

Exploiting the UMLS Metathesaurus in the Ontology Alignment Evaluation Initiative

Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, and Ian Horrocks

Department of Computer Science, University of Oxford
{ernesto, berg, ian.horrocks}@cs.ox.ac.uk

Abstract. In this paper we describe how the UMLS Metathesaurus—the most comprehensive effort for integrating medical thesauri and ontologies—is being used within the context of the Ontology Alignment Evaluation Initiative (OAEI). We also present the obtained results in the Large BioMed track of the OAEI 2011.5 campaign where the reference alignments are based on UMLS. Finally, we propose a new reference alignment based on the harmonisation of the outputs of the systems participating in the OAEI Large BioMed track.

1 Introduction

The Ontology Alignment Evaluation Initiative¹ (OAEI) is an international campaign for the systematic evaluation of ontology matching systems —software programs capable of finding correspondences (called *alignments*) between the vocabularies of a given set of input ontologies [22, 7, 9, 23]. The matching problems in the OAEI are organised in several tracks, with each track involving different kinds of test ontologies [7]. The ontologies in the largest test case in the OAEI 2011 contain only 2,000–3,000 classes; however, ontology matching tools have significantly improved in the last few years and there is a need for more challenging and realistic matching problems for which suitable reference alignments exist [22, 7].

UMLS-Metathesaurus (UMLS) [1] is currently the most comprehensive effort for integrating medical thesauri and ontologies, including the National Cancer Institute Thesaurus (NCI) [12, 11], the Foundational Model of Anatomy (FMA) [19] and the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [24], which are large-scale and semantically rich ontologies. NCI, FMA and SNOMED CT are gradually superseding the existing medical classifications and are becoming core platforms for accessing, gathering, and sharing biomedical knowledge and data. Hence, matching such large ontologies represents a very interesting challenge for the OAEI initiative.

In this paper we describe how the UMLS correspondences between NCI, FMA and SNOMED CT have been used as reference alignments for the new *Large BioMed track*² in the OAEI initiative. Furthermore we present the results obtained in the OAEI 2011.5 campaign for this track and we propose a new reference alignment based on the harmonisation of the outputs of the participating ontology matching systems.

¹ <http://oaei.ontologymatching.org/>

² <http://www.cs.ox.ac.uk/isg/projects/SEALS/oaei/>

Table 1. The notion of “Joint” in the MRCONSO file from the UMLS distribution.

CUI	Language	Source	Entity
C0022417	ENG	FMA	Joint
			Set_of_joints
		SNOMED CT	Joint_structure
		NCI	Joint
			Articulation

Table 2. UMLS-based alignment between FMA, NCI and SNOMED CT for the notion of “Joint”.

Ontology pair	Generated Alignments
FMA ~ NCI	$\langle 1, FMA:Joint, NCI:Joint, 1.0, equiv \rangle$ $\langle 2, FMA:Joint, NCI:Articulation, 1.0, equiv \rangle$ $\langle 3, FMA:Set_of_joints, NCI:Joint, 1.0, equiv \rangle$ $\langle 4, FMA:Set_of_joints, NCI:Articulation, 1.0, equiv \rangle$
FMA ~ SNOMED CT	$\langle 5, FMA:Joint, SNOMED:Joint_structure, 1.0, equiv \rangle$ $\langle 6, FMA:Set_of_joints, SNOMED:Joint_structure, 1.0, equiv \rangle$
SNOMED CT ~ NCI	$\langle 7, SNOMED:Joint_structure, NCI:Joint, 1.0, equiv \rangle$ $\langle 8, SNOMED:Joint_structure, NCI:Articulation, 1.0, equiv \rangle$

2 The UMLS-based reference alignments

Ontology alignments are often conceptualised as tuples with the form $\langle id, e_1, e_2, n, \rho \rangle$, where id is a unique identifier for the mapping, e_1, e_2 are entities in the vocabulary of the integrated ontologies, n is a numeric confidence measure between 0 and 1, and ρ is a relation between e_1 and e_2 , typically subsumption (i.e., e_1 is more specific than e_2) and equivalence (i.e., e_1 and e_2 are synonyms) [8]. The OAEI initiative uses an RDF format to represent the alignments³ [6] containing the aforementioned elements. Alternatively, OAEI alignments are also represented as OWL 2 subclass and equivalence axioms with the mapping identifier (id) and confidence (n) added as OWL 2 annotation axioms [4].

Although the standard UMLS distribution does not directly provide sets of alignments (in the OAEI sense) between the integrated ontologies, it is relatively straightforward to extract alignment sets from the information provided in the distribution files [15]. Concretely, we have processed the *MRCONSO*⁴ file, which contains every entity in UMLS together with its *concept unique identifier* (CUI), its source vocabulary (e.g. FMA), its language (e.g. English), and other attributes not relevant for the OAEI. Table 1 shows an excerpt from the *MRCONSO* file associated to the notion of “Joint”.

It follows from Table 1 that the notion of “Joint” is shared by FMA, SNOMED CT and NCI. In particular, FMA contains the entities *Joint* and *Set_of_joints*, NCI the entities *Articulation* and *Joint*, and SNOMED CT only the entity *Joint_structure*. All these entities have been annotated with the same CUI C0022417 and therefore, according to UMLS’s intended meaning, they are synonyms. Then, for each pair of entities e_1 and e_2 from *different* sources and annotated with the same CUI, we have

³ <http://alignapi.gforge.inria.fr/format.html>

⁴ <http://www.ncbi.nlm.nih.gov/books/n/nlmumls/ch03/>

Table 3. UMLS-based alignments

Ontology pair	Original alignments	Unsatisfiabilities	Refined alignments
FMA ~ NCI	3,024	655	2,898
FMA ~ SNOMED CT	9,072	6,179	8,111
SNOMED CT ~ NCI	19,622	20,944	18,322

Table 4. Results for the Large BioMed track in the OAEI 2011.5 campaign.

System	Size	Unsat.	Refined UMLS			Original UMLS			Time (s)
			P	R	F	P	R	F	
LogMap	2,658	9	0.868	0.796	0.830	0.875	0.769	0.819	126
GOMMA _{bk}	2,983	17,005	0.806	0.830	0.818	0.826	0.815	0.820	1,093
GOMMA _{nobk}	2,665	5,238	0.845	0.777	0.810	0.862	0.759	0.807	960
LogMapLt	3,466	26,429	0.675	0.807	0.735	0.695	0.796	0.742	57
CSA	3,607	$>10^5$	0.514	0.640	0.570	0.528	0.629	0.574	14,068
Aroma	4,080	$>10^5$	0.467	0.657	0.546	0.480	0.647	0.551	9,503
MapSSS	2,440	33,186	0.426	0.359	0.390	0.438	0.353	0.391	$>10^5$

generated the corresponding (equivalence) UMLS-based alignments with a confidence value of 1.0 (see Table 2).

The integration of new resources in UMLS combines expert assessment and sophisticated auditing protocols [1, 3, 10]. However, it has been noticed that UMLS-based alignments lead to a large number of unsatisfiable classes if they are represented as OWL 2 axioms and integrated with the input ontologies [15, 14]. For example the integration of SNOMED CT and NCI via UMLS-based alignments leads to more than 20,000 unsatisfiable classes. To address this problem, we have presented in [14] a refinement of the (original) UMLS-based alignments that do not lead to (many) unsatisfiable classes (see Table 3). This refinement is based on the alignment repair module of the ontology matching system LogMap [14, 16].

3 Results of the Large BioMed track in the OAEI 2011.5

In this section we briefly present the obtained results in the Large BioMed track of the OAEI 2011.5 campaign.⁵ We have only evaluated the FMA-NCI matching problem, where the used versions of FMA and NCI contains 78,989 and 66,724 classes, respectively. The original and refined UMLS-based alignments (see Table 3) has been used as reference to evaluate the efficiency of participating ontology matching systems.

Table 4 summarizes the obtained results where systems has been ordered according to the F-measure against the refined UMLS-based reference alignment. LogMapLt—a simple ontology matcher—has been used as a base-line. Besides precision (P), recall (R), F-measure (F) and runtimes we have also evaluated the coherence of the alignments when reasoning together with the input ontologies.⁶ Note that we have evaluated

⁵ <http://www.cs.ox.ac.uk/isg/projects/SEALS/oei/2011.5/>

⁶ We have used the OWL 2 reasoner HermiT [20]

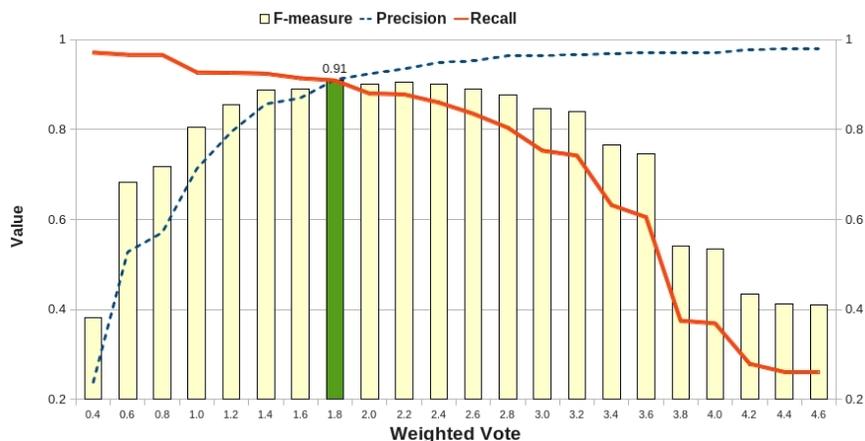


Fig. 1. Harmonised alignments for the FMA-NCI matching problem of the OAEI 2011.5.

GOMMA [17] with two different configurations. $GOMMA_{bk}$ uses UMLS-based background knowledge, while $GOMMA_{nobk}$ has this feature deactivated.

GOMMA (with its two configurations) and LogMap are a bit ahead in terms of F-measure with respect to Aroma [5], CSA [25] and MapSSS [2], which could not top the results of the base-line LogMapLt. $GOMMA_{bk}$ obtained the best results in terms of recall, while LogMap provided the best results in terms of precision and F-measure. The use of the original UMLS-based reference alignment did not imply important variations. Since the original set contains more mappings, precision and recall slightly increases and decreases, respectively. It is worth mentioning, however, that $GOMMA_{bk}$ improves its results when comparing with the original UMLS-based reference alignment and provides the best F-measure.

Regarding mapping coherence, only LogMap generated an ‘almost’ clean output in all three tasks. Although $GOMMA_{nobk}$ also provides highly precise output correspondences, they lead to a huge amount of unsatisfiable classes.

4 Towards a silver standard reference alignment

The original UMLS-based reference alignment, as shown in Section 2, contains errors (i.e. lead to large number of unsatisfiable classes when integrated with the input ontologies). On the other hand, the refined UMLS-based reference alignment is based on the (incomplete) alignment repair techniques of the ontology matching systems LogMap [14, 16], which may fail to detect and discard the appropriate alignments. Thus, in order to turn the extracted UMLS-based reference alignments into an agreed-upon gold standard expert assessment would be needed, which is almost unfeasible for large alignment sets. We have opted to move towards a *silver standard* by harmonising the outputs of different matching tools over the relevant ontologies. Similar silver standards have been developed for named entity recognition problems [21, 13].

We have harmonised the outputs of the systems participating in the OAEI 2011.5 FMA-NCI matching problem. Each system has been associated a weighted vote based on its precision w.r.t. the refined UMLS-based reference alignment (see Table 4). For example, LogMap and MapSSS have been associated the weights 0.868 and 0.426, respectively. Note that systems participating with two versions (e.g. GOMMA and LogMap) have been only considered once in the voting process.

Figure 1 summarises the evolution of the F-measure, Precision and Recall for the harmonised alignment depending on the minimum required votes. For example the harmonised alignment set requiring 4.0 points of weighted votes has a precision of 0.971 and a recall of 0.369 w.r.t. the refined UMLS-based reference alignment. As expected precision increases and recall decreases as the required votes increase.

We have selected the harmonised alignment set with the highest F-measure (0.91) as the “first” silver standard of the FMA-NCI matching problem. This set contains 2,890 alignments that have been “at least” voted by two systems with weight 0.90. Note that this harmonised alignment has not been yet refined and it is known to lead to more than 14,000 unsatisfiable classes when integrated with FMA and NCI.

5 Future work

In the OAEI 2012 campaign⁷ we also intend to evaluate the SNOMED-NCI and FMA-SNOMED matching problems using the correspondent UMLS-based reference alignments (see Table 3). We will also create harmonised silver standards alignments and we will evaluate the participating systems against them. This comparison will be very useful to analyse how different a system is with respect to the others.

Finally, we also intend to combine different reasoning and diagnosis tools such as ALCOMO⁸ [18] to generate error-free refinements of both the UMLS-based reference alignments and the harmonised silver standards.

Acknowledgements

This work was supported by the Royal Society, the EU FP7 project SEALS and by the EPSRC projects ConDOR, ExODA, LogMap and Score!. We also thank the organisers of the OAEI evaluation campaigns for providing test data and infrastructure.

References

1. Bodenreider, O.: The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic acids research* 32 (2004)
2. Cheatham, M.: MapSSS results for OAEI 2011. In: *Proceedings of the 6th Ontology Matching Workshop*. pp. 184–189 (2011)
3. Cimino, J.J., Min, H., Perl, Y.: Consistency across the hierarchies of the UMLS semantic network and metathesaurus. *J of Biomedical Informatics* 36(6) (2003)

⁷ <http://www.cs.ox.ac.uk/isg/projects/SEALS/oaei/2012/>

⁸ <http://web.informatik.uni-mannheim.de/alcomo/>

4. Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U.: OWL 2: The next step for OWL. *Journal of Web Semantics* 6(4), 309–322 (2008)
5. David, J., Guillet, F., Briand, H.: Association Rule Ontology Matching Approach. *Journal of Semantic Web Information Systems* 3(2), 27–49 (2007)
6. David, J., Euzenat, J., Scharffe, F., dos Santos, C.T.: The Alignment API 4.0. *Semantic Web* 2(1), 3–10 (2011)
7. Euzenat, J., Meilicke, C., Stuckenschmidt, H., Shvaiko, P., Trojahn, C.: Ontology Alignment Evaluation Initiative: six years of experience. *J Data Semantics* (2011)
8. Euzenat, J.: Semantic precision and recall for ontology alignment evaluation. In: *Proc. of the 20th International Joint Conference on Artificial Intelligence, IJCAI*. pp. 348–353 (2007)
9. Euzenat, J., Ferrara, A., van Hage, W.R., Hollink, L., Meilicke, C., Nikolov, A., Ritze, D., Scharffe, F., Shvaiko, P., Stuckenschmidt, H., Sváb-Zamazal, O., Trojahn dos Santos, C.: Results of the Ontology Alignment Evaluation Initiative 2011. 6th OM workshop (2011)
10. Geller, J., Perl, Y., Halper, M., Cornet, R.: Special issue on auditing of terminologies. *Journal of Biomedical Informatics* 42(3), 407–411 (2009)
11. Golbeck, J., Fragoso, G., Hartel, F.W., Hendler, J.A., Oberthaler, J., Parsia, B.: The National Cancer Institute’s Thésaurus and Ontology. *J. Web Sem.* 1(1), 75–80 (2003)
12. Hartel, F.W., de Coronado, S., Dionne, R., Fragoso, G., Golbeck, J.: Modeling a description logic vocabulary for cancer research. *Journal of Biomedical Informatics* 38(2) (2005)
13. Jiménez-Ruiz, E., Rebholz-Schuhmann, D., Lewin, I.: Exploitation of cross-references between terminological resources within the CALBC context. In: *1st Intl. Workshop on Exploiting Large Knowledge Repositories, DEXA Workshops* (2011)
14. Jiménez-Ruiz, E., Cuenca Grau, B.: LogMap: Logic-based and Scalable Ontology Matching. In: *Proc. of the 10th International Semantic Web Conference (ISWC)*. pp. 273–288 (2011)
15. Jiménez-Ruiz, E., Cuenca Grau, B., Horrocks, I., Berlanga, R.: Logic-based assessment of the compatibility of UMLS ontology sources. *J Biomed. Sem.* 2 (2011)
16. Jiménez-Ruiz, E., Cuenca Grau, B., Zhou, Y., Horrocks, I.: Large-scale interactive ontology matching: Algorithms and implementation. In: *Proc. of ECAI* (2012)
17. Kirsten, T., Gross, A., Hartung, M., Rahm, E.: GOMMA: a component-based infrastructure for managing and analyzing life science ontologies and their evolution. *Journal of Biomedical Semantics* 2, 6 (2011)
18. Meilicke, C.: Alignment Incoherence in Ontology Matching. Ph.D. thesis, University of Mannheim, Chair of Artificial Intelligence (2011)
19. Mejino Jr., J.L.V., Rosse, C.: Symbolic modeling of structural relationships in the foundational model of anatomy. In: *Proc. of First International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004)*. pp. 48–62 (2004)
20. Motik, B., Shearer, R., Horrocks, I.: Hypertableau Reasoning for Description Logics. *Journal of Artificial Intelligence Research* 36, 165–228 (2009)
21. Rebholz-Schuhmann, D., Jimeno Yepes, A., Van Mulligen, E.M., Kang, N., Kors, J., Milward, D., Corbett, P., Buyko, E., Beisswanger, E., Hahn, U.: CALBC Silver Standard Corpus. *J Bioinform Comput Biol*. pp. 163–179 (2010)
22. Shvaiko, P., Euzenat, J.: Ten challenges for ontology matching. In: *On the Move to Meaningful Internet Systems (OTM Conferences)* (2008)
23. Shvaiko, P., Euzenat, J.: Ontology matching: State of the art and future challenges. *IEEE Trans. Knowl. Data Eng.* 99 (2011)
24. Spackman, K.: SNOMED RT and SNOMED CT. Promise of an international clinical ontology. *M.D. Computing* 17 (2000)
25. Tran, Q.V., Ichise, R., Ho, B.Q.: Cluster-based similarity aggregation for ontology matching. In: *Proc. of 6th Ontology Matching Workshop*. pp. 142–147 (2011)