

Ontology Matching Techniques: A Gold Standard Model

Alok Chauhan, Vijayakumar V , Layth Sliman

Abstract. Typically an ontology matching technique is a combination of much different type of matchers operating at various abstraction levels such as structure, semantic, syntax, instance etc. An ontology matching technique which employs matchers at all possible abstraction levels is expected to give, in general, best results in terms of precision, recall and F-measure due to improvement in matching opportunities and if we discount efficiency issues which may improve with better computing resources such as parallel processing. A gold standard ontology matching model is derived from a model classification of ontology matching techniques. A suitable metric is also defined based on gold standard ontology matching model. A review of various ontology matching techniques specified in recent research papers in the area was undertaken to categorize an ontology matching technique as per newly proposed gold standard model and a metric value for the whole group was computed. The results of the above study support proposed gold standard ontology matching model.

Keywords: Ontology matching, gold standard, metric etc.

I. INTRODUCTION

Reported slowing down in speed of improvement in the field of ontology matching is the motivation behind present work [1]. It requires a fresh look into the field of ontology matching. Ontology matching can be performed at various levels, criteria, and environments leading to different kind of techniques (Fig. 1). It may be done either locally at the element level or globally at the structure level of ontologies. Matching criterion could be semantic, syntactic, terminological, structural and extensional based. Matching environment could be either context or content based. Various legal combinations of matching levels, criteria and environments give rise to whole range of concrete techniques as shown in Fig.1. As all these techniques are complimentary to each other, various matching systems use a combination of these techniques.

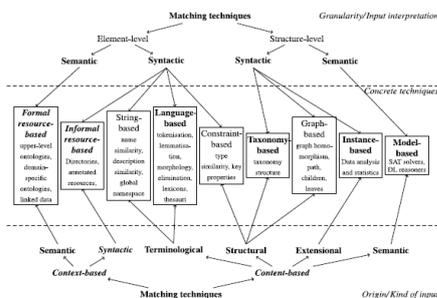


Fig. 1. A model classification of ontology matching techniques (Source: [2], p. 77)

Revised Manuscript Received on December 22, 2018.

Alok Chauhan, School of Computing Science & Engineering, VIT University, Chennai;

Vijayakumar V, School of Computing Science & Engineering, VIT University, Chennai;

Layth Sliman, EFREI, PARIS alok.chauhan@vit.ac.in, vijayakumar.v@vit.ac.in, layth.sliman@efrei.fr

But it leads to following questions with respect to above model classification of ontology matching techniques:

1. Are there any relationship among various matching levels, criteria and environments
2. Can we have any guideline to combine all the different aspects/ dimensions of ontology matching as shown in model classification, assuming it leads to some kind of synergy approximating as gold standard
3. Instead of arbitrary selection of concrete techniques for an ontology matching system, is there any pattern corresponding to a holistic concept of meaning

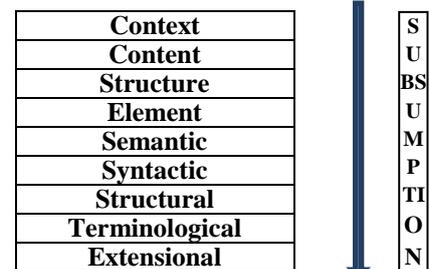
In the following sections, we will address these issues.

II. RELATED WORK

Traditionally, almost all ontology matching surveys [3-6] focussed on classification/disintegration rather than integration, hence model proposed in this paper is a novel idea, though this integrative model would not have been possible in the absence of excellent surveys made earlier. The closest approach to the current work [7] classifies matching techniques in terms of three layers, viz., data layer, ontology layer and context layer which does have some resemblance to proposed model but goals of the two approaches are entirely different.

III. GOLD STANDARD MODEL

Intuitively, on the basis of above mentioned classification, one might even propose a gold standard model for an ontology matching system in terms of a layered view of ontology matching criteria as shown in Fig. 2. This new model is a structured regrouping of generic ontology matching criteria in the form of a layered representation having subsumption relationship among successive layers from top to bottom. In a way the new model is derived from model classification by integrating generic matching criteria leading to a comprehensive state-of-the-art representation of ontology matching or it looks like a compressed/folded version of model classification. As per the proposed model (Fig. 2), by “Gold Standard”, we mean an ontology matching system that includes all the nine layers (Context, Content,, Extensional).



Ontology Matching Techniques: A Gold Standard Model

Above model represents a holistic/universal view of meaning in terms of existing constructs, though at present we do not have any methodology to support this model but an effort in this direction has been made by the authors [8] and process to propose a new foundation ontology is underway.

IV. COVERAGE METRIC

An appropriate metric may also be proposed to measure the coverage of ontology matching criteria by a class of ontology matching systems as per proposed model. Proposed metric called C-measure (Coverage-measure) is as follows:

$$C - Measure = \frac{\sum(\text{frequency of ontology matching techniques})}{9 \times (\text{number of matching systems})}$$

Fig. 3. Formula for coverage metric

V. PROPOSED STUDY

The proposed comparative study of ontology matching techniques is based on the models shown in Fig. 1 and Fig. 2. For this purpose, recent research papers from www.ontologymatching.org are used as sample and titles of these papers are searched manually on the basis of appropriate keywords. The choice of keywords may seem arbitrary but the logic behind using these keywords in such a manner is to cover all the concrete techniques given in Fig. 1 and to avoid overlapping among different layers of the proposed model (Fig. 2). The whole effort is to derive an intuitive ontology matching model from the well known classification of matching techniques (Fig. 1). The approaches mentioned in these papers are classified manually on the basis of models mentioned above. Results of proposed study are found to be in accordance with proposed gold standard ontology matching model.

5.1 Extensional Layer

It is the innermost layer of the proposed model (Fig. 2). The online repository (<http://www.ontologymatching.org/publications.html>) was searched with keyword = "instance". An appropriate search word is helpful in getting more results. Overall some twenty-eight (28) research papers were reviewed in this category and chart below shows frequency distribution of various parameters/elements of the proposed gold standard model (Fig. 4, 5).

A	B	C	D	E	F	G	H	I	J	K
keyword = (Instance)	context	content	structure	element	semantic	syntactic	structural	terminological	extensional	
1. H. Senthilguru, B. Prasad, Deb Nath, M. Aisoo: An efficient metric of automatic weight generation for properties in instance matching technique. <i>IJWest</i> , 1 2015		P	P	P		P	P			
2. S. Zeng, T. Schuster, S. Conrad: Instance-Based Ontology Matching Using Different Kinds of Formalisms World Academy of Science, 12 Engineering and Technology, 2009.		P	P			P				
3. context-dependent semantic similarity among ontology instances. <i>Journal on Data Semantics</i> , 17 2008.		P	P	P	P	P	P			
4. L. Serafini, A. Tamplin: Instance Migration in Heterogeneous Ontology Environments. In: <i>Proceedings of SWACIS</i> , 2007		P	P	P	P	P	P			
5. P. Bernstein, S. Melnik, P. Mork: Interactive Schema Translation with Instance-Level Mapping. In: <i>Proceedings of VLDB</i> , 2005		P	P			P				
6. Instance-based Schema Matching for Web Databases by Domain-specific Query Probing. In: <i>Proceedings of VLDB</i> , 2004		P	P			P				
7. S. Ichise, H. Takeda, S. Hoshida: Integrating Multiple Internet Directories by Instance-based Learning. In: <i>Proceedings of IJCAI</i> , 2003		P	P			P				

Fig. 4. Analysis of extensional layer

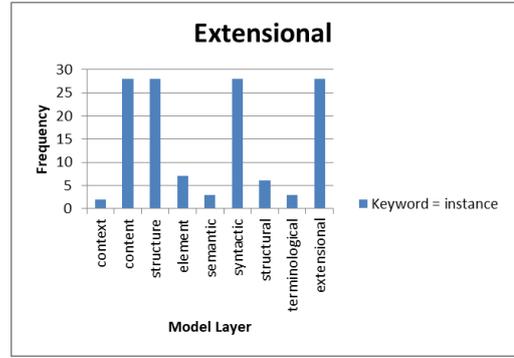


Fig. 5. Coverage of extensional ontology matching techniques

What we can see from above chart is that very few research papers dealing with instance-based techniques really include all the components of the proposed model, i.e., from context to extensional. One can as well see the emphasis on certain components (content, structure, syntactic) as compared to other components. The emphasis on extensional component is obvious in this case.

5.2 Terminological Layer

This layer is above extensional layer and online repository is searched with keywords such as "string", "lexic"ons and "thesaur"i and nineteen (19) research papers were reviewed (Fig. 6, 7).

A	B	C	D	E	F	G	H	I	J	K
keyword = string, lexic, thesaur	context	content	structure	element	semantic	syntactic	structural	terminological	extensional	
1. Y. Sun, L. Ma, S. Wang: A Comparative Evaluation of String Similarity Metrics for Ontology Alignment. <i>PLoS</i> , 2014		P				P				
2. A. Wang, G. Li, D. Deng, Y. Zhang, J. Feng: Two Birds With One Stone: An Efficient Hierarchical Framework for Top-k and Threshold-based String Similarity Search. In: <i>Proceedings of KDD</i> , 2012		P				P				
3. Z. Yang, J. Yu, M. Kitsuragawa: Fast Algorithms for Top-k Approximate String Matching. In: <i>Proceedings of AAAI</i> , 2010		P				P				
4. R. Barua, S. Banerjee, A. Voinov: TUCON: The Intelligent Hunter Agent for Concept Understanding and Lexical Chaining. In: <i>Proceedings of ISMIS</i> , 2009		P				P				
5. T. Sathya, A. Selamat, M. Ashraf, T. Heryawan: Effect of thesaurus size on schema matching quality. <i>Knowledge-Based Systems</i> , 2014		P				P				
6. W. A. van Hege, M. Sim, L. Finch, M. Kulis, G. Schreiber: The OAH food task: an analysis of a thesaurus alignment task. <i>Applied Ontology</i> , 2010		P				P				
7. A. Tordella, B. Compton, G. Schreiber: Thesaurus and Metadata Alignment for a Semantic E-Culture Application. In: <i>Proceedings of I-CAP</i> , 2007		P				P				
8. A. Li, Y. Li, M. Sim: Mapping AGROVOC and the Chinese Agricultural Thesaurus: Definitions, tools, procedures. <i>New Review of Hypermedia and Multimedia</i> , 2006		P				P				
9. A. Li, M. Sim, C. Chum, L. Siling, L. Wainlin, H. Chunhui, L. Kejun: The Mapping Schema from Chinese Agricultural Thesaurus to AGROVOC. <i>New Review of Hypermedia and Multimedia</i> , 2006		P				P				

Fig. 6. Analysis of terminological layer

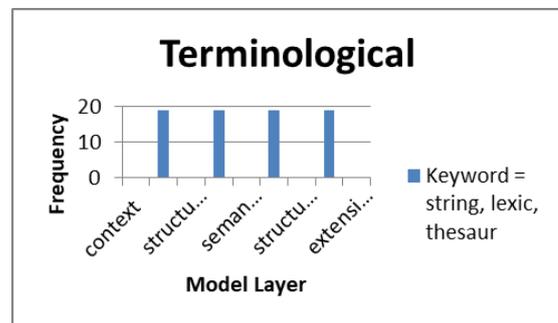


Fig. 7. Coverage of terminological ontology matching techniques

Though no concrete inferences can be made out of above analysis (Fig. 6), it just suggests, intuitively, the inclusion of left over components in ontology matching techniques covered under various layers from a "Gold Standard" point of view. We do find gaps in coverage of various layers in this category (Terminological) of matching techniques.



5.3 Structural Layer

Online repository is searched with keywords such as “structural”, “constraint”, “taxonom”y and “graph”. Largest number of papers (42) were reviewed under this category (Fig. 8). The chart (Fig. 9) below shows the frequency distribution of various matching criteria.

Fig. 8. Analysis of structural layer

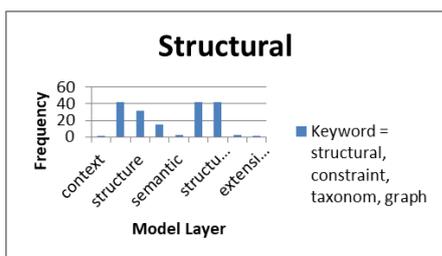


Fig. 9. Coverage of structural ontology matching techniques

Ontology matching techniques under this category seems to favor certain aspects of matching over others and this gap may be filled to achieve better results. Coverage of layers is not as good as it was in extensional techniques.

5.4 Syntactic Layer

Keywords such as “terminological” and “annot”ated were used for the purpose of searching and it was noticed that term “terminological” is being used in a different way (such as a directory) as opposed to what we may infer from Fig. 1 (String based, Language based). Some fifteen (15) research papers were reviewed under this category (Fig. 10). The chart for this layer is shown below (Fig. 11).

Fig. 10. Analysis of syntactic layer

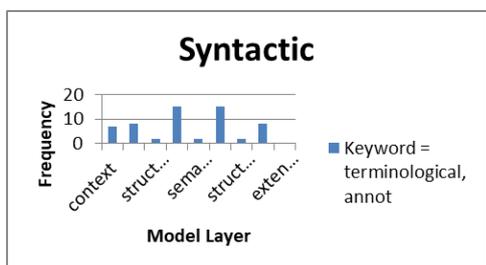


Fig. 11. Coverage of syntactic ontology matching techniques

5.5 Semantic Layer

Online repository (http://www.ontologymatching.org/publications.html) was searched with keywords “background”, “upper”, “context”, and “sat” to get more results (22) for this layer/ category (Fig. 12). Coverage wise, we get the best results here (Fig. 13).

Fig. 12. Analysis of semantic layer

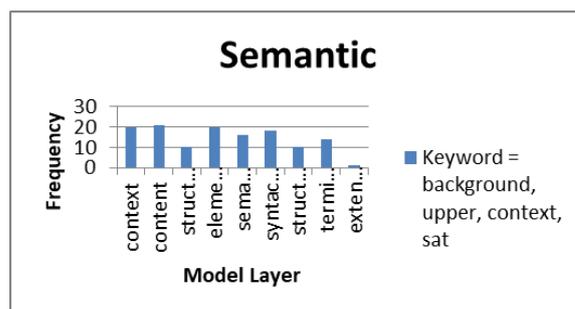


Fig. 13. Coverage of semantic ontology matching techniques

There is no need to search for remaining layers (Element, Structure, Content and Context) of proposed model (Fig. 2) as all the concrete techniques of standard matching technique classification model (Fig. 1) have already been covered in previous layers.

VI. RESULTS AND CONCLUSION

Given below is metric (C-measure) computation of various layers as per above mentioned study:

Sl. No.	Layer	C-measure
1	Extensional	0.527778
2	Terminological	0.444444
3	Structural	0.470899
4	Syntactic	0.437037
5	Semantic	0.656566

Fig. 14. C-measure computation for various types of ontology matching techniques

It is interesting to note that concrete techniques at the beginning (Formal resource-based) and end (Instance-based and Model-based) of the spectrum (Fig. 1) fare much better with respect to proposed coverage metric, as most of the ontology matching systems falling under these categories employ most of the available ontology matching criteria. It appears as if two ends of gold standard model induce a comprehensive coverage of matching criteria due to in-built



subsumption or reverse subsumption relationship. It is followed by Structural and Terminological techniques in that order (defined as per chosen keywords). As per our metric, Syntactic techniques (defined as per selected keywords) come last. Results of study reaffirm our conjecture that ontology matching techniques with higher C-measure score are best placed as per our proposed gold standard model and could be a good indicator towards the need of converging/ unifying various ontology matching efforts towards a common model as proposed by us and not just incremental improvements in individual domains. This ranking has nothing to do with individual techniques' performance as per existing vertical metrics (Precision, Recall, and F-Measure), though it has already been reported in literature that ontology matching using background knowledge does improve match result [9]. The aim of this study is to complement existing vertical metrics with newly proposed horizontal coverage metric. It is the combination of horizontal (C-measure) as well as vertical (recall, precision and F-measure) which is expected to give much better and consistent evaluation of match results. Also, present study in a way reconfirms the standard ontology matching classification model (Fig. 1).

VII. LIMITATIONS OF STUDY

Due to inductive nature of hypothesis, study based on just one repository is sufficient but it can be expanded to include more resources. Similarly, manual review of papers may not be a limitation as automation may lead to compromise with quality of results.

VIII. FUTURE WORK

As future work, proposed model may be applied to popular matching systems and results may be analyzed to prove the utility of coverage metric. OAEI results may also be analyzed from this new perspective and requirement of unification of various ontology matching approaches may be assessed and emphasized.

REFERENCES

1. P. Shvaiko, J. Euzenat: Ontology matching: state of the art and future challenges IEEE Transactions on Knowledge and Data Engineering, 2013
2. J. Euzenat, P. Shaviko: Ontology Matching, Springer-Verlag, Heidelberg (2013)
3. Rahm, E., Bernstein, P.: A survey of approaches to automatic schema matching. VLDB J. 10(4), 334–350 (2001)
4. Shvaiko, P., Euzenat, J.: A survey of schema-based matching approaches. J. Data Semant. IV, 146–171 (2005)
5. Doan, A.-H., Halevy, A.: Semantic integration research in the database community: a brief survey. AI Mag. 26(1), 83–94 (2005)
6. Zanobini, S.: Semantic coordination: the model and an application to schema matching. PhD thesis, University of Trento, Trento, Italy (2006)
7. Ehrig, M.: Ontology Alignment: Bridging the Semantic Gap. Springer, New York (2007)
8. A. Chauhan, V. Vijayakumar, R. Ragalac: Towards a multi-level upper ontology-foundation ontology framework as background knowledge for ontology matching problem Procedia Computer Science, 2015

9. Z. Aleksovski, M. Klein, W. ten Kate, F. van Harmelen: Matching Unstructured Vocabularies using a Background Ontology In Proceedings of EKAW, 2006