OntoMas: a Tutoring System dedicated to Ontology Matching

Mirella Huza[1], Mounira Harzallah[1], Francky Trichet[1]

[1]LINA – University of Nantes
Laboratoire d'Informatique de Nantes Atlantique (FRE-CNRS 2729)
Team « Knowledge and Decision »
2, rue de la houssinière – BP 92208 - 44322 Nantes, France
{mirella.huza,mounira.harzallah,francky.trichet}@univ-nantes.fr

Abstract. Ontology matching is now a core question in most of the applications that require semantic interoperability. To deal with this problem, a lot of methods, classified according to different criteria, are currently developed. However, choosing the most relevant method in a particular context is not an easy task since it requires to know all the methods and their intrinsic properties. The objective of the OntoMas[1] tutoring system (*Ontology Matching Assistant*) is (1) to propose an architecture and to develop an effective knowledge-based system dedicated to a fine-grained description and a classification of the current matching methods and (2) to provide functionalities dedicated to the definition of advices and explanations (for the end-user), in order to facilitate both the choice of the most suitable method for a particular matching problem and the learning of this new domain: ontology matching.

Keywords: Ontology matching, Classification of ontology matching method, Tutoring System, Ontology matching assistant.

1. Introduction

Ontology matching is now a core question in most of the applications that require semantic interoperability such as the Semantic Web. To deal with this problem, which mainly consists in finding semantic links (*e.g.* equivalence, disjointness or subsumption) between the concepts and the relations of two ontologies covering overlapping domains, a lot of methods are currently proposed [1,2,3,4].

The choice of a method or the combination of several ones is not an easy task since (1) it depends on multiple criteria related both (i) to the characteristics of the current matching methods, (ii) to the ontologies that are considered for the matching process (in terms of type, structure, representation format, etc.) and (iii) to the end-user preferences and (2) it requires to know all these different criteria in order to select the best method for the considered matching context. Moreover, as the

[1] This work is supported by the Commission of the European Communities under the Sixth Framework Program: [INTEROP Network Of Excellence – 508011] http://www.interop-noe.org.

Ontology Matching domain is currently booming, it is really difficult to keep an up-to-date overview of the current methods.

This paper introduces the basic foundations of a Tutoring System dedicated to Ontology Matching. This system, called OntoMas[2] (*Ontology Matching Assistant*), aims at facilitating the classification of ontology matching methods and at helping the end-user to select the most relevant method(s) according to his matching problem. It also aims at providing to the end-user relevant advices and explanations on the method(s) that can be used (or can definitively not be used) in his matching context. Thus, OntoMas is not only a tool to select the best matching technique; it is also a tool dedicated to ontology matching learning.

2. OntoMas matching method classification

In most of the current classifications, only the criteria which directly relate to the description of the methods are considered. However, the choice of a matching method also depends on the end-user matching context. Since our objective is to propose a new, complete, up-to-date and easy to use classification of both the current matching methods and the possible matching contexts, we propose to differentiate two categories of classifications (and thus two groups of criteria): a classification based on the characteristics of the matching methods and a classification based on the characteristics of the matching context.

2.1 Matching method-based criteria

For the matching method-based criteria, we reuse the three dimensions of [4] and enrich them with new criteria.

For the *Input dimension*, the criterion **Interlingual matching** is added: is the method able to perform interlingual matching between ontologies, *e.g.* matching an English and a French ontology ? We add also a sub-dimension including the following criteria: (i) **Internal representation** (the method can translate the input models into an internal representation of ontology specification language), (ii) **String normalization on input data** (the method can perform string normalization on the input ontologies, *e.g.* case normalization, diacritics suppression, etc.).

For the *Processing dimension*, three criteria are introduced. The first one is about **User interaction**. The user interaction can be used in different matching phases: preprocessing, configuration of parameters, definition of an initial set of matches; modification of the architecture of the combined matchers and validation of results [2]. The second one is the **Iteration** criterion. The method can offer the possibility of performing iterations in the matching process, *e.g.* the matching algorithm can be launched several times, in this way the quality of the results can rise considerably. The third criterion, called **Alignment extraction**, refers to the process of choosing from all the obtained matching a satisfactory set of correspondences between

[2] The current prototype of OntoMas is available at the following URL: http://www.polytech.univ-nantes.fr/ontomas/

ontologies. It can be achieved by: (i) displaying the entity pairs with their similarity scores and leaving the choice of the appropriate pairs up to the user, (ii) threshold-based filtering, selecting correspondences over a particular threshold, thus retaining only the most similar entity pairs [2].

For the *Output dimension*, we add three others criteria introduced in [2]: (i) **Matches format** (the format in which the matching is represented - it can be: matching pairs, OWL, XML, ASCII, SKOS, text or other format); (ii) **Complete matching** (the method can produce complete matching for one of the ontologies); (iii) **Injective matching** (the method can produce injective, and reversible matching); (iv) **Integration** (using the results of ontology matching for ontology merging).

Finally a G*eneral dimension* concerning the tool description is introduced. Its relevant criteria are: (i) Type of solution (method or tool). When the solution is a tool: (ii) **Programming language** (it was implemented using a programming language); (iii) **Free version** (it can have a free/not free version); (iv) **Installation requirements** (it may have installation requirements: need installation of another software).

All these criteria are method-oriented. Thus, they are more dedicated to the ontology matching engineer point of view.

2.2 Matching context based criteria

The second group of criteria are dedicated to the description of the matching context: (1) *Input ontologies criteria* , (2) *User desired output criteria* and (3) *Method desired characteristics criteria*. These criteria are also related to the matching process, but from the end-user point of view.

The *Input ontologies criteria* are dedicated to the description of the content of the ontologies that are considered: (i) Format; (ii) Language (the natural language used in the representation of ontology); (iii) Content; (iv) **Ontology Domain** (*e.*g. Human Anatomy, Computer Science); (v) **Ontology size** (ontology is very large or not, A very large ontology is an ontology that has more than 1000 concepts); (vi) **Ontology structure** (it concerns the unbalance between the ontologies, one is a deep ontology and the other one is a shallow ontology), (vii) **Ontology version** (the ontologies are different versions of the same ontology), (viii) **Ontology similarity** (the domains covering by the ontologies are very different – they are not strong overlapping domains), (ix) **Ontology depth**, (x) **Ontology Relations** (*e.*g. is-a, part-of). The three first criteria are already introduced in the classifications presented in Section 2.

For the criteria related to the *User desired output*, we propose to consider: (i) Matches cardinality, (ii) Matches format, (iii) Matches Relations, (iv) Matches Ranking, (v) Integration.

The *Method desired characteristics criteria* are: (i) Type of solution (method or tool), (ii) Programming language, (iii) Free version, (iv) Iteration and (v) User Interaction.

3. OntoMas Tutoring System

One of the functionalities of OntoMas is the selection of the most suitable matching methods (for matching two ontologies given by the end-user) from the ones integrated in the knowledge base which currently includes 3 methods (IF-Map - Information Flow based ontology mapping, HCONE-merge and AROMA - Association Rule Ontology Mapping Approach) and 5 tools (CMS - CROSI Mapping System, AUTOMS, TooCoM, Organon and H-MATCH). The principle underlying this selection consists in using a set of **decision rules** defined between (1) methods described in OntoMas knowledge base and (2) matching contexts.

A **decision rule** represents a relation between a criterion related to the method point of view and a criterion related to the context point of view. The selection process is based on all the decision rules of the OntoMas system.

Each characteristic of the matching context, after being described by the end-user, determines a "restriction" imposed on the methods that can be used in the matching context. Each characteristic of the matching context implies a decision rule. Thus, we use 20 rules: 10 for the *Input ontologies criteria*, 5 for the *Method desired characteristics criteria* and 5 for the *User desired output criteria*.

Three types of decision rules are distinguished: compulsory rules, not compulsory rules and optional rules.

(C) Compulsory rules. These rules are related to the criteria used to describe the matching context for which the end-user has to provide values because they correspond to the minimum amount of information which is necessary to perform the decision process.

Example of C: Format

If	Ontology 1 is represented in the format F1 And Ontology2 is represented in The format F2 And F1 and F2 are the same or compatible
Then	SELECT from the knowledge base the methods that can use as input models ontologies that are represented in the F1(F2) format.

(NC) Not compulsory rules. They correspond to the criteria for which the end-user is not obliged to provide (input) information. Example of such a rule: Size.

Example of NC: Size

If	Ontology1 is a very large ontology Or Ontology2 is a very large ontology
Then	Increase with one point the total score of the methods that produce relevant results when the input ontologies are very large.

(O) Optional rules. They allow the end-user to precise preferences ("Only find methods that…" OR "I prefer methods… that"). When the end-user selects the "Only find" option, then the rule becomes compulsory and when the end-user selects the "I prefer" option, the rule becomes not compulsory.

Exemple of O: User interaction

Compulsory

If	The end-user wants to CHOOSE ONLY methods that use (don't use) Interaction
Then	Choose from the knowledge base the methods that use (don't use) in their matching phases user interaction.

Not compulsory

If The end-user wants OntoMas to PREFER methods that use (don't use) Interaction

Then Increase with one point the total score of the methods that use (don't use) in their matching phases user interaction.

After applying all the compulsory rules to the methods of the knowledge base, a set of methods that satisfy the compulsory requirements of the matching context is obtained: setC. Then, each rule of the not compulsory rules is applied to each method m_i of setC. If m_i satisfies the rule, one adds to its importance score one point. If m_i does not satisfy the rule, one decreases its importance score with one point. After applying all the not compulsory rules to the setC, an importance score is obtained for each method from the setC. It is then possible to order the set (more higher the importance score is, more suitable the method is).

5 Conclusion

In this paper, we have underlined the fact that choosing the best method for a particular ontology matching problem is not an easy task since the number of methods is currently booming and the criteria used in the current classifications do not enable the description of the matching context (they are mainly method-oriented). To deal with this problem, we propose to introduce new criteria dedicated to the end-user point of view: (i) description of his ontologies (domain, size, structure, version, similarity, depth, relations) and (ii) description of his preferences in terms of results (matches cardinality, format, relations, ranking and integration) and tools (programming language, free version, iteration and user interaction). Based on these new criteria and on the basic method-oriented criteria, we have defined compulsory, not compulsory and optional decision rules. These rules are used by OntoMas in order to recommend (and order) a set suitable (resp. unsuitable) methods for a specific ontology matching problem.

References

1. A. Doan and A. Halevy. Semantic integration research in the database community: A brief survey. 2005.
2. J. Euzenat, J. Barrasa, P. Bouquet, R. Dieng, M. Ehrig, M. Hauswirth, M. Jarrar, R. Lara,D. Maynard, A. Napoli, G. Stamou, H. Stuckenschmidt, P. Shvaiko, S. Tessaris, S. van Acker, I. Zaihrayeu, and T. L. Bach. D2.2.3: State of the art on ontology alignment. Technical report, NoE Knowledge Web project deliverable, 2004.
3. N. F. Noy. Semantic integration: A survey of ontology-based approaches. SIGMOD Record, 33(4):65–70, 2004.
4. P. Shvaiko and J. Euzenat. A survey of schema-based matching approaches. 3730:146–171, 2005.