FOAM – Framework for Ontology Alignment and Mapping Results of the Ontology Alignment Evaluation Initiative

Marc Ehrig
Institute AIFB
University of Karlsruhe
76128 Karlsruhe, Germany
ehrig@aifb.uni-karlsruhe.de

York Sure
Institute AIFB
University of Karlsruhe
76128 Karlsruhe, Germany
sure@aifb.uni-karlsruhe.de

ABSTRACT

This paper briefly introduces the system FOAM and its underlying techniques. We then discuss the results returned from the evaluation. They were very promising and at the same time clarifying. Concisely: labels are very important; structure helps in cases where labels do not work; dictionaries may provide additional evidence; ontology management systems need to deal with OWL-Full. The results of this paper will also be very interesting for other participants, showing specific strengths and weaknesses of our approach.

1. PRESENTATION OF THE SYSTEM

1.1 State, purpose, general statement

In recent years, we have seen a range of research work on methods proposing alignments [1; 2]. When we tried to apply these methods to some of the real-world scenarios we address in other research contributions [3], we found that existing alignment methods did not suit the given requirements:

- high quality results;
- efficiency;
- optional user-interaction;
- flexibility with respect to use cases;
- and easy adjusting and parameterizing.

We wanted to provide the end-user with a tool taking ontologies as input and returning alignments (with explanations) as output meeting these requirements.

1.2 Specific techniques used

We have observed that alignment methods like QOM [4] or PROMPT [2] may be mapped onto a generic alignment process (Figure 1). Here we will only mention the six major steps to clarify the underlying approach for the FOAM tool. We refer to [4] for a detailed description.

1. Feature Engineering, i.e. select excerpts of the overall ontology definition to describe a specific. This includes individual features, e.g. labels, structural features, e.g. subsumption, but also more complex features as used in OWL, e.g. restrictions.
2. Search Step Selection, i.e. choose two entities from the two ontologies to compare (e_1,e_2).
3. Similarity Assessment, i.e. indicate a similarity for a given description (feature) of two entities (e.g., $sim_{superConcept}(e_1,e_2)=1.0$).
4. Similarity Aggregation, i.e. aggregate the multiple similarity assessments for one pair of entities into a single measure.
5. Interpretation, i.e. use all aggregated numbers, a threshold and an interpretation strategy to propose the alignment (align(e_1)=‘ e_2’). This may also include a user validation.
6. Iteration, i.e. as the similarity of one alignment influences the similarity of neighboring entity pairs; the equality is propagated through the ontologies.

Finally, we receive alignments linking the two ontologies.

This general process was extended to meet the mentioned requirements.

- High quality results were achieved through a combination of a rule-based approach and a machine learning approach. Underlying individual rules such as, if the super-concepts are similar the entities are similar, have been assigned weights by a machine learnt decision tree [5]. Especially steps 1, 3 and 4 were adjusted for this. Currently, our approach does not make use of additional background knowledge such as dictionaries here.
- Efficiency was mainly achieved through an intelligent selection of candidate alignments in 2, the search step selection [4].
- User-interaction allows the user intervening during the interpretation step. By presenting the doubtable alignments (and only these) to the user, overall quality can be considerably increased. Yet this happens in a minimal invasive manner.
- The system can automatically set its parameters according to a list of given use cases, such as ontology merging, versioning, ontology mapping, etc. The parameters also change according to the ontologies to align, e.g., big ontologies always require the efficient approach, whereas smaller ones do not [6].
- All these parameters may be set manually. This allows using the implementation for very specific tasks as well.
- Finally, FOAM has been implemented in Java and is freely available, thus extensible.

1.3 Adaptations made for the contest

No special adjustments have been made for the contest. However, some elements have been deactivated. Due to the small size of the benchmark and directory ontologies efficiency was not used, user-interaction was removed for the initiative, and no specific use case parameters were taken. A general alignment procedure was applied.

The system used for the evaluation is a derivative of the ontology alignment tool used in last year’s contests I3Con [7] and EON-OAC [8].

2. RESULTS

All tests were performed on a standard notebook under Windows. FOAM has been implemented in Java with all its advantages and disadvantages.

The individual results of the benchmark ontologies were grouped. Further, one short section describes the testing of the directory and anatomy ontologies. The concrete results can be found in Section 6.3 of this paper.

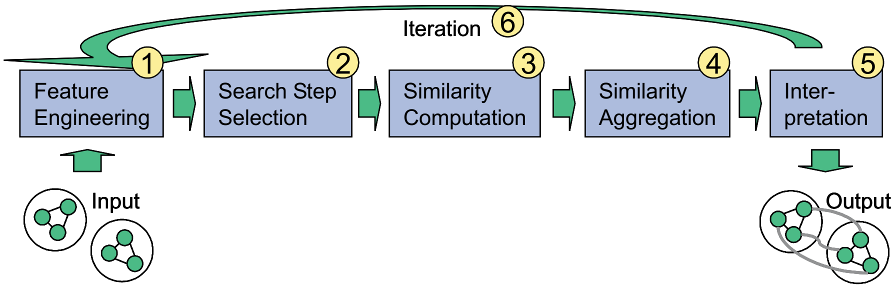

Figure 1: Ontology Alignment Process

2.1.1 Tests 101 to 104

These tests are basic tests for ontology alignment.

As the system assumes that equal URIs mean equal objects an alignment of an ontology with itself always returns the correct alignments. The alignment with and irrelevant ontology does not return any results. Language generalization or restriction does not affect the results. Our approach is robust enough to cope with these differences. Considering the differences which occur in real world ontology modeling this is a very desirable feature.

2.1.2 Tests 201 to 210

Tests 201 through 210 focus on labels and comments of ontological entities.

The labels are the most important feature to identify an alignment. In fact, everything else can be neglected, if the labels indicate an alignment (e.g. also the comments in Test 203). Vice versa, changed labels do seriously affect the outcomes. As our approach currently does not make use of any dictionaries, this is critical. Small changes as occurring through a different naming convention can be balanced-out (Test 204 is only slightly worse than the ideal result). Synonyms or translations, possibly also with removed comments, lower especially recall considerably (between 0.57 and 0.87). Nevertheless, the structure alignment does find many of the alignments, despite the differing labels. For the mentioned recalls, precision stays between 0.80 and 0.96.

2.1.3 Tests 221 to 247

For all these tests the structure is changed.

However, as the labels remain, alignment is very good. Again, this indicates that labels are the main distinguishing feature. Only smaller irritations result from the differing structures. In specific, more false positives are identified resulting in a precision of in the worst case “only” 0.94. Recall stays above0.97. According to the amount of structure also the processing time changes. Please note that first results are returned almost instantaneously (less than 5 seconds). The times presented in the table represent the total time until the approach stops its search for alignments.

2.1.4 Tests 248 to 266

These tests were the most challenging ones for our approach. Labels and comments had been removed and different structural elements as well.

Precision reaches levels of 0.61 to 0.95. Recall is in the range of 0.18 to 0.55. Unfortunately, the evaluation results did not show a clear tendency of which structural element is most important for our alignment approach. It seems that the structural features can be exchanged to a certain degree. If one feature is missing, evidence is collected from another feature. This is a nice result for our approach, as it indicates that the weighting scheme of the individual features has been assigned correctly. One tendency that could be identified was that with decreasing semantic information the found alignments become sparser. However, most of the identified alignments were correct (see precision).

We will briefly mention one test for which our approach performed surprisingly well. Ontology 262 has practically everything removed: no labels; no comments; no properties; no hierarchies. Nevertheless, some alignments have been identified. The only information that remained was the links between instances and their classes. By checking whether instance sets were the same (at least in terms of numbers, the instance labels actually differed), some concepts could be correctly aligned.

2.1.5 Tests 301 to 304

Ontologies 301 through 304 represent schemas modeled by other institutions but covering the same domain of bibliographic metadata. From the evaluation perspective, these real world ontologies combine the difficulties of the previous tests.

Especially test case 301 differs both in terms of structure and labels. Its labels generally use the term “has”, i.e. “hasISBN” instead of “ISBN”. This results in a rather low term similarity, as our approach does not split the strings into individual terms. Combined with the differing structure this results in a rather low quality. Also for the other ontologies, both precision and recall do not reach perfect levels. However, the results are satisfactory. In fact, preliminary tests using our semi-automatic approach showed that results could be noticeably increased with very little effort. The question that will partially also be answered by this initiative, is what can maximally be reached. We hope to gain these insights by comparing our results to other participants’ results.

2.2 Directory Ontologies

The directory ontologies are subsumption hierarchies. They could be easily processed. The evaluation results at the workshop will presumably show the following main effects: Subsumption helps to identify some alignments correctly. Our missing usage of dictionaries misses some alignments. As this dataset only uses subsumption, we cannot rely on the more complex ontology features which our approach normally also tries to exploit. Thus, results will not be ideal.

2.3 Anatomy Ontologies

We were very interested in running our ontology alignment on the big real world anatomy ontologies. Especially for our efficient approach, this would have been a deep evaluation. Unfortunately, the ontologies were modeled in OWL-Full. Our approach is based on the KAON2-infrastructure[1] that only allows for OWL-DL. As this interaction is very deep, it was not possible to change to an ontology environment capable of OWL-Full for the contest. We could not run these tests. One result, for us, was the realization that ontologies will probably not stay in the clean world of OWL-DL. We will have to draw consequences from this.

3. GENERAL COMMENTS

3.1 Comments on the results

An objective comment on strengths or weakness requires the comparison with other participants, which will not be available before the workshop. However, some conclusions can be drawn.

Strengths:

- Labels or identifiers are important and help to align most of the entities.
- The structure helps to identify alignments, if the labels are not expressive.
- A more expressive ontology results in better alignments; an argument in favor of ontologies compared to simple classification structures.
- The generally learnt weights have shown very good results.

Weaknesses:

- The approach cannot deal with consequently changed labels. Especially translations, synonyms, or other conventions make it difficult to identify alignments.
- The system is bound to OWL-DL or lesser ontologies.

3.2 Discussions on the way to improve the proposed system

Possible improvements are directly related to the weaknesses in the previous section.

- Extending the handling of labels (strings) can presumably increase overall effectiveness. Usage of dictionaries is widely applied and will be added to our approach as well.
- The tight interconnection of FOAM with KAON2 restricts the open usage of it. Currently efforts are being made to decouple them by inserting a general ontology management layer.

3.3 Comments on the test cases

The benchmark tests have shown very interesting general results on how the alignment approach behaves. These systematic tests are one good underlying test base. For our approach, the directory tests are less interesting, as they are restricted to subsumption hierarchies, rather than complete ontologies. Many of the specific advantages of our approach cannot be applied. It was very unfortunate, that we could not run the anatomy tests. However, we think it is very important to have some real world ontologies, and we hope to test them at a latter point in time.

For future work, it might be interesting to add some user-interaction component to the tests. It would also be interesting to not only have real world ontologies, but also see which alignment approach performs how for specific ontology alignment applications.

3.4 Comments on the measures

Precision and recall are without any doubt the most important measures. Some balancing measure needs to be added as well, as we have done with the f-measure. Otherwise, it is very difficult to draw conclusions on which approach worked best on which test set. For future evaluation it would also be interesting to make use of some less strict evaluation measure, as presented in [9].

4. CONCLUSION

In this paper, we have briefly presented an approach and a tool for ontology alignment and mapping - FOAM. This included the general underlying process. Further, we have mentioned how specific requirements are realized with this tool. We then applied FOAM to the test data. The results were carefully analyzed. We also discussed some future steps for both our own approach and the evaluation of alignments in general.

The main conclusions from the experiments were:

- It is possible to create a good automatic ontology alignment approaches.
- Labels are most important.
- Structure helps, if the labels are not expressive.
- Due to the importance of labels, our approach needs to be extended with e.g. dictionaries in the background.
- One general conclusion from the real world ontologies, was that an ontology system has to be able to also manage OWL-Full, as the real world does not provide the clean ontologies of OWL-DL.

In general, the evaluation has shown us where our specific strengths and weaknesses are, and how we can continue on improving. The results of other participants will give us some further guidelines.

5. REFERENCES

[1] Agrawal, R., Srikant, R.: On integrating catalogs. In: Proceedings of the Tenth International Conference on the World Wide Web (WWW-10), ACM Press (2001) 603–612

[2] Noy, N.F., Musen, M.A.: The PROMPT suite: interactive tools for ontology merging and mapping. International Journal of Human-Computer Studies 59 (2003) 983–1024

[3] Ehrig, M., Haase, P., van Harmelen, F., Siebes, R., Staab, S., Stuckenschmidt, H., Studer, R., Tempich, C.: The SWAP data and metadata model for

[1] http://kaon2..semanticweb.org

semantics-based peer-to-peer systems. In: Proceedings of MATES-2003. First German Conference on Multiagent Technologies. LNAI, Erfurt, Germany, Springer (2003)

[4] Ehrig, M., Staab, S. QOM - quick ontology mapping. In F. van Harmelen, S. McIlraith, and D. Plexousakis, editors, Proceedings of the Third International Semantic Web Conference (ISWC2004), LNCS, pages 683–696, Hiroshima, Japan, 2004. Springer.

[5] Ehrig, M., Staab, S., Sure, Y. Supervised learning of an ontology alignment process. In Proceedings of the Workshop on IT Tools for Knowledge Management Systems: Applicability, Usability, and Benefits (KMTOOLS) at 3. Konferenz Professionelles Wissensmanagement, Kaiserslautern, Germany, April 2005.

[6] Ehrig, M., Sure, Y. Adaptive Semantic Integration. In Proceedings of the Workshop on Ontologies-based techniques for DataBases and Information Systems at VLDB 2005, Trondheim, Norway, August 2005.

[7] Hughes, T. Information Interpretation and Integration Conference (I3CON) at PerMIS-2004, Gaithersburg, MD, USA, August 2004.

[8] Sure, Y., Corcho, O., Euzenat, J., Hughes, T. (editors), 3rd International Workshop on Evaluation of Ontology based Tools (EON2004), Volume 128, CEUR-WS Publication. Workshop at the 3rd International Semantic Web Conference (ISWC 2004), 7th-11th November 2004, Hiroshima, Japan

[9] Ehrig, M., Euzenat, J. Relaxed Precision and Recall for Ontology Alignment. In Proceedings of the Integrating Ontologies Workshop at K-Cap ’05, Banff, Alberta, Canada, October 2005.

6. RAWRESULTS

6.1 Link to the system and parameters file

The FOAM system may be downloaded at

http://www.aifb.uni-karlsruhe.de/WBS/meh/foam.

The system is continuously improved, so results may slightly differ from the results provided in this paper. The interested reader is encouraged to download, test, and use the system.

6.2 Link to the set of provided alignments (in align format)

The results are also available through the website: http://www.aifb.uni-karlsruhe.de/WBS/meh/foam/results.zip.

6.3 Matrix of results

The following results were achieved in the evaluation runs. As FOAM only allows identifying equality relations, precision and recall only refer to these.

#	Name	Prec.	Rec.	F-measure	Time
101	Reference alignment	1.0	1.0	1.0	2.96
102	Irrelevant ontology	-	-	-	207.14
103	Language generalization	1.0	1.0	1.0	180.95
104	Language restriction	1.0	1.0	1.0	177.63
201	No names	0.90	0.65	0.75	175.99
202	No names, no comments	0.85	0.57	0.68	176.59
203	No comments	1.0	1.0	1.0	174.21
204	Naming conventions	0.96	0.93	0.94	185.09
205	Synonyms	0.80	0.67	0.73	174.46
206	Translation	0.93	0.76	0.84	172.15
207		0.95	0.78	0.86	167.89
208		0.96	0.87	0.92	164.20
209		0.81	0.57	0.67	168.63
210		0.92	0.67	0.77	164.31
221	No specialization	1.0	1.0	1.0	172.92
222	Flattened hierarchy	1.0	1.0	1.0	127.63
223	Expanded hierarchy	0.99	1.0	0.99	142.70
224	No instance	1.0	0.99	0.99	42.09
225	No restrictions	1.0	1.0	1.0	171.13
228	No properties	1.0	1.0	1.0	112.60
230	Flattened classes	0.94	1.0	0.97	137.60
232		1.0	0.99	0.99	45.50
233		1.0	1.0	1.0	110.57
236		1.0	1.0	1.0	12.77
237		1.0	1.0	1.0	87.94
238		1.0	1.0	1.0	106.29
239		0.94	1.0	0.97	73.14
240		0.95	0.97	0.97	84.63
241		1.0	1.0	1.0	11.15
246		0.94	1.0	0.97	51.14
247		0.94	1.0	0.97	70.27
248		0.85	0.48	0.62	251.65
249		0.73	0.46	0.57	150.39
250		0.95	0.55	0.69	114.00
251		0.88	0.41	0.56	132.39
252		0.62	0.34	0.44	145.59

253		0.80	0.44	0.57	83.96
254		0.75	0.18	0.29	103.56
257		0.76	0.48	0.59	28.43
258		0.86	0.39	0.53	133.79
259		0.75	0.45	0.56	149.39
260		0.85	0.38	0.52	71.21
261		0.61	0.33	0.43	82.89

262		0.78	0.21	0.33	21.70
265		0.85	0.38	0.52	70.50
266		0.63	0.36	0.46	81.68
301	BibTeX/MIT	0.78	0.35	0.48	23.43
302	BibTeX/UMBC	0.88	0.74	0.80	21.31
303	Karlsruhe	0.84	0.90	0.87	61.08
304	INRIA	0.94	0.97	0.95	43.32