# DisMatch results for OAEI 2016

Maciej Rybiński *, María del Mar Roldán-García, José García-Nieto, and José F. Aldana-Montes

Dept. de Lenguajes y Ciencias de la Computación, University of Malaga, ETSI Informática, Campus de Teatinos, Malaga - 29071, Spain
maciek.rybinski@lcc.uma.es, mmar@lcc.uma.es, jnieto@lcc.uma.es, jfam@lcc.uma.es

**Abstract.** DisMatch is an experimental ontology matching system based on the use of corpus based distributional measure for approximating semantic relatedness. Through the use of a domain-related corpus, the measure can be applied to a problem focused on the domain of the corpus, here being the Disease and Phenotype track. In this paper, we aim to briefly present the proposed approach and the results obtained in the evaluation, as well as some early conclusions regarding the performance of DisMatch.

**Keywords:** Ontology Matching, Bench-marking, Lexical Semantic Relatedness

## 1 Presentation of the system

### 1.1 State, purpose, general statement

It has been demonstrated that corpus based measures can be used to successfully approximate human judgment, w.r.t. semantic relatedness between pairs of concepts [1,3,4]. DisMatch is an experimental system built for the purpose of evaluating the applicability of a state-of-the-art domain-focused corpus based measure of semantic relatedness, to a task of ontology alignment.

For a pair of ontologies, DisMatch calculates the matrix of semantic relatedness between labels representing their concepts. It then uses this matrix as the input for the classic algorithm of Similarity Flooding [2], in order to incorporate the taxonomic information into our final results.

### 1.2 Specific techniques used

The workflow of DisMatch can be broken down into the following steps:

1. Preprocessing: extraction of the taxonomies and labels of the concepts.
2. Assigning distributional representations to the concepts of the ontologies

* Corresponding author maciek.rybinski@lcc.uma.es

3. Calculating the semantic relatedness for the pairs of concepts of the respective ontologies
4. Calculating the similarity propagation given the taxonomies and initial relatedness scores (SimFlood)
5. Calculating the final similarity scores
6. Filtering

In step (2), we use vector based representations of an ESA (Explicit Semantic Analysis [1]) style approach adapted to the biomedical domain related use. The representations are created for inputs that are the labels of individual concepts. The distributional representations are obtained through a combined use of Wikipedia and a domain-focused corpus of scientific documents, i.e. Medline.

In step (3), we use the vectors from step (2) to calculate the relatedness approximation as the cosine similarity of these vectors. To calculate the similarity propagation in step (4), we use the very basic version of the algorithm applied to the taxonomic structures. We do however restrict the propagation graph size by not including the nodes that do not surpass a certain minimal initial relatedness threshold.

We calculate the final similarity scores (step 5) as an average between the initial scores (semantic relatedness) and the similarity propagation output. This gives more importance to the relatedness score (which is the point of our experiment), and also caters for cases in which Similarity Flooding is poorly applicable.

The filtering is done by: i) accepting only a maximal number of candidate matches per node of an ontology; ii) eliminating candidate matches below a certain similarity threshold; iii) accepting a globally maximal number of candidate matches.

### 1.3 Adaptations made for the evaluation

No specific adaptations were made for the experiments, apart from minor changes of the filtering parameters (i.e. the global number of candidate matches accepted in the final alignment).

### 1.4 Link to the set of provided alignments

The set of provided alignments is available in URL `http://bit.ly/2dPA9H5`

## 2 Results of the Disease and Phenotype track

DisMatch has been evaluated in both tasks of the Disease and Phenotype track: HP-MP (alignment of Human Phenotype Ontology with Mammalian Phenotype Ontology) and DOID-ORDO (alignment of Human Disease Ontology with Orphanet Rare Disease Ontology). A summary of results is reported in the Official site of OAEI 2016::Disease and Phenotype Track[1].

[1] In URL http://oaei.ontologymatching.org/2016/results/phenotype/.

**Table 1.** Unique mappings in the HP-MP task

| OM Algorithm | Unique Equivalence Mappings | Precision (Manual Assessment) | Positive Contribution (TP) | Negative Contribution (FP) |
|---|---|---|---|---|
| AML | 122 | 0.8667 | 8.63% | 1.33% |
| DisMatch | **291** | 0.8333 | **19.80%** | 3.96% |
| FCA_Map | 26 | 0.9615 | 2.04% | 0.08% |
| LogMap | 130 | 0.9330 | 9.90% | 0.71% |
| LogMapLite | 0 | 0.0000 | 0.00% | 0.00% |
| LogMapBio | 176 | 0.9330 | 13.40% | 0.96% |
| LYAM++ | 226 | 0.7000 | 12.91% | 5.53% |
| PhenoMF | 89 | 1.0000 | 7.27% | 0.00% |
| PhenoMM | 85 | 1.0000 | 6.94% | 0.00% |
| PhenoMP | 80 | 1.0000 | 6.53% | 0.00% |
| XMap | 0 | 0.0000 | 0.00% | 0.00% |
| **Totals** | **1225** | | **87.42%** | **12.58%** |

It can be observed that the results of DisMatch are relatively far off the silver standard created in the evaluation process. We believe that this is largely due to setting up the system with parameters that resulted in overly strict filtering that created a relatively low number of mappings. In turn, the low number of mappings led to poor recall, both in the silver standard evaluation and w.r.t. the set of manually created mappings.

The precision of DisMatch in the HP-MP alignment looks quite promising, especially if we consider the number of unique alignments produced by the system. Out of the total of 644 mappings, 353 mappings are confirmed by at least one another system (thus falling into 'correct' category in the silver standard 2). Out of these 353, 293 are confirmed by at least 2 other systems ('correct' in silver standard 3). The remaining 291 mapping are unique to DisMatch. Table 1 presents an overview of unique mappings produced by the respective systems. The precision of the unique mappings produced by Dismatch is estimated at 0.8333, which accounts for a large portion of unique and correct mappings discovered by our system. In this regard, the proposed approach obtained the highest percentage of positive contribution (19.80%), with a relatively low negative contribution (3.96%).

In the case of DOID-ORDO alignment, the performance of our system is limited, as it is affected not only by the low recall related to the poor parameter selection, but also by the inability of our structural mapping component to cope with the structure of the Orphanet ontology. This shortcoming will be addressed in the future versions of DisMatch. Nonetheless, as shown in Table 2, even in this setting, the system managed to produce a considerable number (estimated 40% of 259 is $> 100$) of correct unique mappings.

**Table 2.** Unique mappings in the DOID-ORDO task

| OM Algorithm | Unique Equivalence Mappings | Precision (Manual Assessment) | Positive Contribution (TP) | Negative Contribution (FP) |
|---|---|---|---|---|
| AML | 308 | 0.8667 | 30.40% | 4.68% |
| DisMatch | 259 | 0.4000 | 11.80% | 17.70% |
| FCA_Map | 61 | 0.8330 | 5.79% | 1.16% |
| LogMap | 80 | 0.9000 | 8.20% | 0.91% |
| LogMapLite | 7 | 0.5000 | 0.40% | 0.40% |
| LogMapBio | 144 | 0.9667 | 15.85% | 0.55% |
| LYAM++ | 0 | 0.0000 | 0.00% | 0.00% |
| PhenoMF | 3 | 1.0000 | 0.34% | 0.00% |
| PhenoMM | 0 | 0.0000 | 0.00% | 0.00% |
| PhenoMP | 0 | 0.0000 | 0.00% | 0.00% |
| XMap | 16 | 0.5625 | 1.03% | 0.80% |
| **Totals** | **878** | | **87.42%** | **12.58%** |

## 3 General comments

Relatedness measure seems to capture non-trivial matches better than, for example, string edit distance. At the same time, it still works for the trivial cases, as common words will generate similar distributional representations. The main strength of DisMatch (and its distributional semantic relatedness component) is its ability of finding non-trivial mappings, which seems to be confirmed by the number of unique correct matches generated by the system (and the unique-to-total mappings ratio).

Nonetheless, the structural matching strategy still seems to be an important component of the system, as the relatedness matcher itself will, for example, generate high confidence matches for inputs, such as 'X syndrome' and 'Y syndrome', if X and Y are very rare in the background corpus. The importance of the structural matching step seems to be consistent with the performance gap between HP-MP (where the structural matcher worked) and DOID-ORDO (where it did not work properly) cases.

We believe that DisMatch could be improved substantially through improving the relatedness-structure matching combination, i.e. by employing a better suited structural matcher. Furthermore, our current structural matching strategy relied solely on strictly taxonomic relationships, which is not always enough (i.e. in the case of the OrphaNet ontology).

Furthermore, semantic relatedness module generates candidate mappings that are not necessarily 'equivalent', as the measure does not distinguish between the possible relationship types. It is worth considering adding an additional 'prediction' module to provide a classification output of the relationship type of the mappings.

Moreover, when it comes to improving the performance of the relatedness module itself, it seems that the measure provides more accurate results for

shorter input texts. This points to two possible improvements: (a) in finding a better suited compositional approach for the lexical relatedness measure, or (b) in using shorter inputs (possibly through synonym properties of the ontologies to be aligned).

## 4 Conclusions

The results obtained with the DisMatch system show enough promise to continue the experiments with corpus-based distributional relatedness measures applied to the problem of ontology alignment. We believe, that our focus should now be on providing an optimal set of additional components around the relatedness measure. In addition, we expect that tuning of the filtering parameters will lead the proposed system to reach higher precision with respect to silver standards.

## Acknowledgments

This work has been partially funded by Grants TIN2014-58304-R (Spanish Ministry of Education and Science) and P11-TIC-7529 (Innovation, Science and Enterprise Ministry of the regional government of the Junta de Andalucía) and P12-TIC-1519 (Plan Andaluz de Investigación, Desarrollo e Innovación). José Garía-Nieto is recipient of a Post-Doctoral fellowship of "Captación de Talento para la Investigación" at Universidad de Málaga.

## References

1. Evgeniy Gabrilovich and Shaul Markovitch. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *IJCAI*, volume 7, pages 1606–1611, 2007.
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. Ted Pedersen, Serguei V S Pakhomov, Siddharth Patwardhan, and Christopher G Chute. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of biomedical informatics*, 40(3):288–299, 2007.
4. Maciej Rybinski and José F Aldana-Montes. Calculating semantic relatedness for biomedical use in a knowledge-poor environment. *BMC bioinformatics*, 15(Suppl 14):S2, 2014.