Alignment Results of SOBOM for OAEI 2009

Peigang Xu, Haijun Tao, Tianyi Zang, Yadong, Wang

School of Computer Science and Technology

Harbin Institute of Technology, Harbin, China

xpg0312@hotmail.com, hjtao.hit@gmail.com, tianyi.zang@gmail.com, ydwang@hit.edu.cn

Abstract. In this paper we give a brief explanation of how Anchor Concept and Sub-Ontology based Ontology Matching (SOBOM) gets the alignment results at OAEI2009. SOBOM deal with the ontology from two different views: an ontology with is-a hierarchical structure O' and an ontology with other relationships O''. Firstly, from the O' view, SOBOM starts with a set of anchor concepts provided by linguistic matcher. And then it extracts sub-ontologies based on the anchor concepts and ranks these sub-ontologies according to their depth. Secondly, SOBOM utilizes Semantic Inductive Similarity Flooding algorithm to compute the similarity of the concepts between the sub-ontologies derived from the two ontologies according the depth of sub-ontologies to get concept alignments. Finally, from the O'' view, SOBOM gets relationship alignments by using the concept alignment results in O''. The experiment results show SOBOM can find more alignment results than other compared relevant methods with high degree of precision.

1 System presentation

Currently more and more ontologies are distributedly built and used by different organizations. And these ontologies are usually light-weighted [1] containing lots of concepts especially in biomedicine, such as anatomy taxonomy NCI thesaurus. The Anchor Concept and Sub-ontology based Ontology Matching (SOBOM) is designed for matching light-weight ontologies. It handles an ontology from two views: O'

and O'' that are depicted in Fig. 1. The unique feature of our method is combining sub-ontology extraction with ontology matching.

1.1 State, purpose, general statement

SOBOM is an automatic ontology matching tool. There are three matchers implemented in current version: linguistic matcher I-Sub [2], structure matcher SISF (Semantic Inductive Similarity Flooding) which was inspired by Anchor-Prompt [3] and SF [4] algorithms, and relationship matcher R-matcher which utilizes the results of SISF to get relationship alignments. In addition, a Sub-ontology Extractor (SoE) is integrated into SOBOM to extract sub-ontologies according to the result of I-Sub and rank them. The method of SOBOM is fully sequential, so it does not care how to combine the results of different matchers. The overview of the approach is illustrated in Fig. 2.

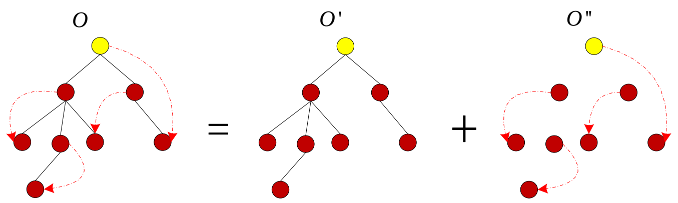

Fig. 1. The construction of ontology in SOBOM

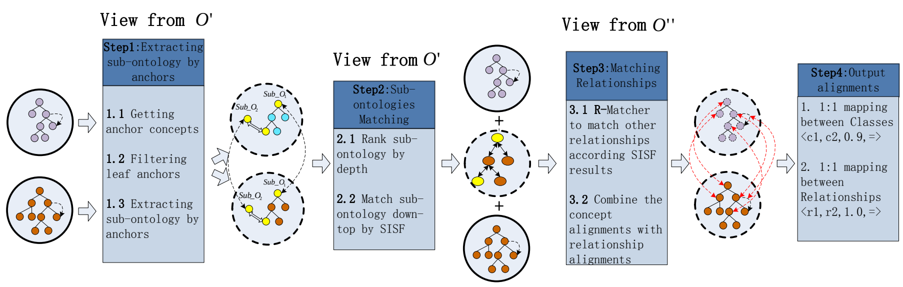

Fig. 2. The process of SOBOM algorithm

For simplicity, we define some notations used in the report.

Ontology: An ontology O consists of a set of concepts C, properties/relations R, instances I, and Axioms A^O. We use entity e to denote either $c \in C$ or $r \in R$. Each relation r has a domain and range defined as following:

$$Domain(r) = \{c_i \mid c_i \in C \ and \ having the relationship \ r\}$$

$$Range(r) = \{c_i \mid c_i \in C \ and \ can be \ value of \ r\}$$

Anchor concept: an anchor concept is the strongest semantic similarity between two entities in different ontologies. It is a pair of concepts from two ontology:

$$a = < c_1, c_2 >, \text{ where } sim(c_1, c_2) \geq \mu$$

Sub-Ontology: Sub_O is a concept hierarchy with $a.c_i$ as root，it satisfied

that $\forall c \in Sub\_O$, $c \in O$ and c is a descendant of $a.c_i$, $Sub\_O \subseteq O$。

1.2 Specific techniques used

SOBOM aims to provide high quality of 1:1 alignments between concept and property pairs. We implemented SOBOM algorithm in java and had integrated three distinguishing constitutional matchers, I-Sub, SISF and R-matcher. They are regarded as independent components in core matcher library of SOBOM. Due to the space limitation, we only describe the key features of them. The details can be found in the related paper [8].

- I-Sub is a light-weighted matcher simply based on the string comparison techniques. The innovation of I-Sub is not only the commonalities between the descriptions of domain entities are calculated but also their differences are examined. Furthermore, it is stable to small diverges from the optimal threshold taking place and intelligent to identify all the differences between strings. In SOBOM, I-Sub is a core component to generate anchor concepts.
- SISF uses the RDF statement to represent the ontology and utilizes the results of I-Sub to inducting the construction of similarity propagation graph from

sub-ontologies. SISF and I-Sub handle the ontology from the view O' and only generate concept-concept alignment.

- R-matcher is a relationship matcher base on the definition of the ontology. It combines the linguistic and semantic information of a relation. From the O'' view, it utilizes the is-a hierarchy to extend the domain and range of a relationship and uses the result of SISF to generate the alignment between relationships.

More importantly, SoE is integrated into the SOBOM and extract sub-ontologies according to the anchor concept [5, 6]. SoE ranks extracted sub-ontologies from the O' view according to their depth. As for ontology matching, the rules of extracting sub-ontology in SoE are as following:

Rule 1: Upwards traversal of the hierarchy: $\forall c' \in O$, if c' is an ancestor of $a.c_i$, then $c' \notin Sub\_O$.

Rule 2: Siblings classes of anchor concepts: $\forall c' \in O$, if c' is a sibling concept of $a.c_i$, then $c' \notin Sub\_O$.

Rule 3: Downwards traversal of the hierarchy: $\forall c' \in O$, if c' is descendant concepts of $a.c_i$, $c' \in Sub\_O$

Rule 4: Other relationships of the anchor concepts: $\forall r \in O$, if r is a relationship in O and $r \neq is\_a$, then $r \notin Sub\_O$

Rule 5: Leaf Concept Nodes: if $a.c_1 \in O_1$, $a.c_2 \in O_2$ and $a.c_1$, $a.c_2$ are leaf nodes respectively in O_1, O_2, then don't extract Sub-Ontology.

After extracting sub-ontologies, SOBOM will match these sub-ontologies according to their depth in original ontology. We first match the sub-ontologies with larger depth value. By using SoE, SOBOM can reduce the scale of ontology and make it easy to operate sub-ontologies in SISF.

1.3 Adaptations made for the evaluation

We don't make any specific adaptation for the tests in the OAEI 2009 campaign. All the alignments outputted by SOBOM are based on the same set of parameters.

1.4 Link to the system and set of provided alignments (in align format)

The current version of SOBOM and the alignment results for OAEI 2009 are available at http://mlg.hit.edu.cn:8080/Ontology/Download.jsp, and the parameters setting is illustrated in the reading me file.

2 Results

In this section, we describe the results of SOBOM algorithm against the benchmark, directory and anatomy ontologies provided by the OAEI 2009 campaign. We use Jena-API to parse the RDF and OWL files. The experiments were carried out on a PC running Windows vista ultimate (32 bit) with Core 2 Duo processors (2.66 GHz) and 4-gigabyte memory.

2.1 Benchmark

On the basis of the nature, we can divide the benchmark dataset into five groups: #101-104, #201-210, #221-247, #248-266 and #301-304. We described the performance of our SOBOM algorithm over each group and overall performance on the benchmark test set in Table 1.

#101-104 SOBOM plays well for these test cases.

#201-210 In this group, some linguistic features of candidate ontologies are discarded or modified. SOBOM is a sequential matcher, if the linguistic matcher get no mappings, then the SISF will produce no mapping too. So in these test, the result is in high precision but low recall.

#221-247 The structures of the candidate ontologies are altered in these tests. However, SOBOM discovers most of the alignments from the linguistic perspective via our linguistic matcher, and both the precision and recall are pretty good.

#248-266 Both the linguistic and structural characteristics of the candidate ontologies are changed heavily. In most cases, SOBOM can get high precision but low recall.

#301-304 This test group are four real-life ontologies of bibliographic references. SOBOM can only find equivalence alignment relations.

Table 1. The performance on the benchmark

	101-104	201-210	221-247	248-266	301-304	Average	H-mean
Precision	0.98	0.99	0.99	1.0	0.86	0.96	0.98
Recall	0.97	0.48	0.95	0.43	0.52	0.67	0.43

2.2 Anatomy

The anatomy real world test bed covers the domain of body anatomy and consist of two ontologies, Adult Mouse Anatomy (2247 classes) and NCI Thesaurus (3304 classes). This type ontologies is what SOBOM suitable for. The experiment result shows in Table 2.

Table 2. The performance of SOBOM on the anatomy test

<table>
<tr><th></th><th>Anchor-concept</th><th>Sub-ontologies</th><th>Alignments</th><th>Time consuming</th></tr>
<tr><td>NCI</td><td rowspan="2">1233</td><td rowspan="2">268</td><td rowspan="2">1249</td><td rowspan="2">19min3s</td></tr>
<tr><td>MA</td></tr>
</table>

2.3 Directory

The directory track requires matching two taxonomies describing the web directories. It includes 4639 matching tasks represented by pairs of OWL ontologies, where classification relations are modeled as *rdfs:subClassOf* relations. But in the experiments, we found there are some ontologies have wrong structure, they have a loop such as 1603, 1704, 2114, 2184, 2241, 2252, 2416, 3045, 3135, 3166, 3183, 3301,3398, 3440, 3556, 3653, 3695, 3711, 4075, 4129, 4544, 851, 118, 148, 1550,1723, 1863, 1967,2, 2000, 2103, 2270, 2271, 2632, 2749, 2803, 3058, 3186,3310, 3455, 3461, 3891, 4048, 4089, 4116, 4341, 4556, 614, 726, 747, totally 50 ontologies. So SOBOM cannot deal with these tests. The experiment results shows in Table 3.

Table 3 The performance of on directory test

Precision	Recall	F-measure
0.5931	0.4145	0.4879

3 General comments

3.1 Comments on the results

Strengths SOBOM deals with ontology from two different views and combines results of every step in sequential way. If the ontologies have regular literals and hierarchical structures, SOBOM can achieve satisfactory alignments. And it can avoid missing alignment in many block matching methods [7].

Weaknesses SOBOM needs the anchor concepts to extract sub-ontologies. So it heavily depends on the anchor concepts. if the literals of concept missed, SOBOM will get bad results.

3.2 Discussions on the way to improve the proposed system

SOBOM can be viewed as a frame of ontology matching. So many independent matchers can be integrated into it. Now anchor concepts generator is a weak matcher, our next plan is to integrate a more powerful matcher to produce anchor concepts or develop a new method to get anchor concepts.

4 Conclusion

This paper reports our first participation in OAEI campaign. We present the alignment process of SOBOM and describe the specific techniques for ontology matching. We also show the performance in different alignment tasks. The strengths and the weaknesses of our proposed approach are summarized and the possible improvement will be made for the system in the future. We propose a brand new algorithm to match ontologies.

References

1. Fausto Giunchiglia and Ilya Aihrayeu : Lightweight Ontologies. Technical Report. (2007) DIT-07-071.
2. G.Stoilos, G. Stamou, S. Kollias: A string metric for ontology alignment, In Proc. Of the 4th International Semantic Web Conference(ISWC'05). (2005) 623-637.
3. N.F. Noy, M.A. Musen: Anchor-PROMPT: using non-local context for semantic matching, In Proc. Of IJCAI2001 Workshop on Ontology and Information Sharing, (2001) 63-70.
4. S. Melnik, H.G. Molina and E. Rahm: Similarity Flooding: A Versatile Graph Matching Algorithm, In Proc 18th Int'l Conf. Data Eng. (ECDE'02) (2002) 117-128.
5. Julian Seidenberg and Alan Rector: Web Ontology Segmentation: Analysis, Classification and Use, WWW2006, (2006).
6. H. Stuckenschmidt and M. Klein: Structure-Based Partitioning of Large Class Hierarchies. In Proc of the 3rd International Semantic Web Conference (2004).
7. W. Hu, Y. Qu: Block matching for ontologies, In Proc of the 5th International Semantic Web Conference, LNCS, vol. 4273, Springer (2006) 300-313.
8. P.G. Xu, H.J. Tao: SOBOM: An Anchor Concept and Sub-ontology based Ontology Matching Approach. To be appear.