Ontology Matching: current status, dilemmas and future challenges

Konstantinos Kotis¹ and Monika Lanzenberger²

¹Dept. of Information & Communications Systems Engineering
University of the Aegean, Karlovassi, Samos 83100, Greece
kotis@aegean.gr

²Vienna University of Technology
Institute of Software Technology and Interactive Systems
Vienna, Austria
lanzenberger@ifs.tuwien.ac.at

Abstract

Ontology matching is still one of the hottest topics of Semantic Web research. The aim of this position statement is three-fold. Firstly, to briefly update the research community about the “where are we now” in ontology matching. Secondly, to trigger discussion on dilemmas and critical questions as these were recently identified in the latest ontology-matching-related research events. Thirdly, to comment on visible challenges that may influence the future of this hot topic and consequently the Semantic Web research in general, pointing on the “where shall we go” in the near future.

1. Current Status in Ontology Matching

Ontology Matching is still one of the hottest topics in the Semantic Web research. Matching ontologies (or schemas) is a critical operation in many application domains, such as Semantic Web, ontology integration, data warehouses, e-commerce, query mediation, etc. It takes as input two ontologies, each consisting of a set of discrete entities and determines as output the relationships (e.g., equivalence, subsumption) holding between these entities. Many diverse solutions to the matching problem have been proposed so far. Although, there is a difference between schema and ontology matching problems, techniques developed for each of them has been of a mutual benefit [1]. Specific criteria has been proposed for evaluating and distinguishing between matching approaches, based on

(a) the different kind of the input of the algorithms, e.g. labels assigned to the entities, entities’ internal structure and the type of their attributes, and relations with other entities, (b) the characteristics of the matching process, e.g. approximate or exact nature of its computation, or the interpretation type of the input data e.g. syntactic, external, semantic, and (c) the output of the algorithms e.g. whether the form of the matching is a one-to-one, one-to-many or many-to-many correspondence between the ontology entities. Other significant distinctions in the output results reported so far are: a) confidence and probability percentages of the resulted mappings, b) the kind of relations between entities a system can provide (equivalence, subsumption, and incompatibility).

Several tools have been developed towards solving the ontology-matching problem, either in a semi-automatic or fully automatic way. Human-involvement during the process is usually in a trade with the precision and recall percentages of the resulted mappings. Fully automated tools are still “looking for” higher accuracy and they are continuously evaluated in international contests e.g. OAEI contest [2]. Still, both automated and semi-automated tools are suffering in their performance. For instance, most of them cannot handle large real-domain ontologies, e.g. Medical or Biology, although more and more realistic test-beds are used to evaluate ontology matching tools (scalability problem).

Beyond ontology matching methods, tools, and evaluation initiatives/frameworks, recent efforts have been made on ontology-matching-tool-design frameworks [3]. Such frameworks allow developers to
use APIs not only for supporting the devising of ontology matching methods but more importantly to provide a mean for synthesizing them towards developing robust tools that produce more accurate mappings.

Most recently, ontology matching methods were used to solve the Semantic Web services matchmaking problem [4], transforming the problem of discovering web services into a matching problem between an ontology description representing a service request to an ontology description representing an offered service. Although there are a few tools implementing such an approach still there is more to be done towards improving the process.

2. Dilemmas and Critical Questions

There has clearly been a large effort from the research community to provide many diverse solutions to the matching problem by developing a variety of tools, yet it does not appear to be a dominant one that is a benchmark for designing other matching tools. This may be due to the fact that there is a variety of domain-specific user or application needs. We conjecture that we could not nominate “the best tool” yet, since we still can identify a variety of important dilemmas and critical questions that users may still have when choosing an ontology matching tool within a specific problem-solving context. Based on users’ preferences, what tool they should choose:

a. A fully or a semi-automated one? What is the grade of human-involvement if any and how this influences the accuracy of the resulted mappings? How much time do they have to spend on validating the resulted mappings? How can they ensure the validity of the resulted mappings without their involvement? At the end, what is more critical for the application they use; to invest on human-involvement and validating resources or on automation of the ontology matching tool which the application integrates?

b. A tool with very high precision and indifferent recall or the other way around? What is the balance between precision and recall that is required by their application? What is the trade-off between precision and recall and how this can be tuned if possible to suite user preferences? At the end, what is more critical for the application they use; to achieve the higher precision the ontology matching tool can give, the higher recall, or an optimum trade-off between them?

c. A tool with very high precision and recall but with high computational complexity (need for powerful computational resources) and rather slow execution time? What is the percentage of precision and recall they are willing to sacrifice in order to speed up the ontology matching process and consequently their application? At the end, what is more critical for the application they use; to achieve the higher precision and recall the ontology matching tool can give or to obtain the mappings as fast as possible?

d. A tool able to support the powerful features of an expressive ontology language such as OWL Full as far as possible but with no computational guarantees? Or a tool, which restricts the type of similarity features making the ontology matching a weaker but more assessable process? In order to bridge the gap between classes and instances OWL Full is necessary, although it causes severe reasoning problems and computational complexity.

e. A tool providing sophisticated artificial intelligence without the need for human intervention? Or a tool doing simple similarity calculations and relying on the users’ knowledge as well? If user interaction is essential we must provide a way for users to analyse the results of an ontology matching process and understand the characteristics of the source ontologies. While many tools generate lists of mappings between entities it is difficult to analyse and validate these mappings without examining every pairwise correspondence in the output files and even then it is an overwhelming task. Is there a need for new human-readable ontology-matching interfaces and how these should be designed?

f. A tool to locate and resolve mismatches between ontologies or a tool to spot the analogies and suppress the other sinister fields? Besides mismatches on the language level, several mismatches on the ontology level may occur [5]. Many of them, e.g., coverage and modelling style, need to be solved manually.

We conjecture that all the above questions and dilemmas should be put in the context of a specific application and user needs in order to be able to answer them. For instance, in a Semantic Search Engine application where ontology matching methods/tools are used to match query-ontologies against OWL documents for delivering the best SWD (semantic web document) ranking, execution time is rather important [6], however it is obvious that precision and recall is also of a major importance. A trade-off between these three parameters should be detected and an optimum solution should be applied. However, in some cases, an optimum solution can be reached by differentiating the importance of the parameters. These cases need to be further investigated and reported. The detection of the trade-off between two or more parameters that influence the performance of an ontology matching tool
within a specific application context is not a trivial task at all. It will require rather exhaustive testing and evaluation before ending up with a secure and optimum solution.

3. Challenges and Future Directions

Although a lot has been done towards tackling the problem of ontology matching [2], the research community still reports open issues that impose new challenges for researchers and underline new directions for the future. This paper briefly comments on such issues however the list may not be exhaustive at the time of reading it:

1. **Scalability:** Most of the implemented and evaluated ontology matching tools suffer from handling large ontologies (e.g. from Medical, Biology domains). Real problems in specific application contexts require scalable solutions as a first priority. Future ontology matching tools should be able to provide such capability.

2. **Speed/Automation/Accuracy tuning:** Tools are currently emphasizing on maximizing specific parameters of their performance such as speed, automation or accuracy. The most common case is the maximization of the performance for one of these parameters, and the other parameters are neglected or largely influenced in a negative manner. Future directions should be towards a fine tuning of all parameters such as the overall performance of an ontology matching tool (and of the application that is integrated in) to be leveraged.

3. **Background knowledge:** Experiments of matching an ontology to another while using one (or more) much larger and detailed ontology of the same domain as background knowledge have been recently announced [2]. The extensive use of domain-related background knowledge in the ontology matching process has positive effects on recall, but does not seem to scale well with large ontologies [2]. The challenge in this case is to adopt such an approach without sacrificing the overall tools’ performance. Future directions may be towards discovering an optimum point in the background-knowledge/performance trade-off.

4. **Ontology matching frameworks:** Although ontology-matching-tool-design frameworks have been already developed [3] supporting developers towards not only devising ontology matching methods but also synthesizing them towards realizing new and robust tools that produce more accurate mappings, their performance need to be further investigated. Scalability, speed, and input ontology type compatibility should be further investigated towards delivering a model-framework that could be used by the research community to device ontology matching tools for any user or application preferences.

5. **Ontology matching visualization:** Development and advancement of automated tools is a constitutive goal in the Semantic Web research community and such tools support fundamental steps in ontology matching. Nevertheless, we currently assume that it is not possible to fully automate the process of ontology matching. However, depending on the specific goals and preferences of the user, more or less automation will be feasible. But still several issues in ontology matching need to be performed and decided by humans in order to ensure the quality, appropriateness, and relevance of the matching results. Interpreting an entity of one ontology in the context of the knowledge of another ontology is a cognitively difficult task since it requires the understanding of semantic relations among entities of different ontologies [7]. At this point visualization comes in. The use of visualization techniques to graphically display data from ontology matching results facilitates user understanding of the meaning and consequences of the ontology matching. Largeness, complexity, and heterogeneity of ontologies and their matching results bring about a bundle of challenges, which need to be addressed in future research.

4. Conclusion Statement

Although major advances in the ontology matching research topic have been presented so far, our position is that there are still important open issues to be tackled towards a more efficient and effective use and integration of ontology matching tools with current applications. This position statement identifies specific issues and challenges that future topical research should consider. Focusing on specific research questions raised in this article should advance the
ontology matching research and the realization of the Semantic Web consequently.

References


