Description and Evaluation of Algorithms for Ontology Matching

Mario Blanco-Alonso, Francisco J. Rodríguez-Martínez and Lorena Otero-Cerdeira
LIA2 Group - Computer Science Department, University of Vigo, Ourense, Spain

Keywords: Didactic Framework, MaasMatch, CODI, LogMap, Ontology, Matching.

Abstract: In this paper, we present a state of the art about Ontology Matching Algorithms and we propose a general classification of them. A selection of three algorithms to work with a concrete platform is presented: *CODI*, *LogMap* and *MaasMatch*.
In addition we propose a testbed divided in three groups of tests to evaluate the algorithms .
These algorithms were tested and evaluated to verify which was the most suitable for this problem.

1 INTRODUCTION

The existence of computer systems that handle information and entities related to the real world has caused a growing interest in techniques and processes that can build connections among them. This shows the need to identify and link the heterogeneous information present in multiple resources. Since this requirement was detected, several proposals appeared to solve and facilitate, especially, the treatment of the ambiguity on these relationships. This paper focuses on semantic techniques since, in Ontology Matching they are the most prominent among the available ones.

An ontology defines a vocabulary that describes a domain of interest and a specification of the meaning of terms used in the vocabulary (David et al., 2010). Depending on both, the accuracy of the specification and, its purpose and scope, the notion of ontology ranges from, groups of terms and classifications, to database schemas (Euzenat and Shvaiko, 2007).

There are different environments where Ontologies and Ontology Matching(alignment of terms) can be used. Based on these ontologies, various systems arose to align in such way that it was possible to generate relations in comparable domains.

We propose a specific case to apply Ontology Matching, both in a theoretical and practical way, related to a didactic platform. This platform has two main types of users, "Teacher" and "Student", who create and use contents like: templates, tables, models, exercises, etc.

This platform is used in several schools that generate sets of terms that could refer to the same meaning but with different names. Using the right tools its possible to create labels to tag and identify contents, users and units. Thus, to assume that system is mapped and to make it easier to create groups under some rules and conditions, we need to manually tag the specific domain. In such way the system can model the knowledge with a set of primitives called Ontologies.

In the context of this platform, the goal is to map the semantic information to use it as a base to create relations between the Ontologies that define the concepts. Therefore its possible to get automatic interconnections that are discernible similarity relations. For example, a teacher may create new content and, to identify it, tag it using the label "subject" while in another school, another teacher could define new enhanced content about the same topic, but adding a label "course" rather than "subject". Those two concepts are related and, therefore, they should belong to the same group. The alignment "subject-course" does exactly that, it classifies a concept that should be grouped using a relationship in an alignment.

This is why it seems mandatory that once the platform is created with its own ontological representation, it will be possible to interconnect the knowledge, even when its allocated in different geographic locations and managed by different people. As a restriction to the algorithm that would align the ontologies, it must also align the knowledge consistently keeping a great degree of coherence.

With regard to Ontology Matching, it has achieved a satisfactory level of success but still has challenges at the operational and computational level which lead

Blanco-Alonso M., Rodríguez-Martínez F. and Otero-Cerdeira L..
Description and Evaluation of Algorithms for Ontology Matching.
DOI: 10.5220/0004253304920495
In *Proceedings of the 5th International Conference on Agents and Artificial Intelligence* (ICAART-2013), pages 492-495
ISBN: 978-989-8565-39-6

to a very open research line.

In recent decades diverse solutions have been proposed for Ontology Matching (Batini and M.Lenzerini, 1986; Spaccapietra et al., 1992), several recent surveys (Bellahsene et al., 2011; Shvaiko and Euzenat, 2008; Sakarkar and Upadhye, 2010) and some books have also been written (Euzenat and Shvaiko, 2007; Bellahsene et al., 2011) on the subject.

The rest of the paper is organized as follows.In Section 2 the basis of the platform are specified, in section 3 a brief review of the best existing algorithms based on Ontology Matching. Section 4 shows the benchmarks and tests. And finally In section 5 there is a conclusion about the general results and the behavior of each algorithm.

2 PLATFORM BASIS

The main focus of this didactic platform IntellTec5.1 is to create a virtual environment for interactive e-learning, promoting the development of a creative, autonomous and independent pedagogy.
We are not limited to upload traditional materials to the cloud, in fact we provide:

- A domain to elaborate learning materials
- Support tools
- Data and knowledge sources
- Self-assessment activities
- Collaborative learning

To introduce and explain the platform mechanisms, we have to distinguish between two potential users: Teachers and Students. Teachers can manage activities and the progress of every student. Therefore with this specialized training and teaching it is easy to empower the students' motivation to become specialized professionals taking advantage of the new technologies and improving their own skills to draw upon the e-learning capabilities and internet potential.

About the built-in management system that each user has in IntellTec5.1:

- **Teacher.** This kind of user has a set of tools to develop its course topics, units, agenda and the on-line activities to improve its new acquired skills. They can evaluate students on-line, check their specific and over-all progress, take care of individuals, etc.
  Each teacher can create and upload its own content, activities, exercises using the integrated tool to create the content using templates, completely editable and configurable. For those teachers who don't want to create new content, we have a database with a recompilation of every template and activity skeleton created by any teacher in the world. The focus of this approach is to share and grow the knowledge and good ideas, so any student in any school is able to take advantage of the best activities and applications.
  Then a teacher can manage students, create digital contents and new interactive, traditional and auto-evaluable tasks.

- **Student.** The whole system has the student as the main target to improve the e-learning experience and focusing on the motivation and inspiration to create and use the digital advantages and the internet potential.
  We encourage the exploration of the data kept in our knowledge base with the interactive activities approach making a great enhancement to self-taught learning and provoking a profile of student who is eager to explore by himself.
  The platform lets the student enjoy a collaborative domain, the use of an easy but complete set of tools to develop the digital skills to create, use and explore digital content and internet by himself.

The focus of this paper is to study a possible solution to some side problems like knowledge representation, knowledge interaction and grouping, suggestion functionalities, intelligent sharing and an easy way to show and use a vast source of specialized information.

3 ALGORITHMS THAT GENERATE ALIGNMENTS

Regarding the measure effectiveness of an Ontology Matching algorithm, there are several ways to select which one is better, depending on the type of tests, data and complexity used to compare them. It will change which algorithm is the better in every case. There are evaluators and generic tests to evaluate and compare algorithms and determine which one is the best in each field or goal.
CODI(Huber et al., 2011), LogMap(Jiménez-Ruiz et al., 2011) and MaasMatch(Schadd and Roos, 2011) are algorithms that were designed to solve problems similar to the didactic platform proposed in this research paper. They are specialized to solve alignments prioritizing the logic and coherence in the resulting ontologies.

3.1 CODI

Combinatorial Optimization for Data Integration (CODI) leverages terminological structure for Ontology matching. The current implementation produces mappings between concepts, properties, and individuals. The system combines lexical similarity measures with schema information to completely avoid incoherence and inconsistency during the alignment process(Huber et al., 2011).
CODI is based on instance classification, so it is grouped as an instance-based technique with the addition of the lexical similarity measures.

3.2 LogMap

LogMap was developed to be capable of computing scalable and logic-based Ontologies. It's able to take advantage of diagnosis techniques, make alignments and output mappings that don't lead to logical inconsistencies when integrated with the input ontologies. Therefore, the resulting alignments don't have incoherences. The second focus of this algorithm is to achieve a low runtime.

3.3 MaasMatch

MaasMatch is an Ontology Matching tool that focuses on resolving terminological heterogeneities, such that entities with the same meaning but differing names and entities with the same name but different meanings are identified as such and matched accordingly.

4 BENCHMARK TEST

4.1 Benchmark Tracks

The goal of the benchmark data set is to provide a stable and detailed picture of each algorithm. For that purpose, algorithms are run on systematically generated test cases.

4.1.1 Bonus Tests

- **Anatomy.** This test confronts the existing matching technology with a specific type of ontologies from the biomedical domain. In this domain, many ontologies have been build covering different aspects of medical research. We focus on fragments of two biomedical ontologies which describe the human anatomy and the anatomy of the mouse(Shvaiko et al., 2011).

Table 1: Anatomy results.

Anatomy	CODI	LogMap	MaasMatch
Precision	0.965	0.948	0.995
Recall	0.825	0.846	0.287
F_1 score	0.879	0.894	0.445

- **Conference.** The main features are that the ontologies were developed independently and based on different resources using different terminologies and points of view, and most ontologies were equipped with OWL DL axioms of various kinds opening a way to use semantic matchers. In addition, we tested the runtime of the algorithms(Shvaiko et al., 2011).

5 CONCLUSIONS AND FUTURE WORK

In this paper, an evaluation, classification and selection of Ontology Matching algorithms is proposed. The main pre-condition to select the algorithms for this study is that they must be able to generate coherent alignments (without disjointness), which is only achieved by three of the studied algorithms(over 40). This work was made to distinguishing relevant algorithms applicable to a real project with some restrictions and real life issues. The proposed algorithms to face the problem were *CODI, LogMap* and *MaasMatch*.

The experimental results of this paper show that there are algorithms that yield great success with the proposed restrictions and seem to be promising. Based on the results, CODI seems like it is the best one of them with a very good F-measure, coherent alignments and capable of aligning large ontologies.

LogMap could be selected if it didn't yield incoherent alignments since its runtime is the best. Testing and using these algorithms gave a good expertise of what is done and what should be enhanced.

MaasMatch is an average algorithm: it has better runtime than CODI, but worse than LogMap. It has a good F-measure, but worse than CODI and LogMap. It can handle large ontologies as the other two. MaasMatch has more incoherences than CODI and LogMap, therefore this algorithm is excluded since it was outperformed in every test.

Based on the research made and results evaluated, we have concluded that we have to use LogMap approach and CODI approach to implement three ways to handle the issues detected in the didactic platform. First, we will use LogMap as a real-time suggestion tool for the teacher users, then they can type down or

Table 2: Conference results.

Conference	Prec	Rec	$F_{0.5}$-Meas	F_1-Meas	F_2-Meas
CODI	0.74	0.57	0.7	0.64	0.6
LogMap	0.85	0.5	0.75	0.63	0.54
MaasMatch	0.83	0.42	0.69	0.56	0.47

Table 3: Average size of alignments, number of incoherent alignments, and average degree of incoherence.

Coherence	Size	Inc. Alignments	Degree of Inc.	Reasoning problems
CODI	9.5	0/91	0%	0
LogMap	8	8/91	2%	0
MaasMatch	7.5	21/91	4%	0

select a set of proposed tags to use in the new or modified content. Second, we will use CODI as a lazy alignment tool server side, this way we make alignments and Ontology enrichment with only our data base information, therefore we reach a coherent set of knowledge. Third, with the database knowledge properly tagged, grouped, and with its own representation (as ontologies) we will assign or suggest tags, content and activities by LogMap in real time when it's necessary, therefore we finally set up a mixed strategy with CODI and LogMap advantages. After this test is done, the next step will be to mix our schema, tags, ontologies and generated taxonomies with a set of taxonomies made of every verb and name of interest in our case, it's like a WordNet but specialized and limited within our target field of interest.

About the algorithms tested, there is still a lot of work to do. Firstly, in the sense of runtime-coherence we need to look for a good method to reduce the runtime like LogMap does, but preventing the coherence loss like it's done by CODI. This is a great challenge and a key step to really using Ontology Matching in real systems.

Secondly, almost every algorithm depends on mapped knowledge, only a few does auto-mapping, and those who do it they have a very low confidence on the generated data. With machine learning technologies this can be improved by creating more useful and reliable ontologies. Thirdly, to improve precision-recall ratios and to improve capabilities to face new problems and scalability.

These three challenges are the main and permanent lines of improvement in every algorithm that has been made, so they can't be removed from the scope.

ACKNOWLEDGEMENTS

This work is part of the project *"TSI-090500-2011-36 - Ministerio de Industria, Turismo y Comercio"*, and was also supported by Sandra Castro, Noelia Gil and J.M. Castro from Intellectia Bank S.A.

REFERENCES

Batini, C. and M.Lenzerini (1986). A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys*, 18(4).

Bellahsene, Z., Bonifati, A., Duchateau, F., and Velegrakis, Y. (2011). On Evaluating Schema Matching and Mapping.

David, J., Euzenat, J., Scharffe, F., and dos Santos, C. T. (2010). The Alignment API 4.0. *IOS Press*, 1:1 – 8.

Euzenat, J. and Shvaiko, P. (2007). *Ontology Matching*. Springer.

Huber, J., Sztyler, T., Nößner, J., and Meilicke, C. (2011). Codi: Combinatorial optimization for data integration: results for oaei 2011. In (Shvaiko et al., 2011).

Jiménez-Ruiz, E., Morant, A., and Cuenca Grau, B. (2011). LogMap results for OAEI 2011. In *Proc. of the 6th International Workshop on Ontology Matching (OM)*, volume 814. CEUR Workshop Proceedings (CEUR-WS.org). http://ceur-ws.org/Vol-814/.

Sakarkar, G. and Upadhye, S. (2010). A survey of software agent and ontology. *International Journal of Computer Applications*, 1(7).

Schadd, F. C. and Roos, N. (2011). Maasmatch results for oaei 2011. In (Shvaiko et al., 2011).

Shvaiko, P. and Euzenat, J. (2008). Ten challenges for ontology matching.

Shvaiko, P., Euzenat, J., Heath, T., Quix, C., Mao, M., and Cruz, I. F., editors (2011). *Proceedings of the 6th International Workshop on Ontology Matching, Bonn, Germany, October 24, 2011*, volume 814 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Spaccapietra, S., Parent, C., and Dupont, Y. (1992). Model Independent Assertions for Integration of Heterogeneous Schemas. *VLDB Journal*, 1:81 – 126.