MaasMatch results for OAEI 2012

Frederik C. Schadd, Nico Roos

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

Abstract. This paper summarizes the results of the participation of MaasMatch in the Ontology Alignment Evaluation Initiative (OAEI) of 2012. We provide a brief description of the techniques that have been applied, with the emphasis being on the utilized similarity measures and the performed improvements over the system that participated in the year 2011. Additionally, the results of the 2012 OAEI campaign will be discussed.

1 Presentation of the system

1.1 State, purpose, general statement

Sharing and reusing knowledge is an important aspect in modern information systems. Since multiple decades, researchers have been investigating methods that facilitate knowledge sharing in the corporate domain, allowing for instance the integration of external data into a company's own knowledge system. Ontologies are at the center of this research, allowing the explicit definition of a knowledge domain. With the steady development of ontology languages, such as the current OWL language [5], knowledge domains can be modelled with an increasing amount of detail.

The initial research of the MaasMatch framework focused on resolving terminological heterogeneities between ontology concepts, which is reflected in its initial selection of similarity measures. Recent research focused on further developing these techniques, while increasing its spectrum of similarity measures such that the system can be applicable in a wider area of matching tasks. The supported matching domain of ontologies for MaasMatch are limited to semi-large, meaning up to ~2000 concepts per ontology, mono-lingual OWL ontologies, thus yielding predictable results for the Library and Multifarm tracks.

1.2 Specific techniques used

Various similarity measures covering differing categories have been applied in the current system. This subsection provides a brief explanation of each measure and how these are combined to extract the final alignment.

Syntactic Similarity MaasMatch currently utilizes a token-based measure for the purpose of determining the syntactic similarity between concepts. More specifically, concept names and labels are compared by computing the 3-grams [10] of their names and determining their similarity using the Jaccard [3] measure.

Structural Similarity As structural similarity a Name-Path similarity is used. Given a concept c, such a similarity collects the name of c and all ancestors of c, which is subsequently used as a basis for comparison. Given the nature of these strings, a hybrid similarity has been selected for this purpose. A hybrid similarity is defined as any similarity that relies on another similarity measure for its computation. Cohen et al. [1] researched a token-based framework for a hybrid distance. Given two strings s and t, the set of tokens $a_1, a_2, ... , a_K$ into which string s can be divided into and the set of tokens $b_1, b_2, ... , b_L$ into which string t can be divided into, a hybrid distance can be computed as follows:

$$sim(s,t) = \frac{1}{K} \sum_{i=1}^{K} \max_{j=1}^{L} sim'(a_i, b_j) \tag{1}$$

The hybrid similarity in MaasMatch utilizes the Levenshtein [4] similarity, to which a substring-based extension is applied. This extension functions similarly to the Winkler [11] extension, however is not limited to the size or location of the substring. This setup has been shown to outperform other variations of measures on the conference dataset and a record matching dataset [2]. Given two strings s and t, the longest common substring of s and t defined as $LCS(s,t)$ and a scaling factor S, sim' of our hybrid distance is computed as follows:

$$sim'(s,t) = Levenshtein(s,t) + \frac{LCS(s,t)}{\min(s,t)} \cdot S \cdot (1 - Levenshtein(s,t)) \tag{2}$$

Virtual Document Similarity A new similarity that is deployed in MaasMatch is the comparison of virtual documents representing ontology concepts, which are created by gathering the information contained within a concept and the information of its related neighbours according to a specific model. This approach has been pioneered by Qu et al. [7]. In essence, this approach uses a weighted combination of descriptions of concepts. A description of a concept is a weighted document vector describing the terms that occur in the concept description. The model of creating such a description allows for certain types of terms, such as the concept name, label or comments, to be weighted differently according to their perceived importance. Descriptions of related concepts are added to the description of a particular concept by multiplying the term weights of the related descriptions with a diminishing factor before merging the vectors. For a full description of this process, we recommend the reader to consult the works of Qu et al. [7].

Lexical Similarity This similarity has seen improvements, compared to its counterpart of the 2011 competition, with regard to its computing time. The similarity uses Word-Net as a basic lexical resource, however utilizes virtual document similarities between ontology concepts and WordNet synsets in order to only assign synsets to concepts which accurately describe the meaning of that concept. Given two ontologies O_1 and O_2 that are to be matched, O_1 contains the sets of entities $E_x^1 = \{e_1^1, e_2^1, ..., e_m^1\}$, where x distinguishes between the set of classes, properties or instances, O_2 contains the sets

of entities $E_x^2 = \{e_1^2, e_2^2, ..., e_n^2\}$, and $C(e)$ denotes a collection of synsets representing entity e, the essential steps of our approach, performed separately can be described as follows:

1. For every entity e in E_x^i, compute its corresponding set $C(e)$ by performing the following procedure:
    (a) Assemble the set $C(e)$ with synsets that might denote the meaning of entity e.
    (b) Create a virtual document of e, and a virtual document for every synset in $C(e)$.
    (c) Calculate the document similarities between the virtual document denoting e and the different virtual documents originating from $C(e)$.
    (d) Discard all synsets from $C(e)$ that resulted in a low similarity score with the virtual document of e, using some selection procedure.
2. Compute the WordNet similarity for all combinations of $e^1 \in E_x^1$ and $e^2 \in E_x^2$ using the processed collections $C(e^1)$ and $C(e^2)$.

Figure 1 illustrates steps 1.b - 2 of our approach for two arbitrary ontology entities e^1 and e^2:
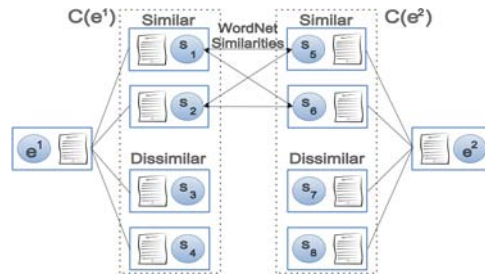

Fig. 1. Visualization of step 1.b-2 of the proposed approach for any entity e^1 from ontology O_1 and any entity e^2 from ontology 2.

Further details of the particular steps of this approach are illustrated in the works by Schadd et al. [9].

Aggregation and Extraction In our system, similarity matrices are aggregated by computing the average similarity measure of each pairwise combination of concepts, based on the computed similarity cube. The Naive descending extraction algorithm [6] is applied on the aggregated similarity matrix in order to determine the final alignment. At this point a confidence threshold can be applied in order to avoid producing alignments which do not satisfy a determined degree of confidence.

1.3   Adaptations made for the evaluation

While for practical applications it is recommended to apply a confidence boundary in the extraction step, this has been omitted for the evaluation system in order to provide

the possibility for the experimenters to conduct a more thorough analysis of the produced alignments, even if these have a low confidence value and would not be included in the final alignment under normal circumstances.

1.4 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 OAEI2012 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. Note that the evaluations of some of the tracks do not determine the optimal confidence threshold of the produced alignments such that correspondences with low confidence values are incorporated into the evaluations as well, resulting in lower performance measures compared to a normal execution environment.

2.1 Benchmark

The benchmark data set consists of several base ontologies which are matched with automatically altered versions of themselves. This makes it possible to establish under what condition a matcher performs well or poorly. Previous competitions used only a single ontology as base, with the alterations being done by hand. The current data set consists of several base ontologies such that a more varied spectrum of knowledge domains is utilized. The results of MaasMatch on the benchmark data set can be seen in Table 1.

Test Set	Precision	Recall	F-Measure
biblio	0.54	0.57	0.56
2	0.6	0.6	0.6
3	0.53	0.53	0.53
4	0.54	0.54	0.54
finance	0.59	0.6	0.59

Table 1. Aggregated harmonic means of the benchmark test sets.

From Table 1 it is observable that the results set a stark contrast in comparison to the competition of 2011 [8]. The continued development of our system was successful in increasing the recall of the produced alignments, however this came at a cost of reduced recall, yielding a similar f-measure when compared to the previous year. However, this evaluation does not take into account the confidence values provided with the

alignments, resulting in alignments with low confidence value being included in the evaluation. In a realistic scenario a pruning mechanism, for instance a simple cutoff rate, would be applied such that matches with low confidence values would not be included. As reported by the experimenter, pruning the alignments results in f-measure gains between 0.07 to 0.15, mostly due to a significant gain in precision, thus yielding significantly improved results over the MaasMatch system of 2011.

2.2 Anatomy

The anatomy data set consists of two large real-world ontologies from the biomedical domain, with one ontology describing the anatomy of a mouse and the other being the NCI Thesaurus, which describes the human anatomy. The results of this data set can be seen in Table 2.

Test Set	Precision	Recall	F-Measure
mouse-human	0.434	0.784	0.559

Table 2. Results of the anatomy data set.

Also the results of the anatomy data set have seen some drastic changes compared to the results of the previous year. The recall has been significantly improved, albeit at the cost of a significant proportion of precision. Overall, the f-measure has been improved by 0.11 over the results of the previous year [8].

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.

Test Set	Precision	Recall	F1-Measure
ra1	0.63	0.57	0.60
ra2	0.60	0.50	0.56

Table 3. Results of the conference data set.

For this data set, MaasMatch produced alignments of fairly balanced quality. The comparison to the standard reference alignments resulted in an f-measure of 0.6, which is a significant improvement compared to the same evaluation of the previous year. The evaluation using reference alignments which have been pruned using a consistency reason resulting in the recall being more affected than the precision of the alignments.

2.4 Large Biomedical Ontologies

This data set consists of several large scale ontologies, containing up to tens of thousands of concepts. While ontologies of such scale are not in the target domain of Maas-Match, due to the high computation complexity, some evaluation could still be performed, visible in Table 4.

Test Set	Precision	Recall	F-Measure
FMA-NCI Original UMLS	0.622	0.765	0.686
FMA-NCI Clean UMLS (LogMap)	0.606	0.778	0.681
FMA-NCI Clean UMLS (Alcomo)	0.597	0.788	0.679

Table 4. Results of the Large Biomedical Ontologies data set.

Among the varying evaluation methods, MaasMatch produced fairly consistent alignments when matching the FMA and NCI ontologies, all resulting in f-measures of approximately 0.68. Unfortunately, the remaining ontologies of this data set are outside of the supported complexity range, such that an alignment could not be computed within the given time frame. However, the results of the completed tasks indicate that our system is already capable of producing alignments of high quality in this domain, thus improving its efficiency, for instance by applying partitioning techniques, should result in an overall satisfying performance during the next evaluation.

2.5 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. Currently, the similarities employed by MaasMatch are not suitable in a multi-lingual matching problem, thus yielding predictably poor results.

Test Set	Precision	Recall	F-Measure
type I	0.02	0.14	0.03
type II	0.14	0.14	0.14

Table 5. Aggregated results of the Multifarm data set.

In Table 5, aggregation measures are separated into heterogeneous ontologies translated into different languages (type I) and homogeneous ontologies translated into different languages (type II). While the recall is unchanged for both matching types, the precision if positively influenced for homogeneous matching tasks.

3 General comments

3.1 Comments on the results

Overall, our system has seen improvements across various tracks, aided by the incorporation of additional similarity measures as well as the further development of the already existing measures. While the results of the previous year were high in precision and low in recall, the results of this year's participation demonstrate a more balanced measure of precision and recall, with both measures usually having a similar value.

3.2 Discussions on the way to improve the proposed system

The first area of improvement would consist of expanding the supported domain of matching problems, such that large scale or multi-lingual ontologies can be matched as well. Matching large scale ontologies would require the development of partitioning techniques in order to reduce the computational complexity of a matching task, preferably without impacting the results.

3.3 Comments on the SEALS platform

While the SEALS platform is a convenient tool for competition purposes, it would be nice to see its capabilities expanded such that evaluations can be automatically performed for research purposes, such that for instance any matching tool that is uploaded is automatically evaluated on the different available data sets.

3.4 Comments on the OAEI 2011 procedure

This years competition has seen some confusion whether or not the participants should omit post processing measures, such as cutoff based alignment pruning, given that some tracks perform automatic thresholding in order to generate the best possible alignments. However, the reported results of the benchmark data set did not include automatic thresholding, thus yielding the impression that the systems performs worse than it actually does. It would be preferable to have a clear statement on this matter and that each track is being evaluation according to the same policy.

3.5 Comments on the OAEI 2011 measures

An important part of the scientific method is the ability of recreating experimental results. Some tracks aggregate precision, recall and f-measure using the harmonic mean. However, given that the ranges of these 3 values lie in the interval of $[0, 1]$, it is possible that values of 0 would be incorporated in the evaluation, which in turn would yield a division by 0 due the reciprocal being computed of these values. It is currently unclear how this is circumvented and how exactly the measures are aggregated, making it very difficult to replicate experiments outside the OAEI environment. Thus it would be preferable to incorporate a detailed explanation on the computation and especially aggregation of the computed measures, even if this means including the same text in each year's proceedings.

4 Conclusion

This paper describes the 2012 participation of MaasMatch in the OAEI campaign, in which considerable improvements have been observed in the benchmark, anatomy and conference tracks, which have been evaluated in the previous year. New tracks were introduced with matching problems outside of the currently supported matching domain, however we intend to expand the capabilities of our system such the new types of problems can be tackled as well.

References

1. W. W. Cohen, P. Ravikumar, and S. E. Fienberg. A comparison of string distance metrics for name-matching tasks. In *Proc. IJCAI-03 Workshop on Information Integration on the Web (IIWeb-03)*, pages 73–78, 2003.
2. M. Hermans and F. C. Schadd. A generalization of the winkler extension and its application for ontology mapping. In *Proceedings Of The 24th Benelux Conference on Artificial Intelligence (BNAIC 2012)*, 2012.
3. P. Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37:547–579, 1901.
4. V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. Technical Report 8, 1966.
5. D. L. McGuinness and F. van Harmelen. OWL web ontology language overview. W3C recommendation, W3C, February 2004.
6. C. Meilicke and H. Stuckenschmidt. Analyzing mapping extraction approaches. *The Second International Workshop on Ontology Matching*, 2007.
7. Y. Qu, W. Hu, and G. Cheng. Constructing virtual documents for ontology matching. In *Proceedings of the 15th international conference on World Wide Web*, WWW '06, pages 23–31, New York, NY, USA, 2006. ACM.
8. F. C. Schadd and N. Roos. Maasmatch results for oaei 2011. In *Proc. 6th ISWC workshop on Ontology Matching (OM)*, pages 171–178, 2011.
9. F. C. Schadd and N. Roos. Coupling of wordnet entries for ontology mapping using virtual documents. In *Proceedings of the ISWC'12 International Workshop OM-2012*, 2012. Accepted Paper.
10. C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, January 2001.
11. W. E. Winkler. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. Technical report, 1990.