Lily Results on SEALS Platform for OAEI 2011

Peng Wang

School of Computer Science and Engineering, Southeast University, China
pwang@seu.edu.cn

Abstract. This paper presents the alignment results of Lily on SEALS platform for the ontology alignment contest OAEI 2011. Lily is an ontology matching system. In OAEI 2011, Lily submited the results for three matching tasks on the SEALS platform: benchmark, anatomy, conference. The specific techniques used by Lily are introduced. The matching results of Lily are also discussed.

1   Presentation of the system

Lily is an ontology matching system for solving the key issues related to heterogeneous ontologies, and it uses hybrid matching strategies to execute the ontology matching task. Lily can be used to discovery alignments for both normal size ontologies and large scale ontologies. In past ontology alignment contests, Lily showed it is a good ontology matching system and obtained good results in some tasks [1-3].

1.1   State, purpose, general statement

The core principle of the matching strategy in Lily is utilizing the useful information effectively and correctly. Lily combines several novel and efficient matching techniques to find alignments. Lily has three main matching functions: (1) Generic Ontology Matching (GOM) is used for common matching tasks with normal size ontologies. (2) Large scale Ontology Matching (LOM) is used for the matching tasks with large size ontologies. (3) Ontology mapping debugging is used to verify and improve the alignment results.

   The matching process mainly contains three steps: (1) In preprocess, Lily parses ontologies and prepares the necessary information for subsequent steps. (2) In similarity computation step, Lily uses special methods to calculate the similarities between elements from different ontologies. (3) In post-process, alignments are extracted and refined by mapping debugging.

   In OAEI2011, we redesign or modify some key algorithms, and we also make it can be run on the SEALS platform.

1.2 Specific techniques used

Lily aims to provide high quality 1:1 concept pair or property pair alignments. The main specific techniques used by Lily are as follows.

Semantic subgraph: An element may have heterogeneous semantic interpretations in different ontologies. Therefore, understanding the real local meanings of elements is very useful for similarity computation, which are the foundations for many applications including ontology matching. Therefore, before similarity computation, Lily first describes the meaning for each entity accurately. However, since different ontologies have different preferences to describe their elements, obtaining the semantic context of an element is an open problem. We proposed the semantic subgraph to capture the real meanings of ontology elements [4]. To extract the semantic subgraphs, a hybrid ontology graph is used to represent the semantic relations between elements. An extracting algorithm based on an electrical circuit model is then used with new conductivity calculation rules to improve the quality of the semantic subgraphs. We have showed that the semantic subgraphs can properly capture the local meanings of elements [4].

Based on the extracted semantic subgraphs, we can build more credible matching clues. Therefore it can reduce the negative affection of the matching uncertain.

Generic ontology matching method: The similarity computation is based on the semantic subgraphs, i.e. all the information used in the similarity computation is come from the semantic subgraphs. Lily combines the text matching and structure matching techniques.

Semantic Description Document (SDD) matcher measures the literal similarity between ontologies. A semantic description document of a concept contains the information about class hierarchies, related properties and instances. A semantic description document of a property contains the information about hierarchies, domains, ranges, restrictions and related instances. For the descriptions from different entities, we calculate the similarities of the corresponding parts. Finally, all separate similarities are combined with the experiential weights.

Matching weak informative ontologies: Most existing ontology matching methods are based on the linguistic information. However, some ontologies have not sufficient or regular linguistic information such as natural words and comments, so the linguistic-based methods cannot work. Structure-based methods are more practical for this situation. Similarity propagation is a feasible idea to realize the structure-based matching. But traditional propagation does not take into consideration the ontology features and will be faced with effectiveness and performance problems. We analyze the classical similarity propagation algorithm Similarity Flood and propose a new structure-based ontology matching method [5]. This method has two features: (1) It has more strict but reasonable propagation conditions which make matching process become more efficient and alignments become better. (2) A series of propagation strategies are used to improve the matching quality. We have demonstrated that this method performs well on the OAEI benchmark dataset [5].

However, the similarity propagation is not always perfect. When more alignments are discovered, more incorrect alignments would also be introduced by the similarity propagation. So Lily also uses a strategy to determine when to use the similarity propagation.

***Large scale ontology matching*:** Matching large ontologies is a challenge due to the high time complexity. We propose a new matching method for large ontologies based on reduction anchors [6]. This method has a distinct advantage over the divide-and-conquer methods because it does not need to partition large ontologies. In particular, two kinds of reduction anchors, positive and negative reduction anchors, are proposed to reduce the time complexity in matching. Positive reduction anchors use the concept hierarchy to predict the ignorable similarity calculations. Negative reduction anchors use the locality of matching to predict the ignorable similarity calculations. Our experimental results on the real world data sets show that the proposed method is efficient for matching large ontologies [6].

Ontology mapping debugging Lily uses a technique called ontology mapping debugging to improve the alignment results [7]. Different from existing methods, which focus on finding efficient and effective solutions for the ontology mapping problem, mapping debugging emphasis on analyzing the mapping result to detect/diagnose the mapping defects. We proposed a technique called debugging ontology mappings [7]. During debugging, some types of mapping errors, such as redundant and inconsistent mappings, can be detected. Some warnings, including imprecise mappings or abnormal mappings, are also locked by analyzing the features of mapping result. More importantly, some errors and warnings can be repaired automatically or can be presented to users with revising suggestions.

1.3   Adaptations made for the evaluation

Lily is fully automatic in OAEI2011, and there is no any parameter turning during the matching. Lily has a simple strategy to choose the right matching method.

1.4   Link to the system and the set of provided alignments

Lily system for OAEI 2011 is available at http://cse.seu.edu.cn/people/pwang/ software/Lily/lily-package.zip

2   Results

2.1   benchmark

The Benchmark2010 dataset can be divided into five groups: 101-104, 201-210, 221-247, 248-266 and 301-304.

101-104 Lily plays well for these test cases.

201-210 Lily can produce good results for this test set. Even without right labels and comments information, Lily can find most correct alignments through making use of other information such as instances. Using few alignment results obtained by the basic methods as inputs, the similarity propagation strategy will generate more alignments.

221-247 Lily can find most correct alignments using the labels and comments information.

248-266 This group is the most difficult test set. Lily first uses the GOM method to find alignments, and then use matching weak informative method to discover more alignments.

301-304 This test set are the real ontologies. Lily only finds the equivalent alignment relations.

The following table shows the average performance of each group and the overall performance on the Benchmark2010 dataset.

Table 1. The results on the Benchmark2010

	101-104	201-210	221-247	248-266	301-304	Average
Precision	1.00	0.92	0.98	0.81	0.71	0.86
Recall	1.00	0.82	0.99	0.51	0.69	0.66
F1-Measure	1.00	0.85	0.97	0.59	0.70	0.71

The BenchmarkII2011 dataset can be divided into three groups: 101-102, 221-247 and 248-266. It seems that this task is more difficult than Benchmark2010. Our alignment results have small decrease in average.

The following table shows the average performance of each group and the overall performance on the BenchmarkII2011 dataset.

Table 2. The results on the BenchmarkII2011

	101-202	221-247	248-266	Average
Precision	0.92	0.96	0.74	0.79
Recall	0.65	0.98	0.55	0.63
F1-Measure	0.70	0.97	0.61	0.67

2.2 anatomy

The anatomy track consists of two real large-scale biological ontologies. Lily can handle such ontologies smoothly with LOM method. Task#1 means that the matching system has to be applied with standard settings to obtain a result that is as good as possible. Table 3 shows the performance of the task #1 on anatomy dataset.

Compared to the result in OAEI2009, our result has little increase. However, it has obvious gap to the results of other matching system in OAEI2010.

Table 3. The performance on the anatomy

	Runtime	Precision	Recall	F1-Measure
Task#1	20min	0.80	0.72	0.76

2.4 conference

This task contains 16 real world ontologies about conference. We only get part of alignments from SEALS platform, which is showed in Table 4. The heterogeneous

character in this track is various. It is a challenge to generate good results for all ontology pairs in this test set.

Table 4. The performance on the conference based on reference mappings

Ontology Pair	Precision	Recall	F1-Measure
cmt-confOf	0.46	0.38	0.41
cmt-conference	0.21	0.25	0.23
cmt-edas	0.29	0.54	0.38
cmt-ekaw	0.25	0.45	0.32
cmt-iasted	0.18	0.50	0.27
cmt-sigkdd	0.27	0.25	0.26
confOf-edas	0.57	0.42	0.48
confOf-ekaw	0.72	0.65	0.68
confOf-iasted	0.38	0.67	0.48
confOf-sigkdd	0.09	0.14	0.11
conference-confOf	0.55	0.73	0.63
conference-edas	0.16	0.29	0.20
conference-ekaw	0.38	0.32	0.35
conference-iasted	0.39	0.50	0.44
conference-sigkdd	0.32	0.47	0.38
edas-ekaw	0.50	0.52	0.51
edas-iasted	0.31	0.47	0.37
edas-sigkdd	0.47	0.53	0.50
ekaw-iasted	0.30	0.70	0.42
ekaw-sigkdd	0.33	0.45	0.38
iasted-sigkdd	0.43	0.67	0.53

3. General comments

We redesign some key algorithms of Lily this year, but it does not produce better alignment results than previous versions. For example, we try to use mapping debugging technique to improve the precision of results. However, for the reason that the generated results are 1:1 equivalent mappings, the mapping debugging can only find few wrong alignments.

The SEALS platform is very important for ontology matching research. It provides a way to examine new matching method and compare to other matching systems.

4 Conclusion

We briefly introduce our ontology matching tool Lily. The matching process and the special techniques used by Lily are presented. The alignment results are carefully analyzed.

References

1. Peng Wang, Baowen Xu: Lily: ontology alignment results for OAEI 2009. In The 4th International Workshop on Ontology Matching, Washington Dc., USA (2009)
2. Peng Wang, Baowen Xu: Lily: Ontology Alignment Results for OAEI 2008. In The Third International Workshop on Ontology Matching, Karlsruhe, Germany (2008)

3. Peng Wang, Baowen Xu: LILY: the results for the ontology alignment contest OAEI 2007. In The Second International Workshop on Ontology Matching (OM2007), Busan, Korea (2007)
4. Peng Wang, Baowen Xu, Yuming Zhou: Extracting Semantic Subgraphs to Capture the Real Meanings of Ontology Elements. Journal of Tsinghua Science and Technology, vol. 15(6), pp. 724-733 (2010)
5. Peng Wang, Baowen Xu: An Effective Similarity Propagation Model for Matching Ontologies without Sufficient or Regular Linguistic Information, In The 4th Asian Semantic Web Conference (ASWC2009), Shanghai, China (2009)
6. Peng Wang, Yuming Zhou, Baowen Xu: Matching Large Ontologies Based on Reduction Anchors. In The Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011), Barcelona, Catalonia, Spain (2011)
7. Peng Wang, Baowen Xu: Debugging Ontology Mapping: A Static Approach. Computing and Informatics, vol. 27(1), pp. 21–36 (2008)

Appendix: Raw results

The final results of benchmark task are as follows.

Matrix of results

#	Comment	Prec.	Rec.	#	Comment	Prec.	Rec.
101	Reference alignment	1.00	1.00	251		1.00	0.08
103	Language generalization	1.00	1.00	251-2		0.93	0.83
104	Language restriction	1.00	1.00	251-4		0.91	0.74
201	No names	0.96	0.96	251-6		0.94	0.67
201-2		1.00	1.00	251-8		0.88	0.49
201-4		1.00	1.00	252		0.29	0.02
201-6		0.98	0.98	252-2		0.92	0.82
201-8		1.00	1.00	252-4		0.89	0.79
202	No names, no comment	0.43	0.03	252-6		0.90	0.80
202-2		0.95	0.86	252-8		0.92	0.82
202-4		0.94	0.76	253		0.33	0.02
202-6		0.94	0.67	253-2		0.94	0.80
202-8		0.92	0.51	253-4		0.92	0.71
203	Misspelling	0.98	0.98	253-6		0.95	0.64
204	Naming conventions	1.00	1.00	253-8		0.91	0.44
205	Synonyms	1.00	0.99	254		0.00	0.00
206	Translation	1.00	0.99	254-2		0.84	0.64
207		1.00	0.99	254-4		0.95	0.55
208		0.96	0.93	254-6		0.92	0.36
209		0.71	0.53	254-8		0.86	0.18
210		0.71	0.53	257		1.00	0.03
221	No specialisation	1.00	1.00	257-2		0.94	0.88
222	Flattened hierarchy	1.00	1.00	257-4		0.93	0.76
223	Expanded hierarchy	0.98	0.98	257-6		0.75	0.55

ID	Description	V1	V2	ID	Description	V3	V4
224	No instances	1.00	1.00	257-8		0.92	0.36
225	No restrictions	1.00	1.00	258		0.57	0.04
228	No properties	1.00	1.00	258-2		0.93	0.83
230	Flattened classes	0.94	1.00	258-4		0.92	0.75
231	Expanded classes	1.00	1.00	258-6		0.97	0.69
232		1.00	1.00	258-8		0.94	0.52
233		1.00	1.00	259		0.29	0.02
236		1.00	1.00	259-2		0.91	0.81
237		1.00	1.00	259-4		0.91	0.81
238		0.97	0.97	259-6		0.89	0.79
239		0.97	1.00	259-8		0.89	0.79
240		0.94	0.97	260		0.50	0.03
241		1.00	1.00	260-2		0.85	0.79
246		0.97	1.00	260-4		0.87	0.69
247		0.91	0.94	260-6		0.88	0.52
248		0.43	0.03	260-8		0.85	0.38
248-2		0.93	0.79	261		0.50	0.03
248-4		0.95	0.73	261-2		0.83	0.76
248-6		0.95	0.64	261-4		0.83	0.76
248-8		0.90	0.44	261-6		0.83	0.76
249		0.71	0.05	261-8		0.84	0.79
249-2		0.95	0.86	262		0.00	0.00
249-4		0.91	0.74	262-2		0.88	0.67
249-6		0.96	0.68	262-4		0.89	0.52
249-8		0.91	0.49	262-6		0.92	0.36
250		1.00	0.03	262-8		0.86	0.18
250-2		0.94	0.88	265		0.50	0.03
250-4		0.93	0.76	266		0.50	0.03
250-6		0.83	0.61	301	BibTeX/MIT	0.82	0.76
250-8		0.92	0.36	302	BibTeX/UMBC	0.54	0.42
				303	Karlsruhe	0.57	0.65
				304	INRIA	0.89	0.93