needed until the values converged. Using the resulting vertex weights the results alignment is extracted.

The previous version of MaasMatch utilized four similarity measures which rely on the names, labels and comments of the concept definitions. One of which is a syntactical similarity (Jaccard), two a type of structural similarity (Name-Path, Virtual Document
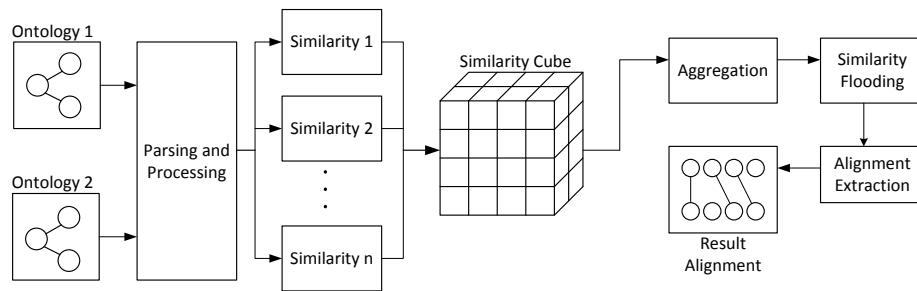
Fig. 1. Visualization of the MaasMatch architecture.

Similarity) and one a lexical similarity. The details of these can be found in the report paper of the previous year [4].

The aim of this year's development was to increase the utilization of other ontology feature, with the hope that the resulting system will be more robust to distortions and produce alignments of better quality. To achieve this, the system now also utilizes a internal structural similarity and a instance similarity, while also a similarity flooding procedure after the aggregation step in order to discover additional mappings.

When comparing classes, the internal structural similarity gathers all properties whose inferred domains correspond with the given classes. Then the a maximum correspondence between these two sets of properties are computed according to the similarities between the data-types of the properties. Comparing properties involves a combinations of two similarities. First, the data-types of the properties themselves are compared. Second, the other properties in the immediate neighbourhood are compared using the maximum correspondence of the two property sets.

The instance similarity compares the asserted instances of concepts using information retrieval techniques. For classes, all instances that are asserted to belong to their corresponding class are gathered, where all values that are asserted in each of these instances are collected in a document. For properties, all values that are asserted using these properties are gathered in a document instead. The similarity between classes and properties is then determined by the similarity of their instance documents.

It is important to note that with the attempt of making the system more robust by adding more similarities and procedures, the runtime of the system will be negatively impacted, especially since the full similarity cube is computed. Also, the similarity flooding procedure entails the process of computing a pairwise connectivity graph of the two input ontologies. This means that, given two large input ontologies, the resulting graph will have many nodes and vertices, which will have to be stored in memory. Hence, the memory requirements for large matching tasks will be quite high. Given both of these issues, future endeavours will likely entail some methodologies to reduce the memory requirements and the amount of comparisons between concepts.

1.2 Adaptations made for the evaluation

For this year's evaluation we have re-introduced a alignment cut-off based on preliminary evaluations, since not all tracks perform a thresholding procedure during the evaluation, yielding results that do not reflect the alignment quality. For the similarity flooding procedure the vertex weights are updated using the increment method C [2]. We also added secondary matcher, based on our anchor-profile approach [5], to the bridge for the evaluation using partial input alignments. Unfortunately, this year's competition did not run this specific sub-track, meaning that we were not able to observe its performance in the field. The functionality however is still available.

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 OAEI2013 results achieved by MaasMatch. Evaluations utilizing ontologies exceeding the supported complexity range, such as the Library track, will be excluded from the discussion for the sake of brevity.

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:

Test Set	Precision	F-Measure	Recall
biblio2	0.6	0.6	0.6
biblioc	0.84	0.69	0.59

Table 1. Harmonic means of the benchmark test sets.

Overall, we can see an improvement over last year's performance [4]. While in the previous year the highest achieved f-measure was at 0.6 among the different sets, this year this is actually the lowest achieved f-measure, with the system scoring significantly higher on the *biblioc* set.

Unfortunately, according to the experimenter the system did not produce any output for the tasks 254 and higher. Upon hearing about this issue, we evaluated the tool locally using the SEALS client to replicate the issue, using both the client from last year and the current 'v4i' version. With both evaluation clients, MaasMatch ran normally and

produced output for all tasks of the test sets. Furthermore, we also observed that other systems, namely LogMap, ServOMap and MapSSS, also had these issues, even though these and also MaasMatch performed without error in last year's competition. From this we must conclude that this error stems from the SEALS platform, and given a proper evaluation the MaasMatch system could have performed much higher.

In addition to this evaluation, another benchmark run was performed using the *onlira* ontology, with the intention of performing an evaluation for which the participants do have access to the dataset in advance. While the results of this evaluation will likely not be published, due to many participating systems not being able to cope with the matching task, it is interesting to see how well MaasMatch performed with this base ontology:

Test Set	Precision	F-Measure	Recall
onlira	0.94	0.74	0.61

Table 2. Harmonic means of the benchmark test set using the onlira base ontology.

From Table 2 we can see that the performance of MaasMatch is consistent with the performance of the standard benchmark set, with a higher emphasis on precision than recall.

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 3 displays the results of this dataset.

Test Set	Precision	F-Measure	Recall
mouse-human	0.359	0.409	0.476

Table 3. Results of the anatomy data set.

This year we can observe a drop in performance, specifically with regard to the recall of the alignment. The most likely reason behind this is that this dataset does not contain the features that the newly added similarities use, namely instances and properties, such that the distinction between the positive and negative correspondences becomes smaller. The overall similarity values will be lower, since two similarities will not produce any positive values, such that it is more likely that correct correspondences will be dismissed due to their similarity value being lower than the re-introduced threshold.

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 4.

Test Set	Precision	F-Measure	Recall
ra1	0.29	0.38	0.54
ra2	0.29	0.37	0.53

Table 4. Results of the conference data set.

Similarly tot he anatomy dataset, we observe that the additions to the system had a detrimental effect to the alignment quality, in this case with more pronounced effects on the precision. Similarly to the anatomy track, this dataset also does not contain instances, yielding the instance similarity redundant. However, properties are present, yielding the interesting observation that while the internal structural similarity showed itself to be of positive influence on the benchmark dataset, its basic intuition which it exploits is not applicable to the conference dataset.

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 5.

	cn-cz	cn-de	cn-en	cn-es	cn-fr	cn-nl	cn-pt	cn-ru	cz-de	cz-en	cz-es	cz-fr	cz-nl	cz-pt	cz-ru	de-en	de-es	de-fr
P	.13	.11	.12	.13	.12	.11	.14	.13	.15	.16	.15	.15	.16	.16	.10	.20	.15	.16
F	.12	.11	.12	.13	.12	.11	.13	.12	.14	.16	.15	.14	.15	.15	.10	.19	.14	.15
R	.12	.10	.11	.12	.11	.10	.12	.12	.14	.15	.14	.13	.14	.14	.10	.18	.14	.14
	de-nl	de-pt	de-ru	en-es	en-fr	en-nl	en-pt	en-ru	es-fr	es-nl	es-pt	es-ru	fr-nl	fr-pt	fr-ru	nl-pt	nl-ru	pt-ru
P	.20	.15	.12	.18	.20	.19	.18	.14	.19	.17	.22	.12	.17	.19	.12	.15	.12	.12
F	.19	.14	.12	.17	.19	.18	.17	.14	.18	.16	.21	.12	.17	.18	.12	.14	.12	.11
R	.18	.14	.11	.16	.18	.17	.17	.13	.17	.15	.20	.11	.16	.17	.11	.14	.11	.11

Table 5. Results of the multifarm dataset.

Compared to the results of the previous year [1], we can see an overall improvement on nearly every task. While in the previous year a very large portion of the tasks resulted in an f-measure of 0.1 or below, this year we can see that in all tasks MaasMatch produced an alignment with an f-measure of .1 or greater. While we can observe

that the addition of language independent similarities did aid the performance of our system, further development is still required in order to reliably produce alignments of significant quality.

3 General comments

3.1 Comments on the results

This year we have observed mixed results for MaasMatch. While the performance of some tracks has seen improvements thanks to our modifications (benchmark, multi-farm), these came at a cost of performance in other tracks (conference, anatomy).

3.2 Discussions on the way to improve the proposed system

This year we added a wider range of similarities in order to make the system more robust. Unfortunately, this caused a detriment in performance for mapping tracks which did not contain the ontology features which the new similarities exploit. From this, we can conclude that an important improvement to our system would be the automatic detection of ontology features and automatic selection of appropriate similarities.

Furthermore, the runtime of MaasMatch is too high in order to realistically tackle huge mapping tasks. This is mostly due to the computation of the full similarity cube. To remedy this, another addition could be some kind of partitioning method, such that larger mapping tasks also become feasible.

We did see improvements in the multifarm dataset. However, this was achieved without any preprocessing step on the ontologies. An obvious improvement on this end would be the addition of a preprocessing step which automatically detects the natural language in which the ontology is written and translating it to a standard lingua-franca, for instance English.

3.3 Comments on the OAEI 2013 procedure

This year's run on the benchmark trajectory saw numerous systems, including Maas-Match, consistently having troubles producing alignments. While the participants have been notified before the results publication of this issue, they were left with only a limited amount of time to address the issue, while the organizers did not investigate the issue themselves at all. This is especially troubling since our own local evaluations using the SEALS clients did not result in these errors, giving a strong indication that the problem lies within the SEALS infrastructure, thus unfairly casting the affected systems in a negative light. We suggest to re-introduce a three week testing period to the evaluation procedure, similar to the 2011 OAEI competition. That way participants can be notified sufficiently early about potential technical issues and giving them enough time to address these.

3.4 Comments on the OAEI 2013 measures

The evaluation of ontology mapping quality is commonly done using the standard measures of precision, recall and f-measure, these methods do not take into account the confidence values associated with the individual correspondences. Recently, two techniques have seen deployment to take the confidences into account, being thresholding and confidence weighted measures. While these developments are appreciated, it is important to communicate which of these techniques have been applied in the evaluation process in order to facilitate the accurate replication of evaluation results.

4 Conclusion

This paper describes the 2013 participation of MaasMatch in the OAEI campaign. We briefly describes the overall setup of the system and the new techniques which were added to it for this evaluation. Those techniques were mainly aimed at improving the robustness of the system by utilizing a more varied range of ontological features. While this main goal has been achieved, evidenced by higher performances in the benchmark and multifarm evaluation, this surprisingly came to the detriment in performance in the remaining tracks, where the newly exploited types of features are not present in the test ontologies. We conclude that, now that MaasMatch possesses a varied spectrum of similarities, there needs to be computation step before the similarity calculation, which analyses the input ontologies with regards to its features. According to this analysis, only appropriate similarities would then be selected for the mapping procedure.

References

1. J.L. Aguirre, B.C. Grau, K. Eckert, J. Euzenat, A. Ferrara, R.W. van Hague, L. Hollink, E. Jimenez-Ruiz, C. Meilicke, A. Nikolov, D. Ritze, P. Shvaiko, O. Svab-Zamazal, C. Trojahn, and B. Zapilko. Results of the ontology alignment evaluation initiative 2012. In *Proc. of the 7th ISWC workshop on ontology matching*, pages 73–115, 2012.
2. Sergey Melnik, Hector Garcia-Molina, and Erhard Rahm. Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *Data Engineering, 2002. Proceedings. 18th International Conference on*, pages 117–128. IEEE, 2002.
3. F.C. Schadd and N. 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. F.C. Schadd and N. Roos. Maasmatch results for oaei 2012. In *Proceedings of The Seventh ISWC International Workshop on Ontology Matching*, pages 160–167, 2012.
5. F.C. Schadd and N. Roos. Anchor-profiles for ontology mapping with partial alignments. In *Proceedings of the 12th Scandinavian AI conference (SCAI)*, 2013. Accepted paper.