Towards a Flexible Schema Matching Approach for Semantic Web Service Discovery

Sana Sellami, Omar Boucelma
LSIS, UMR CNRS 7296
Aix-Marseille University
Marseilles, France
Email: sana.sellami@univ-amu.fr, omar.boucelma@univ-amu.fr

***Abstract*—Semantic web service discovery has attracted a lot of attention in the last decade. Research conducted in this area can be (mainly) summarized as follows: (1) "monolith" matchmaking algorithms (and systems), and (2) schema matching-based techniques. In this paper we describe a flexible approach that takes leverage of existing schema matchers, leading to a multiple choice strategy for semantic service discovery. The approach has been implemented and validated in using the data collection provided by the S3 (Semantic Service Selection) community, and led to promising preliminary results.**

I. Introduction

Semantic web services discovery (SWSD) has gained momentum during the last decade, and has led to the proposal of several semantic matchmaking techniques and systems [1]. Despite commonalities shared with the schema matching research area [1], those techniques have been proposed along a separate and independent pathway. Nevertheless, during the past few years, some schema matching approaches have also been proposed to address WS discovery [2].

In this paper, we propose an "unbundled" approach [2] (versus a monolith one) for SWSD. The approach relies on an extensible set of schema/ontology matching techniques and primitives that may be combined in various ways, accordingly with the specification of a service request. Matching approaches depend on the description of services and can be applied on element or structure levels; they may be terminological, structural, or semantic [3].

II. Flexible Semantic Web Service Discovery

Our approach takes leverage of existing schema matchers, leading to a multiple choice strategy for SWS matchmaking. These matchers can be: terminological, structural and semantic. In the sequel, we describe the different matchers that can be applied for the discovery of web services annotated with SAWSDL [3].

A. Schema and Ontology Matchers

Each SAWSDL interface exposes a set of operations (input, output). For each pair of operations, we compare the input/output and their composed elements (defined in the XSD sequence types) together with SAWSDL annotations that are considered as descriptions of those elements. If there is no correspondence, SAWSDL semantic annotations are considered as separate ontological concepts. We use terminological, structural and semantic-based matchers to find the similarities between those concepts . Based on the classification proposed in [3], we select and combine these matchers according to the nodes that need to be matched. We have identified different matching use cases together with suitable matchers summarized below.

1) Element or Concept-based Matchers: These matchers consists of: (a) string-based techniques, which are used in order to match names and name descriptions of schema/ontology elements; (b) language-based techniques such as tokenization, lemmatization and elimination, (c) constraint-based techniques that deal with the internal constraints of entities definition(e.g., types, cardinality of attributes, and keys), (d) Information retrieval techniques (e.g., for keyword extraction), (e) Linguistic resources such as common knowledge, dictionary or thesauri.

2) Structural-based Matchers: These are graph algorithms which take as input labeled graphs to perform structural similarity. This similarity relies on the context elements and on the hypothesis that two elements are structurally similar if theirs structural neighbours are similar. The context includes: ancestor, sibling, immediate child and leaf nodes.

3) Semantic-based Matchers: Specify a similarity function in the form of a semantic relation (e.g., hyperonym) between concepts. Semantic matching may invoke external information found in lexicons, thesauruses or reference ontologies, incorporating semantic knowledge (mostly domain-dependent) into the process.

B. Matching Strategy

We adopted a hybrid approach in combining syntactic and semantic matching . The idea is to combine the strengths of these different matchers. The combination of the results of individual matchers often improves the matching result quality [4]. Based on SAWSDL descriptions, the matching process consists of three steps:

- Step1: for all elements in the services and request, compute the terminological, constraint-based and structural matching between input/output's labels.

[1] http://www-ags.dfki.uni-sb.de/~klusch/s3/index.html
[2] The term is coined with unbundling database services
[3] http://www.w3.org/TR/sawsdl-guide/

- Step 2: for all ontological concepts referenced in the ontology, we compute constraint-based matching and semantic relations between concepts using structure level matching and/or semantic based matching.
- Step3: we combine the different matcher similarity results [4].

C. Experimentation

We picked 404 services from the SAWSDL-TC1 collection with annotations referring a single ontology and 12 request services from different domains: communication, food, economy, medical, travel and education. Each request returned a relevant set of services.

1) Matcher's Description: We did run different matchers against the SAWSDL-TC1 collection. We compared schema matching techniques through Hybrid context matcher against matchmakers' techniques (logic, textual and hybrid). The configuration of these matchers is described as follows:

- Hybrid Context matcher: combines 2 element levels matching (name and nametype) and 3 structural matchers (namepath, children and leaves) (threshold=0.5). We have selected this configuration because it offers the best results in terms of precision and recall on XML schema matching [5]. We choose combination strategy that proved to be the aggregation method as it could best compensate the shortcomings of individual matchers.
- Logic-based matchmaker: the definition of filters is the same as in SAWSDL-MX1 [6]. For logic-based matching we select subsumed-by filter.
- Textual-based matchmaker is a non logic one that uses Information Retrieval techniques. We did choose Extended Jaccard similarity (threshold=0.5)
- Hybrid (cos)-based matchmaker combines logic and non logic based matchers. We selected cosine measure, with a threshold=0.5.

2) Evaluation results: We used precision and recall measures to evaluate the effectiveness of our approach. We compute average precision (AP) [7] and macro-average of precision (MAP) [8]. AP is the average of precisions computed at each point of the relevant documents in the returned ranked list. MAP corresponds to the precision values for answer sets returned by the matchmaker for all queries in the test collection at standard recall levels. As the evaluation of a single query is not sufficient to make a significant observation, MAP is computed over many queries giving equal weight to each user query. Ceiling interpolation is used to estimate precision values at each standardized recall level, since each query likely has a different number of relevant services.

We compared our Hybrid Context matcher vs logic, textual and hybrid configurations. We tuned them in order to execute different strategies. Table I and figure 1 show the average precision of our hybrid context matcher, logic based matcher, hybrid (cosine) and IR based on Jaccard coefficient, and the macro-averaged precision, respectively.

	Textual	Hybrid	Logic	Hybrid Context
average.AP	0.56	0.6	0.43	0.72

TABLE I
AVERAGED AP VALUES

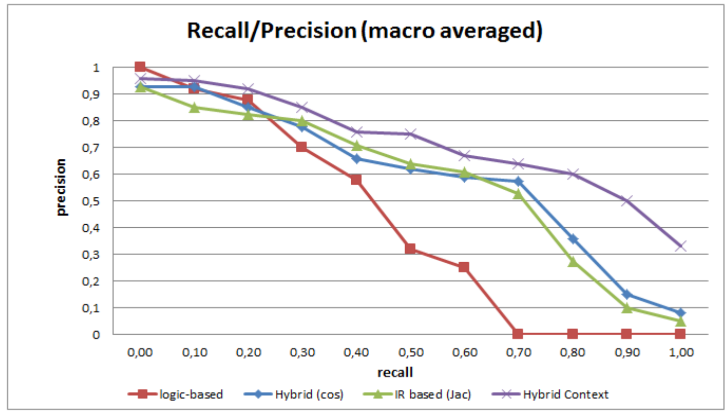

Fig. 1. Recall/Precision Macro Averaged

III. CONCLUSION

In this paper, we described an approach based on schema matching techniques for SWS discovery. The contribution of the paper is twofold: (1) an "unbundled" matchmaking approach based on schema matching, and (2) an experimental evaluation performed by using the well known SAWSDL-TC1 service retrieval test collection [4]. While the experimental results prove satisfactory, we believe that there is room for assessing the quality of the set of services returned by any SWSD method.

REFERENCES

[1] Bellahsene, Z., Bonifati, A., Rahm, E., eds.: Schema Matching and Mapping. Springer (2011)
[2] Algergawy, A., Nayak, R., Siegmund, N., Köppen, V., Saake, G.: Combining schema and level-based matching for web service discovery. In: ICWE. (2010) 114–128
[3] Shvaiko, P., Euzenat, J.: A survey of schema-based matching approaches. Journal on Data Semantics IV **3730** (2005) 146–171
[4] Peukert, E., Maßmann, S., König, K.: Comparing similarity combination methods for schema matching. In Fähnrich, K.P., Franczyk, B., eds.: Informatik 2010: Service Science - Neue Perspektiven für die Informatik, Beiträge der 40. Jahrestagung der Gesellschaft für Informatik e.V. (GI). Volume 175., GI (2010) 692–701
[5] Do, H.H., Melnik, S., Rahm, E.: Comparison of schema matching evaluations. In: Web, Web-Services, and Database Systems, NODe 2002 Web and Database-Related Workshops, Erfurt, Germany, October 7-10, 2002, Revised Papers. Volume 2593 of Lecture Notes in Computer Science., Springer (2003) 221–237
[6] Klusch, M., Kapahnke, P., Zinnikus, I.: Adaptive hybrid semantic selection of sawsdl services with sawsdl-mx2. Int. J. Semantic Web Inf. Syst. (2010)
[7] Raghavan, V.V., Bollmann, P., Jung, G.S.: Retrieval system evaluation using recall and precision: Problems and answers. In: SIGIR'89, 12th International Conference on Research and Development in Information Retrieval, Cambridge, Massachusetts, USA, June 25-28, 1989, Proceedings. (1989) 59–68
[8] Küster, U.: An evaluation methodology and framework for semantic web services technology. PhD thesis (2011)

[4] Available at http://www.semwebcentral.org/