

# Lily Results for OAEI 2015

Wenyu Wang<sup>1,2</sup>, Peng Wang<sup>1</sup>

<sup>1</sup> School of Computer Science and Engineering, Southeast University, China

<sup>2</sup> Chien-Shiung Wu College, Southeast University, China

{ms, pwang} @ seu.edu.cn

**Abstract.** This paper presents the results of Lily in the ontology alignment contest OAEI 2015. As a comprehensive ontology matching system, Lily is intended to participate in four tracks of the contest: benchmark, conference, anatomy, and instance matching. The specific techniques used by Lily will be introduced briefly. The strengths and weaknesses of Lily will also be discussed.

## 1 Presentation of the system

With the use of hybrid matching strategies, Lily, as an ontology matching system, is capable of solving some issues related to heterogeneous ontologies. It can process normal ontologies, weak informative ontologies [5], ontology mapping debugging [7], and ontology matching tuning [9], in both normal and large scales. In previous OAEI contests [1–3], Lily has achieved preferable performances in some tasks, which indicated its effectiveness and wideness of availability.

### 1.1 State, purpose, general statement

The core principle of matching strategies of Lily is utilizing the useful information correctly and effectively. Lily combines several effective and efficient matching techniques to facilitate alignments. There are five main matching strategies: (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) Instance Ontology Matching (IOM) is used for instance matching tasks. (4) Ontology mapping debugging is used to verify and improve the alignment results. (5) Ontology matching tuning is used to enhance overall performance.

The matching process mainly contains three steps: (1) Pre-processing, when Lily parses ontologies and prepares the necessary information for subsequent steps. Meanwhile, the ontologies will be generally analyzed, whose characteristics, along with studied datasets, will be utilized to determine parameters and strategies. (2) Similarity computing, when Lily uses special methods to calculate the similarities between elements from different ontologies. (3) Post-processing, when alignments are extracted and refined by mapping debugging.

In this year, some algorithms and matching strategies of Lily have been modified for higher efficiency, and adjusted for brand-new matching tasks like Author Recognition and Author Disambiguation in the Instance Matching track.

## 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. The semantic subgraph was proposed 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. It has been shown that the semantic subgraphs can properly capture the local meanings of elements [4].

Based on the extracted semantic subgraphs, more credible matching clues can be discovered, which help reduce the negative effects of the matching uncertainty.

**Generic ontology matching method** The similarity computation is based on the semantic subgraphs, which means all the information used in the similarity computation comes 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, the similarities of the corresponding parts will be calculated. Finally, all separated similarities will be combined with the experiential weights.

**Matching weak informative ontologies** Most existing ontology matching methods are based on the linguistic information. However, some ontologies may lack in regular linguistic information such as natural words and comments. Consequently the linguistic-based methods will not work. Structure-based methods are more practical for such situations. Similarity propagation is a feasible idea to realize the structure-based matching. But traditional propagation strategies do not take into consideration the ontology features and will be faced with effectiveness and performance problems. Having analyzed the classical similarity propagation algorithm, *Similarity Flood*, we proposed a new structure-based ontology matching method [5]. This method has two features: (1) It has more strict

but reasonable propagation conditions which lead to more efficient matching processes and better alignments. (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 its significant time complexity. We proposed 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 datasets show that the proposed methods are efficient in matching large ontologies [6].

**Ontology mapping debugging** Lily utilizes a technique named *ontology mapping debugging* to improve the alignment results [7]. Different from existing methods that focus on finding efficient and effective solutions for the ontology mapping problems, mapping debugging emphasizes on analyzing the mapping results to detect or diagnose the mapping defects. 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.

**Ontology matching tuning** Lily adopted ontology matching tuning this year. By performing parameter optimization on training datasets [9], Lily is able to determine the best parameters for similar tasks. Those data will be stored. When it comes to real matching tasks, Lily will perform statistical calculations on the new ontologies to acquire their features that help it find the most suitable configurations, based on previous training data. In this way, the overall performance can be improved.

Currently, ontology matching tuning is not totally automatic. It is difficult to find out typical statistical parameters that distinguish each task from others. Meanwhile, learning from test datasets can be really time-consuming. Our experiment is just a beginning.

### 1.3 Adaptations made for the evaluation

For benchmark, anatomy and conference tasks, Lily is totally automatic, which means Lily can be invoked directly from the SEALS client. It will also determine which strategy to use and the corresponding parameters. For a specific instance matching task, Lily needs to be configured and started up manually, so only matching results were submitted.

### 1.4 Link to the system and parameters file

SEALS wrapped version of Lily for OAEI 2015 is available at <https://drive.google.com/file/d/0B4fqkE38d3QrS1Zta0pPSFpqXzA/view?usp=sharing>.

### 1.5 Link to the set of provided alignments

The set of provided alignments, as well as overall performance, is available at each track of the OAEI 2015 official website, <http://oaei.ontologymatching.org/2015/>.

## 2 Results

### 2.1 Benchmark track

There are two datasets in different sizes: *Biblio* and *energy*. The former one, which will be matched using Generic Ontology Matching, is generally small, while the latter one is so much that it has to be matched by Large scale Ontology Matching.

There are five groups of test suites in each dataset. Each test suite has 94 matching tasks. The overall results of one test suite will be represented by the mean value of Precision, Recall and F-Measure. Test suites were generated from the same seed ontologies, which means they are all equal. Thus, the harmonic mean values of all test suites will be used to evaluate how well Lily worked.

The detailed results are shown in Table 1.

**Table 1.** The performance in the Benchmark track

Test suite	Precision	Recall	F-Measure
biblio-r1	0.96	0.83	0.89
biblio-r2	0.96	0.83	0.89
biblio-r3	0.97	0.84	0.90
biblio-r4	0.97	0.83	0.89
biblio-r5	0.97	0.84	0.90
<b>H-mean</b>	<b>0.97</b>	<b>0.83</b>	<b>0.90</b>
energy-r1	0.90	0.76	0.82
energy-r2	0.90	0.77	0.83
energy-r3	0.90	0.77	0.83
energy-r4	0.89	0.76	0.82
energy-r5	0.91	0.77	0.83
<b>H-mean</b>	<b>0.90</b>	<b>0.77</b>	<b>0.83</b>

As Table 1 has shown, Lily handles Benchmark datasets well in both small and large scales, although the results of the *energy* dataset are slightly worse as the expense of better performance. According to the Benchmark results of OAEI2015<sup>1</sup>, Lily has the highest overall F-Measure among 11 matching systems that generated alignments for the *Biblio* dataset. However, the public results show that Lily failed to produce alignments for *energy* dataset. That is because the *energy* dataset is a replacement for its former dataset *IFC*. The substitution also brought about format changes of ontology description files. Consequently, Lily and some other systems were not able to parse ontologies correctly. After the issue was fixed, we evaluated Lily on only *energy* dataset with SEALS client and obtained the results.

## 2.2 Anatomy track

The anatomy matching task consists of two real large-scale biological ontologies. Table 2 shows the performance of Lily in the Anatomy track on a server with one 3.46 GHz, 6-core CPU and 8GB RAM allocated. The time unit is second (s).

**Table 2.** The performance in the Anatomy track

Matcher	Runtime	Precision	Recall	F-Measure
Lily	266s	0.87	0.79	0.83

Compared with the result in OAEI 2011 [8], there is a small improvement of Precision, Recall and F-Measure, from 0.80, 0.72 and 0.76 to 0.87, 0.79 and 0.83,

<sup>1</sup> <http://oaei.ontologymatching.org/2015/results/benchmarks/index.html>

respectively. One main reason for the improvement is that we found the names of classes not semantically useful, which would confuse Lily when the similarity matrix was calculated. After the names were excluded, better alignments were generated. Besides, there is a significant reduction of the time consumption, from 563s to 266s. This is not only the result of stronger CPU, but also because more optimizations, like parallelization, were applied to the algorithms in Lily.

However, as can be seen in the overall result, Lily lies in the middle position of the rank, which indicates it is still possible to make further progress. Additionally, some key algorithms have not been successfully parallelized. After that is done, the time consumption is expected to be further reduced.

### 2.3 Conference track

In this track, there are 7 independent ontologies that can be matched with one another. The 21 subtasks are based on given reference alignments. As a result of heterogeneous characters, it is a challenge to generate high-quality alignments for all ontology pairs in this track.

Lily adopted ontology matching tuning for the Conference track this year. Table 3 shows its latest performance.

**Table 3.** The performance in the Conference track

Test Case ID	Precision	Recall	F-Measure
cmt-conference	0.53	0.6	0.56
cmt-confof	0.80	0.25	0.38
cmt-edas	0.64	0.54	0.58
cmt-ekaw	0.55	0.55	0.55
cmt-iasted	0.57	1.00	0.73
cmt-sigkdd	0.70	0.58	0.64
conference-confof	0.67	0.53	0.59
conference-edas	0.41	0.41	0.41
conference-ekaw	0.62	0.64	0.63
conference-iasted	0.67	0.43	0.52
conference-sigkdd	0.71	0.67	0.69
confof-edas	0.69	0.47	0.56
confof-ekaw	0.79	0.75	0.77
confof-iasted	0.46	0.67	0.55
confof-sigkdd	0.17	0.14	0.15
edas-ekaw	0.67	0.52	0.59
edas-iasted	0.50	0.37	0.42
edas-sigkdd	0.63	0.33	0.43
ekaw-iasted	0.50	0.80	0.62
ekaw-sigkdd	0.50	0.46	0.48
iasted-sigkdd	0.56	0.67	0.61
<b>Average</b>	<b>0.59</b>	<b>0.53</b>	<b>0.56</b>

Compared with the result in OAEI 2011 [8], there is a significant improvement of mean Precision, Recall and F-Measure, from 0.36, 0.47 and 0.41 to 0.59, 0.53 and 0.56, respectively. Besides, all the tasks share the same configurations, so it is possible to generate better alignments by assigning the most suitable parameters for each task. We will continue to enhance this feature.

## 2.4 Instance matching track

We submitted alignments for two tasks in the IM track of OAEI 2015: Author Disambiguation Task and Author Recognition Task. For the other three tasks, there is currently no specific strategy available, so Lily will not produce alignments for them.

For each task, there are two matching subtasks with different scales. The *sandbox* scale is around 1,000 instances, which was provided as the test dataset. The *mainbox* scale is around 10,000 instances. The results will be analyzed for each task.

**Author Disambiguation Task** Lily utilized a different strategy for this task, as we found several features of the dataset: one author’s name in ontology A usually contains the corresponding name in ontology B, and a slight difference of one property may distinguish publications in two ontologies. The result is shown in Table 4.

**Table 4.** The performance in the author-dis task

Matcher	Scale	Precision	Recall	F-Measure
Lily	sandbox	0.98	0.98	0.98
Lily	mainbox	0.96	0.96	0.96

As can be seen in Table 4, the strategy is practical. Most correct matches can be found with high precision in both *sandbox* and *mainbox* subtasks. According to overall results, Lily scores highest in this task. However, there are still some missing matches. After analyzing the reference alignments and matching ontologies, we found that some matched authors had actually no publication in common, and that accounts for many matches missed by Lily.

**Author Recognition Task** Quite different from the previous task, this task requires computations over the source ontology, whose results will be matched with the target ontology. Lily will first follow the requirement to generate an intermediate, statistical ontology from the source ontology. Then, string properties and numeric properties of that ontology and the target ontology will be compared in different methods. Finally, all the similarities will be combined. The result is shown in Table 5.

**Table 5.** The performance in the author-rec task

Matcher	Scale	Precision	Recall	F-Measure
Lily	sandbox	1.00	1.00	1.00
Lily	mainbox	0.99	0.99	0.99

As can be seen in Table 5, the strategy is practical as well, especially for the *sandbox* subtask.

### 3 General comments

In this year, a lot of modifications were done to Lily for both effectiveness and efficiency. The performance has been improved as we have expected. The strategies for new tasks have been proved to be useful.

On the whole, Lily is a comprehensive ontology matching system with the ability to handle multiple types of ontology matching tasks, of which the results are generally competitive. However, Lily still lacks in strategies for some newly developed matching tasks. The relatively high time and memory consumption also prevent Lily from finishing some challenging tasks.

### 4 Conclusion

In this paper, we briefly introduced our ontology matching system Lily. The matching process and the special techniques used by Lily were presented, and the alignment results were carefully analyzed.

There is still so much to do to make further progress. Lily needs more optimization to handle large ontologies with limited time and memory. Thus, techniques like parallelization will be applied more. Also, we have just tried out ontology matching tuning. With further research on that, Lily will not only produce better alignments for tracks it was intended for, but also be able to participate in the interactive track.

### 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. 2136 (2008)
- [8] Peng Wang: Lily results on SEALS platform for OAEI 2011. Proc. of 6th OM Workshop, pp. 156-162 (2011)
- [9] Yang, Pan, Peng Wang, Li Ji, Xingyu Chen, Kai Huang, Bin Yu: Ontology Matching Tuning Based on Particle Swarm Optimization: Preliminary Results. In The Semantic Web and Web Science, pp. 146-155 (2014)