Towards an Adaptive Tool and Method for Collaborative Ontology Mapping

Ramy Shosha, Christophe Debruyne, Declan O'Sullivan

CNGL Center for Global Intelligent Content, Knowledge and Data Engineering Group,
School of Computer Science and Statistics, Trinity College Dublin, Ireland
{shoshar,debruync,declan.osullivan}@scss.tcd.ie

Abstract. Linked Data makes available a vast amount of data on the Semantic Web for agents, both human and software, to consume. Linked Data datasets are made available with different ontologies, even when their domains overlap. The interoperability problem that rises when one needs to consume and combine two or more of such datasets to develop a Linked Data application or mashup is still an important challenge. Ontology-matching techniques help overcome this problem. The process, however, often relies on knowledge engineers to carry out the tasks as they have expertise in ontologies and semantic technologies. It is reasonable to assume that knowledge engineers should require help from the domain experts, end users, etc. to contribute in the validation of the results and help distilling ontology mappings from these correspondences. However, the current design for the ontology-mapping tools does not take into consideration the different types of users expected to be involved in the creation of Linked Data applications or mashups. In this paper, we identify the different users and their roles in the mapping involved in the context of developing Linked Data mashups and propose a collaborative mapping method in which we prescribe where collaboration between the different stakeholders could, and should, take place. In addition, we propose a tool architecture based on bringing together an adaptive interface, mapping services, workflow services and agreement services that will ease the collaboration between the different stakeholders. This output will be used in an ongoing study to constructing a collaborative mapping platform.

Keywords: Semantic Web-based Knowledge Management, Semantic Mashups, Ontology Mapping, Ontology Mapping Engineering, Collaborative Mapping.

1.   Introduction

The Linked Data (LD) initiative is making available a huge amount of data on the Web of which some of are sharing, to some extent, parts of their (application domain) [1]. The LinkedMDB[1] and DBpedia[2], for instance, have different ontologies to state

---

[1] http://www.linkedmdb.org/

[2] http://dbpedia.org/

things about movies, actors, directors and so forth. Since Linked Data datasets do not necessarily use the same ontologies, one is faced with the problem of managing heterogeneity in meaning and representation when information from different sources need to be merged, integrated or combined. Ontology matching is concerned with tackling the problem of this semantic heterogeneity by proposing solutions to find correspondences between semantically related entities of ontologies [2]. These correspondences are then curated and refined to create mappings that prescribe how ,instances of, one ontology's concepts relate to ,instances of, another ontology's concepts. Mapping management methodologies are concerned with the activities related to the creation of ontology mappings from the identification of the ontologies to be aligned and the discovery of correspondences to rendering the mappings in artifacts that computer-based systems can execute [3].

Though rooted in related work in the fields of, amongst others, database [4] and XML schema [5] matching, ontology matching has its own peculiarities such as the assumptions adopted on the Semantic Web and the ontology languages adopted for reasoning tasks. Ontology matching has been around for over a decade and yet still has many challenges to overcome [6]. Some of these challenges include finding appropriate matchers or configuration thereof fit for particular tasks, designing ways to involve users in the matching process that reduce the burden involved, and supporting interactions between users to support social and collaborative matching and mapping and the managing of the created mappings.

The engineering of ontologies and the engineering of ontology mapping processes are very similar in that both are cognitive intensive activities and in that both artifacts are created for a particular purpose. Furthermore, that purpose will further influence various variables in the process and is driven by a community of stakeholders with varying degrees of expertise in knowledge engineering, IT literacy and (application) domain expertise. In addition, similar to (collaborative) ontology engineering, the tools for ontology matching and mapping are often developed for users with expertise in knowledge engineering [7].

In this paper, we propose to leverage user involvement in collaborative ontology mapping to tackle the social and collaborative matching problem [6]. We can benefit from the collaboration to enhance the matching results. The user input in the matching process will help in generating results that are more accurate and speed the process in the same time. This will be achieved by focusing on adapative interfaces tailored for the different types of stakeholders. We start by identifying the various types of stakeholders in Section 2 and by introducing a collaborative ontology mapping lifecycle in Section 3. We will also indicate in which activities the stakeholders are involved and where collaboration could and should take place. The cross referencing of user types and tasks helps us identify where different perspectives (i.e., interfaces) on the tasks are needed, which in turn motivates the adoption of an adaptive user interface for use in the collaborative ontology-mapping environment. Section 4 presents the proposed tool framework behind the ideas proposed in this paper. Our ideas will be exemplified with the use case of finding correspondences between datasets on the Linked Data Web for the creation of mashups. Section 5 compares our approach with the current state of the art and, finally, Section 6 concludes our paper by indicating next steps. Where necessary, we integrate background information in the relevant sections.

2.   Stakeholders in Collaborative Ontology Mapping

The Linked Data Web is available for all and allows for innovative and creative application be developed on top of very simple, standardized technologies such as URIs, RDF, SPARQL and the HTTP platform. The Linked Data Web is actually a complex socio-technical environment in which various types of participants are important. Take, for instance, the creation of semantic mashups, which are application that combine multiple LD datasets, use RDF as the data model and SPARQL to support task execution [8]. In this setting, we not only have knowledge engineers, but also (web) developers, domain experts, and end users. We already stated that ontology-matching tools help find similarities and differences between ontologies and knowledge bases by identifying potential correspondences. In [9], however, the authors considered a fully automated mapping impracticable. Human intervention will be crucial to guarantee a certain quality in the mapping process. User involvement in the mapping process is still however considered a challenge [6], even more so since ontology-matching tools are typically designed for use by knowledge engineers [7].

Knowledge engineers have expertise in knowledge modeling and the technical skills needed to use the ontology-matching tools and execute the process. Knowledge engineers are, however, not necessarily completely familiar with the universe of discourse. Domain experts, on the other hand, have that domain knowledge, but are not necessarily familiar with knowledge modeling or programming. This makes it typically hard for domain experts to participate in the mapping process without the help of a knowledge engineer. In addition, in the context of creating Linked Data mashups the following types of users are also of concern: web developers and end users. Web developers are acquainted with developing applications and, though not necessarily acquainted with the domain or semantic technologies, can provide valuable input by their knowledge gained while developing the application. It will also be the developer that will execute the created mappings. End users will use the mashup, but who are not necessarily "tech savvy".

In this paper, we thus distinguish four types of stakeholders who can contribute to the mapping process. Again, similar to ontology engineering, the creation of mappings is not a trivial task. From scoping the mapping project to the creation of an executable mapping, all are driven by the social interactions within the *community* of stakeholders that lead to the necessary *agreements* and decisions to proceed from one-step to another. Since ontology mappings are created to suit a need of a particular community of stakeholders, we need to promote those stakeholders as first class citizens in the matching and mapping processes. Designing the mapping engineering process as a collaborative effort necessitates the formulation of tasks (the method) and the creation of adequate tools and interfaces tailored to each of the different types of users (tools). By doing so, we are able to leverage both matching and mapping tasks for all stakeholders. To aid that community in reaching those agreements, one has to prescribe and orchestrate the different steps, activities and tasks that need to take place. In our LD mashup scenario, each type of stakeholder will have a specific set of tasks to perform in the mapping process, some of which are collaborative in nature. A method for collaborative mapping engineering will be proposed in the next section.

3.   A Collaborative Ontology Mapping Engineering Method

Here, we propose a collaborative mapping method that allows different types of users to work in collaborative settings to discover correspondences between data ontologies to build and enrich the data model in the semantic mashups applications. Our approach differs from the state of the art (see Section 5) in that it addresses the problem of user involvement by looking for adequate means to leverage participation. We do this not only by defining the mapping and matching process explicitly as a collaborative undertaking, but also by addressing specific usability requirements by developing interfaces that are tailored for each type of user for each specific task. We will thus propose an ontology mapping workflow in which we prescribe the different activities and link those to the different user types. Whenever different user types are linked to an activity or different user types are involved between two subsequent activities, some form of collaboration will be necessary.

The mapping process as shown in Fig. 1 has two scenarios when dealing with mappings between the ontologies: creating a new set of mappings and reusing an existing mapping. The creation of a mapping starts with defining data sources that need to be combined. After grouping these sources, a feasibility study is then conducted to identify the set of requirement for the mashups applications.

The next set of activities in the mapping process is to look for and assess reusable mappings in order to force reuse. Two alternative types of reuse can take place in the process; the first type is to use the mapping *as is* (which is a decision that is taken when the retrieved mapping fits the application requirement); the other type is to feed those mappings to the matchers in order to produce better results. Following on this, the process continues if the mapping were found to be reusable (to feed to the matchers) or if they were created from scratch. The activity set starts by retrieving matchers, assessing those matchers, finding a suitable set for execution, configuring the matchers and then running them.

The next set of activities will be dedicated to evaluating correspondences generated by the matchers, reconfiguring matchers for better results and amending results in order to proceed with the mapping creation or to be used as an input for the matchers.

After that, the flow starts with rendering the mapping to the stakeholders for assessment and evaluation and reaching an agreement as to whether to proceed with the creation or going back to certain a particular step for generating better results. The roles of the different stakeholders in this collaborative method are as follows:

- The knowledge engineer (K) is involved in all the process activities as he has expertise in ontology mappings and thus his input is valuable. In the first activity in the process, knowledge engineers are able to look for (additional) data sources to be merged, integrated or aligned. Following on the knowledge engineer is able to assess whether the project is actually a semantic heterogeneity problem. In addition, the knowledge engineer is involved in all the data modeling tasks including matchers' retrieval, selection and configuration, and mapping curation, storage and publish.
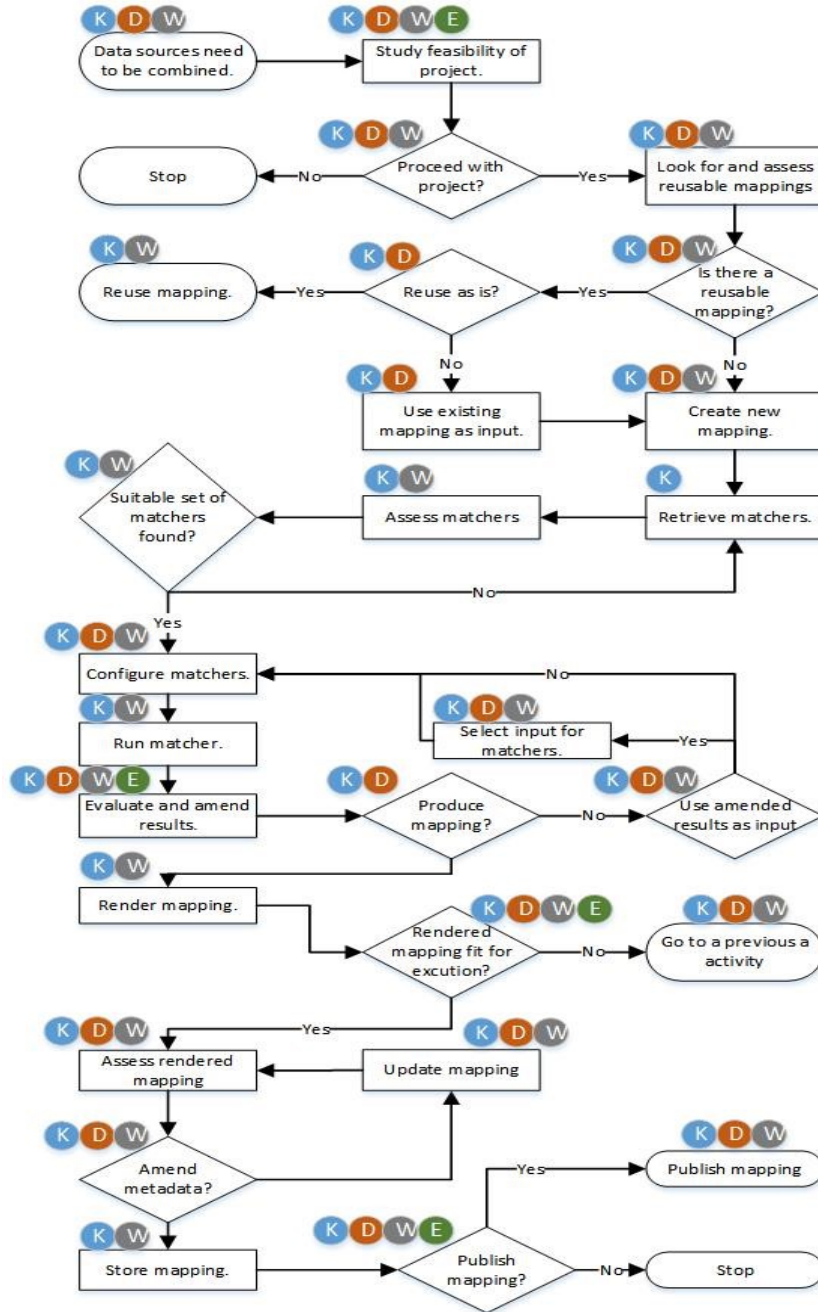
Fig. 1. The collaborative mapping process. Each activity is annotated with symbols that represent knowledge engineers (K), web developers (W), domain experts (D) and end users (E). Whenever an activity has two or more types of stakeholders, collaboration should take place.

- The domain expert (D) has the knowledge in a particular area of endeavor, his input in particular area or topic is valued due to his domain knowledge. The domain expert might also be a (business) domain expert; someone involved the application development that is not necessarily a developer. A manager, for instance. The domain expert participates in the project feasibility study and evaluating the results of the mapping process. In other activities like matchers configuring, knowledge engineer and web developer only consult the domain expert in order to enhance the matchers and results.
- Web developers (W) know in which environment (e.g., runtime environment, operating system, etc.) the mappings will be executed, and thus their input is valuable in choosing the appropriate tools. The web developer will mostly contribute to the requirements that the mapping should comply with. In addition, their expertise in development allows them to partake in some of activities regarding matchers and mapping handling.
- The end user (E), the user who will use the mashup in the end, might have some expertise in the domain but not necessarily on ontology matching or development activity. The audience and type of application is not necessarily known beforehand. Is this a mashup open to all, or for a specific community? Therefore, the end user input is necessary in the feasibility study activity. In addition, we can reuse some of the artifacts (case studies, requirements analyses, etc.) as input in the same activity.

Agreements are reached by means of discussions, voting, etc. Discussions will be structured using Issue-Based Information Systems (IBIS) [10], where people reach consensus via *examples*, *justifications*, or *evaluations* that can be either *objecting* or *supporting*. A particular useful example for adopting IBIS is that one can connect an *objecting example,* – i.e., counterexample – to a correspondence to refute it.

The construction of mapping is challenging as mappings are built for a specific purpose and is a collaborative undertaking involving all stakeholders is crucial. Take, for example, the collaboration that takes place for configuring the matchers, several parameters needs to be set by the knowledge engineers for instance weights and thresholds. A discussion has to take place between all three to come up with the best configuration for these parameters.

All metadata, documentation and collaborative activities generated throughout the process are rendered and saved to enable tracing back the process of decision-making.

4.   Towards an Adaptive Tool for Collaborative Ontology Mapping

We aim to build a web-based ontology-mapping construction tool based on our collaborative ontology mapping method described in the previous section. Most ontology mapping tools and environments have been developed as standalone applications that need to be installed on one's machine. Web-based applications have the benefit that one only needs to have a browser installed. There are exceptions activities. [11], for instance, created a web-based environment for sharing and discovering mappings.

Another challenge with current design of the ontology-matching tools is that it often appears cluttered, presenting irrelevant and relevant information together, which makes mapping construction difficult even for knowledge engineers as asserted in [7],[9], Cluttered interfaces negatively affect user usability [12, 13]. Knowledge engineers and web developers are usually working with development environments and are used to process a lot more information on one screen. In order to involve domain experts as well as end users in the mapping process, however, we believe that usability needs to be significantly improved. We argue that interfaces of mapping-development tools need to be adaptive in order to provide better support for the range of stakeholder users that will be involved.

Thus, the tool under development aims to enhance the matching process from two perspectives: enhanced usability and support for a collaborative mapping. By designing the mapping process in a way that makes it a collaborative undertaking and by adapting (the interfaces of) the tasks to the expertise and role of users, we aim to leverage the expertise of the different stakeholders and reduce the cognitive load overall.

An adaptive user interface will solve the problem of clutter and will facilitate the involvement of different users in the ontology development tasks. An overview of the design is presented in Fig. 2. The User Interface layer interacts with the Adaptation Engine to generate an interface based on the user's profile, role being played and tailored to the particular task involved. The Adaptation Engine thus orchestrates the activities and relies on or mediates between the users and the following components: 1) Mapping services constituting all the tools needed for identifying correspondences (with the matchers), and creating and managing mappings.2) Workflow services, which is responsible for orchestrating and monitoring the tasks (not the interfaces) and the generation of metadata concerned with each task in the ontology mapping process. 3) Agreement services, which are components that will aid and capture the agreements, and thus decisions, made by the community of stakeholders while constructing a mapping.
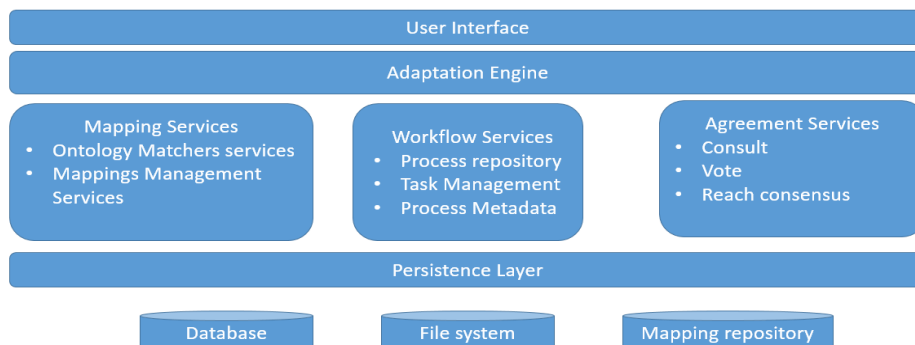

Fig. 2. The tool's architecture

Note that capturing the interactions not only renders the process more traceable, but also more transparent. Decisions are linked to the different parts of an ontology mapping and can thus be looked up for inspection.

The intended key benefit of the adaptive tool is that it will stand out as a collaborative mapping environment that allows domain experts, web developers, and end users to be involved in the mapping process.

5.   Related Work

Ontology matching is an important part of ontological engineering activities, and should thus be supported by adequate methods and tools. Yet, at present, most of the approaches to ontology-matching lack support for collaboration at both the level of method and tool. Few approaches have focused on methodology [14]. There are some methodological guidelines proposed for ontology matching [15], but those guidelines do not treat the ontology mapping process as a collaborative undertaking and only focuses on the creation and management of mappings; the guidelines do not take into account activities such as feasibility studies when the need for creating an ontology mapping arises. As for the state of the art, throughout sections 1 to 3 we have stated that most methods either do not take into account collaboration or focus only on a few of the activities (e.g., discovery and sharing [11]).

Collaboration has been studied in, for instance, in [16] the focus was in circulating the correspondences between communities. The shared correspondences will have information about the author's name, application domain and trust in a community. This approach is mainly designed for domain experts validating the produced correspondences, edit and share them. It does not take into consideration other types of users that might be take place in the mapping process.

There is also the work done by Colm Conroy, reported in [17–19], in converting the mapping process from a one-time task to one performed over several sessions in an iterative manner. Mapping occurs in smaller sessions over time that allows the workload and time per session to be reduced. A positive consequence of an iterative process is it will allow users to see the impact of their mapping decisions between sessions. A downside is that an iterative process will only develop partial mappings that will continue to be refined over time. Colm Conroy engages non-technical users (Or as denoted in his research as end users) through the Tag-Based Mapping tool.

Another work done to find alignment and mappings between vocabularies cultural heritage collections through a web based repository service [11]. While designing the repository service, it was taking into consideration the need for vocabulary experts in the mapping activities. The tool support expert and novice users in finding concepts in a given vocabulary by performing semantic search across multiple collections and browse indexed metadata attached to the alignments within the collections. Metadata is used to find alignments that link any two vocabularies. In such cases, users might amend some of mapping result manually, add missing mapping or modify the existing ones, this raises a conflict between users that can be solved through some given agreement mechanisms, which is missed in this tool.

6.   Conclusions and Future Work

The creation of ontology mappings on the Linked Data Web is not a trivial task as such mappings are the result of numerous agreements between different types of stakeholders each with their own skill sets and valuable input. The stakeholders in building a Linked Data mashups are knowledge engineers, domain experts, web developers and end users. In this paper, we have identified several gaps in the state of the art. Firstly, user engagement is a big challenge as tools are often created for knowledge engineers. Another reason that user engagement is a challenge is that tools are often created as standalone applications, which limits collaboration. Secondly, tools that do provide a (web-based) tool often only look a few activities of the ontology mapping lifecycle. Thirdly, a collaborative tool is not sufficient. In order to accommodate and engage the different users in the collaborative undertaking, the different stakeholders will need different perspectives on the same tasks. The challenge here is to define how these interfaces and tasks should be orchestrated. Adoption of adaptive interfaces would be suitable to tackle this challenge that lead to ease the involvement of other stakeholders besides the knowledge engineer. This will allow distributing the workload between users and give the opportunity to enhance what others have done.

Finally, to encourage traceability and render the ontology mapping process transparent, one needs to structure the interactions between the stakeholders so that one can relate parts of the ontology mapping with the discussions; e.g. to explain why certain choices were made.

Next to identifying the challenges, we proposed a collaborative method supported with adaptive interfaces in which we relate the different stakeholders to the different activities. This exercise identified which activities should be collaborative in nature and hence adaptive interfaces could play a role. A general idea of the tool is presented in which the different components are identified. This tool would bit in a broader Linked Data mashup method (agile, unified process, etc.). Future work consists of using and validating the adaptive tool in a realistic Linked Data mashup creation scenario.

References

1.   Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. Semantic Services, Interoperability and Web Applications: Emerging Concepts. 205–227 (2009).

2.   Euzenat, J., Shvaiko, P.: Ontology matching. Springer, Heidelberg (2007).

3.   Bellahsene, Z., Bonifati, A., Rahm, E.: Schema matching and mapping. Springer, Heidelberg (DE) (2011).

4.   Batini, C., Lenzerini, M., Navathe, S.B.: A comparative analysis of methodologies for database schema integration. ACM computing surveys (CSUR). 18, 323–364 (1986).

5. Bernstein, P.A., Madhavan, J., Rahm, E.: Generic schema matching, ten years later. Proceedings of the VLDB Endowment. 4, 695–701 (2011).

6. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. IEEE Transactions on Knowledge and Data Engineering. 25, 158–176 (2013).

7. Falconer, S.M.: Cognitive support for semi-automatic ontology mapping. Doctoral dissertation, University of Victoria. (2009).

8. Heath, T., Bizer, C.: Linked data: Evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology. 1, 1–136 (2011).

9. Noy, N.F.: Semantic integration: a survey of ontology-based approaches. ACM Sigmod Record. 33, 65–70 (2004).

10. Kunz, W., Rittel, H.W.J.: Issues as elements of information systems. Institute of Urban and Regional Development, University of California Berkeley, California (1970).

11. Van der Meij, L., Isaac, A., Zinn, C.: A web-based repository service for vocabularies and alignments in the cultural heritage domain. In: 7th Extended Semantic Web Conference (ESWC). pp. 394–409. Springer, Heraklion, Crete, Greece (2010).

12. Dahl, Y., Svendsen, R.-M.: End-user composition interfaces for smart environments: A preliminary study of usability factors. In: Design, User Experience, and Usability (DUXU). Theory, Methods, Tools and Practice. , Orlando, FL, USA (2011).

13. Wesson, J.L., Singh, A., Van Tonder, B.: Can adaptive interfaces improve the usability of mobile applications? In: Human-Computer Interaction. pp. 187–198. Springer (2010).

14. Mochol, M.: The methodology for finding suitable ontology matching approaches. Doctoral dissertation, Freie Universität Berlin, Germany. (2009).

15. Euzenat, J., Le Duc, C.: Methodological guidelines for matching ontologies. In: Ontology engineering in a networked world. pp. 257–278. Springer (2012).

16. Noy, N.F., Griffith, N., Musen, M.A.: Collecting community-based mappings in an ontology repository. In: 7th International Semantic Web Conference (ISWC). pp. 371–386. Springer (2008).

17. Conroy, C., O'sullivan, D., Lewis, D.: Ontology Mapping Through Tagging. In: Complex, Intelligent and Software Intensive Systems, 2008. CISIS 2008. International Conference on. pp. 886–891. IEEE (2008).

18. Conroy, C.: Towards semantic mapping for casual web users. In: 7th International Semantic Web Conference, ISWC. pp. 907–913. Springer, Karlsruhe, Germany (2008).

19. Conroy, C., Brennan, R., O'Sullivan, D., Lewis, D.: User evaluation study of a tagging approach to semantic mapping. In: The Semantic Web: Research and Applications. pp. 623–637. Springer (2009).