Visual Analytics for Ontology Matching Using Multi-Linked Views

Jillian Aurisano[1], Amruta Nanavaty[2], and Isabel F. Cruz[2]

[1] EVL Lab, Department of Computer Science, Univ. of Illinois at Chicago, USA
jillian.aurisano@gmail.com

[2] ADVIS Lab, Department of Computer Science, Univ. of Illinois at Chicago, USA
aybgr8@gmail.com, isabelcfcruz@gmail.com

Abstract. Ontology matching is the key to data integration on the Semantic Web. Advanced ontology matching systems incorporate a variety of algorithms. However, they do not always guarantee a complete and correct alignment (set of mappings). Hence, user involvement in the matching process is essential for complex ontologies. In this paper, we explore the power of multi-linked views, where actions in one view affect the display of the other views, thereby extending significantly the state of the art in ontology matching visualization in general and that of visual analytics for ontology matching in particular. A preliminary assessment of our approach that uses the ontologies of the OAEI Conference Track points to the effectiveness of our approach.

1 Introduction

Data integration provides the ability to manipulate data transparently across multiple data sources. At the heart of data integration are ontologies and the ability to establish semantic mappings among them using ontology matching [10].

Semi-automatic approaches to ontology matching allow for experts to intervene by validating or eliminating results that were automatically determined and then iteratively incorporating that feedback into the matching process [7, 3, 4]. To perform this determination, analytical reasoning is needed, which, when supported by an interactive visual interface, is called *visual analytics* [1]. In this paper, we propose the AlignmentVis visualization tool, which uses the AgreementMaker ontology matching system [5], but can be easily adapted to other advanced matching systems with a comparable architecture. We describe next the terminology associated with ontology matching systems and describe the architectural components of AgreementMaker.

The process of ontology matching finds semantic mappings between different entities (classes and properties) of a source and target ontologies, by using a wide range of lexical, syntactic, and structural automatic matching algorithms called *matchers*. A matcher produces a similarity matrix where each row represents a source entity, each column represents a target entity, and each cell contains the *confidence score* for the source-target pair. In AgreementMaker, matchers include the Base Similarity Matcher (BSM), the Parametric String based Matcher (PSM), the Vector-based Multi-word Matcher (VMM), the Lexical Synonym

Matcher (LSM), and the Descendant Similarity Inheritance (DSI) matcher [8, 6]. The Linear Weighted Combination (LWC) matcher combines similarity matrices as produced by the automatic matchers using weights determined by a local quality measure [6]. For each mapping, the combined confidence score is stored in the corresponding element of the LWC matcher similarity matrix. Finally, a set of mappings, called an *alignment*, is selected from this matrix according to an optimization criteria [6]. The performance of an ontology matching system is evaluated by comparing the obtained alignment against a gold standard, also called *reference alignment*, created by domain experts.

We interviewed ontology matching experts to identify the analytic tasks that need to be supported by an advanced visualization tool, as summarized next:

Matcher's performance evaluation Expert users need to evaluate the performance of individual matchers and the quality of the final alignment with respect to the reference alignment. Users also want to characterize the mappings into true positives (correct mappings), false positives (incorrect mappings), and false negatives (missed mappings). When no reference alignment is available, the techniques outlined below may be necessary.

Mapping clusters In addition to a high-level evaluation of the performance of each matcher, expert users may take advantage of clusters of mappings that are grouped according to different statistics and then analyze each cluster in order to assess the performance of an individual matcher.

Exploration and comparison The evaluation of the performance of a matcher makes use of exploration and comparison tasks. Views of entity details, through meaningfully designed explorative interactions and through comparative views of the results across different matchers, should help in identifying potential sources of error.

Diagnosis Once errors are identified by using exploration and comparison, this complex task will help to identify the cause of the errors. It is not an individual task, but rather a combination of the previously outlined tasks as users will iterate through them to arrive to a determination.

For these analytic tasks, in this paper we explore the power of multi-linked views, where actions in one view affect the display of the other views [20, 2], therefore extending significantly the state of the art in ontology matching visualization in general and that of visual analytics for ontology matching in particular.

This paper is organized as follows. In Section 2, we outline the most relevant approaches to ontology matching visualization with a focus on visual analytics. In Section 3, we describe in detail all the views we have created, the tasks they fulfill, and how they are linked to one another. In Section 4, we describe the dataset on which we tested AlignmentVis and the environment in which it was developed. In Section 5, we point to a few examples that demonstrate the kind of anomalies that the interface can help detect. Finally, in Section 6, we draw brief conclusions and point to future work that will quantify the benefits of a visual analytics tool like AlignmentVis.

2 Related Work

A recent survey on user involvement for large ontology matching covers several visualization tools [16]. However, those tools do not support fully the necessary requirements laid out by the authors. For those domain expert users that rely on visualization tools for ontology matching, much more functionality is needed including debugging the obtained alignment (set of mappings), observing similar characteristics in a group of mappings, and assessing the contribution of individual matching algorithms to the final alignment. Essentially, those users need a tool that allows them to detect those mappings that are incorrect and confirm the mappings that are correct. In spite of their limitations, we cover next some of the visualization tools in the aforementioned survey and add to them a couple more, which are especially relevant given their focus on visual analytics.

A representation that is cluster based shows both detailed and general information of the matching results and provides in addition a JTree-like visualization [18]. Users can select the level at which they want to cluster the results. For the visualization of each ontology this approach uses a spring-embedded graph drawing algorithm [11, 9]. A drawback of this approach is that only the results of a single matching algorithm can be visualized. Another approach based on a spring-embedded technique was developed for the AgreementMakerLight system [15], which extends AgreementMaker [5] to very large ontologies; it provides a single visualization where both ontologies and the mappings between classes are displayed. However, it is not intended to display more than a few mappings at a time [21]. This technique also does not allow for the users to compare the results of more than one matching algorithm at once.

PROMPT+COGZ is an advanced visualization tool that supports multiple visualizations, including one based on TreeMaps and another one that displays pie charts [14]. TreeMaps have the advantage that they can be used to visualize large amounts of data, but fit in a small area. However, this tool does not seem to be able to show concurrent displays of more than one matching algorithm and also does not provide analytical details about the mappings or about the contribution of an individual matcher to the alignment process. A recent highly interactive visualization based solely on pie charts has two important features: it scales to very large ontologies and can compare different matching algorithms [19]. Its focus on scalability makes it a possible complement to the multi-linked visualization approach of this paper.

A matrix visualization where the classes of both ontologies are placed along the X and Y axes provides a more comprehensive view of the matching process as compared with other methods because it allows for the whole mapping space to be visualized with equal detail. We know of two such visualizations: the one provided by iMERGE [12] and the visual analytics panel provided by AgreementMaker [7]. Both systems support multiple visualizations, including a traditional JTree-like visualization for each ontology with connections between the two ontologies showing the mappings. AgreementMaker has the distinct capability of allowing for the comparison of different matching algorithms side by side and simultaneous navigation across the various similarity matrices. In AlignmentVis, we want to preserve the unique characteristic of AgreementMaker to display the matching results across several algorithms and its applicability to visual analytics for ontology matching [7]. However, we also want to support multiple views in the same panel, including a matrix view.

3 AlignmentVis Design

AlignmentVis addresses the cognitive support requirements for ontology alignment systems, which are meant to facilitate user involvement, by presenting the mapping results in four linked views. First we describe the three views that are related to the same individual matcher, then the fourth view compares all the matchers. The views display: (1) an overview of the mappings obtained between all the entities in the source ontology and in the target ontology, as presented in the *Matcher Output Grid View*; (2) the behavior of the entities of the source and target ontologies with respect to various statistics, as provided by the *Entity Mapping Characteristics Scatter Plot View*; (3) the mappings between entities in the source and target ontologies, which uses the interactive *Ontology Tree View*; (4) the results for all the matchers alongside the reference alignment (when available) for comparative analysis, as enabled by the *Parallel Coordinate View*. The interface of the AlignmentVis tool is shown in Figure 1.

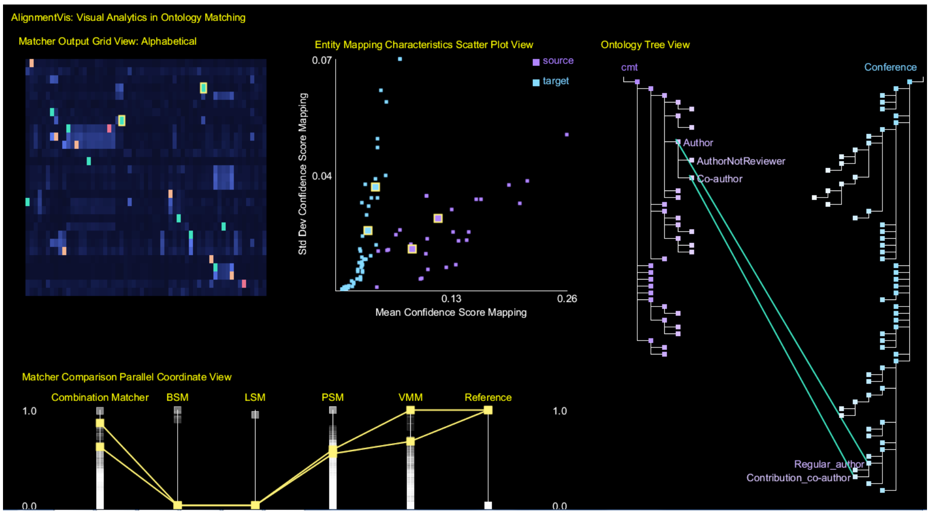


Fig. 1: AlignmentVis user interface.

3.1 Matcher Output Grid View

The Matcher Output Grid View displays a two dimensional matrix where each row represents a source entity, each column represents a target entity, and each cell value represents the confidence score of the selected matcher for a source-target pair. That score ranges from 0 to 1 where values close to 1 indicate high similarity between the source and target entities and values close to 0 indicate high dissimilarity. The confidence score of a mapping sets a color gradient from black for a score of 0 to bright blue for a score of 1. If a cell is colored green then it is a correct mapping. It means that the corresponding mapping is present both in the alignment that is computed by the algorithm and in the reference alignment. If a cell is colored red it is a false negative or missed mapping, which means that the mapping is present in the reference alignment but not in the final alignment. If a cell is colored orange, it is a false positive, which indicates that the mapping is present in the final alignment but not in the reference alignment. The color scheme aims to make the overall performance of the selected matcher immediately evident.

Users can hover over the view to see the confidence score and the labels of the participating source and target entities of the selected mapping. Moreover, as the view is linked to other views, the cell representing the corresponding mapping in the matrix is highlighted by a yellow box whenever a corresponding mapping or participating source and/or target entities are selected in other views.

If an individual source (or target) entity is selected in the other views of AlignmentVis, then its corresponding row or column in the Matcher Output Grid View is highlighted. The Grid View helps users to rapidly explore individual mappings and to observe how each entity from the source ontology is related to the entities of the target ontology.

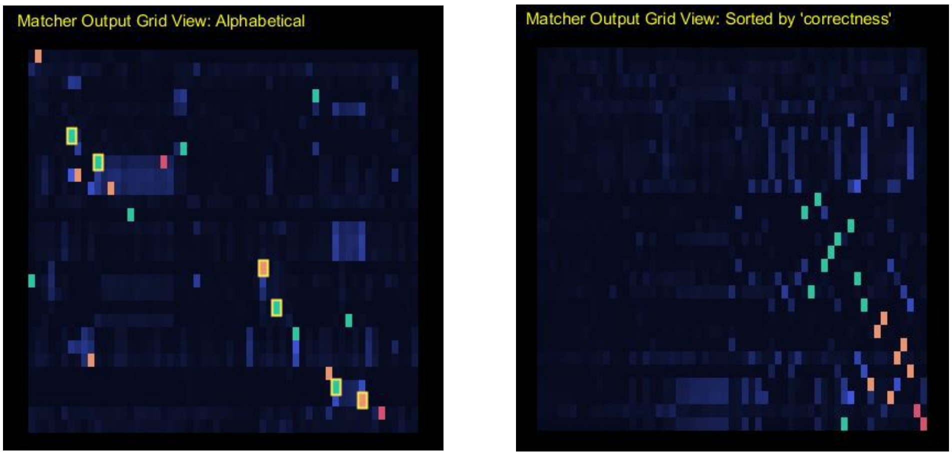
Several reordering features are available for the source and target entities to facilitate the recognition of patterns associated with the detected or missed mappings:

Alphabetical order The labels of the source and target ontology class entities are ordered alphabetically in ascending order. The corresponding rows and columns in the Grid View are rearranged accordingly as shown in Figure 2a.

Ascending order of the mean value of the confidence scores of the corresponding class entity As mentioned earlier, each row represents a source entity and its relation to the target entities. The mean value is computed for each row and then the rows are reordered in ascending order of their mean value. Similar computation and reordering can be performed for each column.

Ascending order of the standard deviation value of the confidence scores of the corresponding class entity The procedure for reordering is as in the previous case, but instead of the mean, the standard deviation is calculated.

Mapping categorization The entities are reordered by first displaying the source entities that are not related to any of the target entities followed by those that are present in the reference alignment. Thereafter, the source entities that are present in the false positive mappings are displayed and lastly the source entities that are involved in the missed mappings are displayed. The same reordering is available for the target entities. This kind of reordering displays distinct mapping clusters with similar characteristics. Users can then explore these entities and associated mappings and look for similar characteristics in the other views. The mapping categorization view is shown in Figure 2b.



(a) Reordered view in ascending alphabetical order.

(b) Reordered view in ascending correctness order.

Fig. 2: Matcher Output Grid View.

3.2 Entity Mapping Characteristics Scatter Plot View

An entity can be described by a vector, where each element indicates a confidence score of the mapping between the entity and all the entities in the other ontology. Various statistics like mean, standard deviation, and correctness can be computed from that vector. These statistics can give an insight into the potential mappings associated with that individual entity. In the Scatter Plot View, which is displayed in Figure 3, entities of the source and target ontology are displayed as nodes in a scatter plot with respect to any of these two statistics, where one of them is displayed in the X axis and the other one in the Y axis. Users can switch between the chosen statistics and exchange the X and the Y axes.

A node is colored depending on whether the representative entity belongs to the source or to the target ontology. The Scatter Plot View helps to identify different characteristics of the source and target ontologies. Users can interact with this view by hovering over the nodes, which become highlighted in the other views. In addition, when users select nodes in another view, they are highlighted in the Scatter Plot View. This view also allows for comparing the performance of an individual matcher with that of other matchers with respect to the computed statistics.

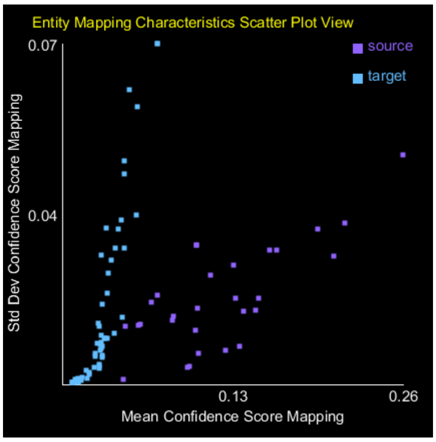


Fig. 3: Entity mapping characteristics using the Scatter Plot View.

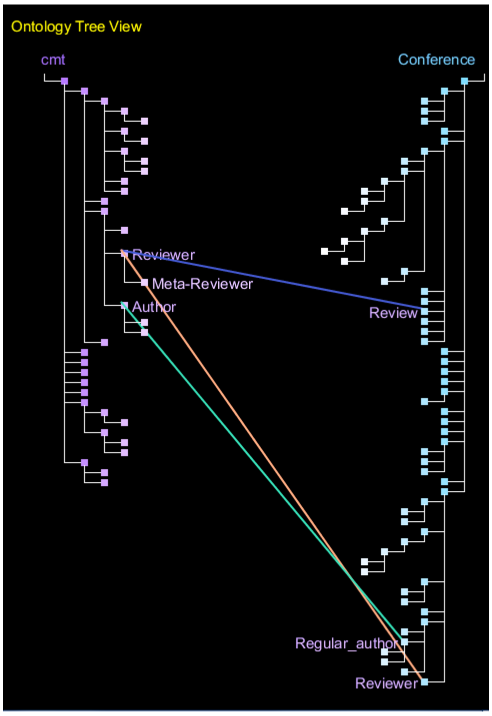


Fig. 4: Ontology Tree View displaying the source and the target ontologies.

3.3 Ontology Tree View

In the Ontology Tree View, which is shown in Figure 4, the hierarchical structure of the source and of the target ontologies are displayed using trees. Users can hover over a section of the tree in order to view the mappings involving the entities under the selected section. Only those mappings that have a confidence score above a predefined threshold for the selected matcher are displayed by a colored line between the source and target trees. The color scheme is the same as in the Matcher Output Grid View. Mappings are available on demand to facilitate the users' focus on entities of interest and to avoid information overload. The related information about the displayed mapping can be viewed in other views due to the multi-linked view feature of AlignmentVis.

3.4 Comparative Analysis of Matchers Using a Parallel Coordinate View

The Parallel Coordinate View, which is shown in Figure 5, is at the heart of the AlignmentVis interface. Each vertical axis represents a matcher on which rectangles associated with the mappings are positioned relative to their confidence score. This allows for users to quickly compare the confidence score associated with a mapping across all the matchers. The minimum value on each axis is 0 and the maximum value is 1. When hovering over any of the vertical axes, the mappings in that area are highlighted and lines are drawn connecting the highlighted mappings across the rest of the vertical axes. The confidence score related to the current position of the mouse on the selected vertical axis is also displayed. The Parallel Coordinate View also helps users identify which matcher plays a dominant role in identifying the mapping. This identification is possible because one of the vertical axes represents the combination matcher. In turn, it is easy to compare the result produced by the combination matcher with the reference alignment. The related information to the highlighted mapping is displayed in the other linked views. Hovering over the Matcher Output Grid View or the Ontology Tree View produces yellow colored lines drawn across all the vertical axes for the selected mappings. In addition, by linking this view to other views, users can analyze whether mappings having similar confidence scores across various matchers tend to have distinct characteristics in the other views or not.

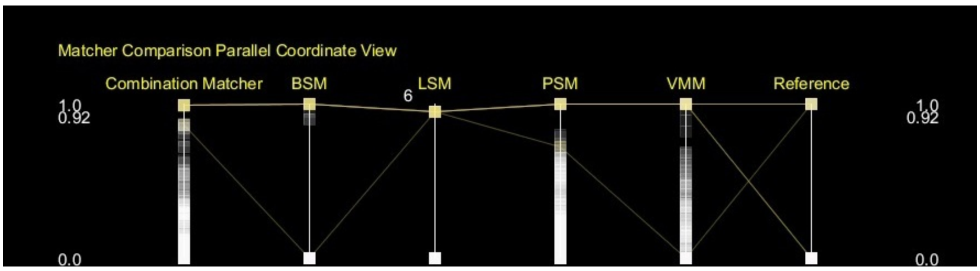


Fig. 5: Parallel Coordinate View.

4 Dataset and Implementation Language

The datasets used for testing and evaluating this interface are from the Conference Track of the Ontology Alignment Evaluation Initiative (OAEI), which is an annual international campaign for the systematic evaluation of ontology matching systems.[3] The Conference Track uses 16 ontologies from the conference organization domain from three types of underlying resources:

1. Actual conferences and their web pages. For example, the SIGKDD ontology is based on the organization of the ACM conference with the same name.
2. Actual software tools for conference organization. For example, the OpenConf ontology is designed using high level concepts from the tool with the same name that was developed for peer-review, abstract, and conference management.
3. People's experience based on their participation in the organization of an actual conference.

These ontologies are suitable for the ontology matching task because of the homogeneity of their domain of interest and of the heterogeneity of their organization, given their very different origins. Each ontology contains less than 200 concepts.

We have used AgreementMaker to perform the ontology matching task for these ontologies and used the similarity matrix and alignment that was produced by AgreementMaker for each of the matchers. We note that AgreementMaker has been the winner for this track, therefore it produces high quality mappings on this dataset [13]. Thus, user interaction and visual analytics can play an important role even when the automatically obtained results are of high quality.

AlignmentVis is implemented in Processing. Processing is an open source programming language and integrated development environment (IDE) built for the electronic arts, new media art, and visual design communities with the purpose of teaching the fundamentals of computer programming in a visual context, and to serve as the foundation for electronic sketchbooks.[4] Processing is built on the Java language, but it uses a simplified syntax and graphics programming model. It allows for quick prototyping and is easy to learn.

5 Evaluation

We tested AlignmentVis with the ontologies of the Conference Track of the OAEI. Each ontology contains less than 200 entities. Till now, most of the ontology matching systems have focused on different ways of visualizing the alignment and very few have made an effort to apply visual analytics to support the involvement of users in the ontology alignment task, therefore is not a standard way to evaluate the benefits provided by tools such as ours. In the absence of an established evaluation methodology, we tested extensively our user interface to evaluate the benefits provided by the multi-linked views to analyze the performance of single matchers and of their combination to produce a final alignment

[3] http://oaei.ontologymatching.org/
[4] https://processing.org/

for the Conference Track. We describe a couple of interesting examples and observations.

In the Ontology Tree View of Figure 4, there is an incorrect mapping highlighted in orange between the source entity *Reviewer* and the target entity *Reviewer* and a correct mapping between the source entity *Author* and the target entity *Regular_author*. Another mapping, in blue, shows a potential mapping between *Reviewer* and *Review*, the only mapping whose value is above a set threshold. Here the domain expert analyzes first the tree view, to see that the distance between *Reviewer* and *Author* in the source ontology is much smaller (they are siblings) than the distance between *Review* and *Regular_author* in the target ontology, a possible indication of an incorrect mapping [17]. In comparison, the green and orange mappings (even if not preserving the sibling relationship), appear acceptable. The expert then analyzes the corresponding Parallel Coordinate View of Figure 6, to discover that all the matchers show high confidence for the mapping between *Reviewer* and *Reviewer*, only contradicted by the reference alignment. This example indicates a possible error in the reference alignment of the Conference Track, which is, in fact, currently undergoing a revision.

For another example that shows how two views can provide complementary information, we focus on Figures 5 (Parallel Coordinate) and 2a (Grid). The former shows that the LSM matcher produces heavily split confidence scores (that is, either 1 or 0). The latter shows the six mappings detected by LSM, of which the majority (four) are true positive mappings. Further interaction will allow for the detailed analysis of each of these mappings in comparison with the results provided by the other matchers.

The Scatter Plot View of Figure 3 shows that the source and target entities display distinct mean and standard deviation statistics. It would be valuable to see whether a similar difference exists between the source and target ontologies of the other OAEI tracks, or whether it is unique to the Conference Track. The Scatter Plot View can contribute to the determination of the intrinsic quality of mapping, given that a high standard deviation may point to the existence of a target entity for which the matcher has a clear preference over the other target entities [6]. This indication can be cross-investigated by the multiple perspectives that are made possible by the unique multi-linked functionality of AlignmentVis.

6 Conclusions

Ontology matching is a key component of data integration. Various lexical, syntactic, and structural automatic matching algorithms contribute to the set of mappings between two ontologies. However, as these algorithms do not guarantee 100 percent accuracy, user involvement is required. Expert users can make real-time decisions for a set of candidate mappings during the ontology matching process, so as to validate or eliminate those mappings. To make such decisions, they benefit from the visualization of the mappings and of the results produced by the various matchers by focusing on the performance of each of them, allowing for statistics to be displayed, mapping clusters to be visualized, and enabling exploration and comparison, so as to diagnose any anomalies in the ontology matching process or to confirm mappings.

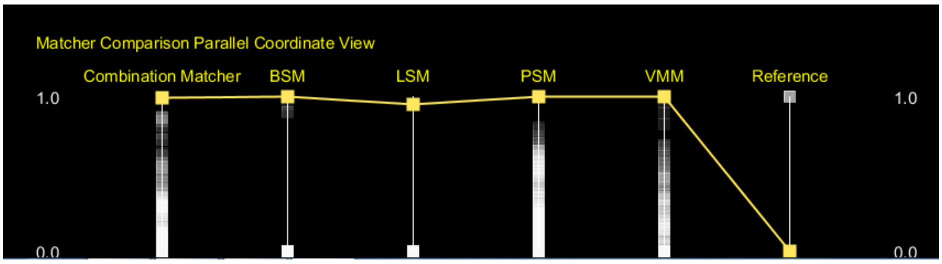


Fig. 6: Parallel Coordinate View for the mapping between the source entity *Reviewer* and the target entity *Reviewer*, which corresponds to the Ontology Tree View of Figure 4.

AlignmentVis provides users with an interactive visual interface, allowing them to conduct analytical reasoning, the two key components of a visual analytics process. In our interactive visual interface, we explore the use of multi-linked views, a known technique in the field of information visualization, yet till now seldom used in the realm of Ontology Matching. Our initial evaluation indicates that the multi-linked views of the interface satisfy important cognitive and interactive user requirements necessary for the ontology matching task. Future work will attempt to quantify the improvement in performance that is obtained from using AlignmentVis.

Acknowledgments

This research was partially supported by NSF Awards IIS-1143926, IIS-1213013, and CCF-1331800.

References

1. Bertini, E., Lalanne, D.: Investigating and Reflecting on the Integration of Automatic Data Analysis and Visualization in Knowledge Discovery. SIGKDD Explorations Newsletter 11(2), 9–18 (May 2010), http://doi.acm.org/10.1145/1809400.1809404
2. Cruz, I.F., Huang, Y.F.: A Layered Architecture for the Exploration of Heterogeneous Information Using Coordinated Views. In: Symposium on Visual Languages and Human-Centric Computing (VL/HCC). pp. 11–18. IEEE Computer Society (2004)
3. Cruz, I.F., Loprete, F., Palmonari, M., Stroe, C., Taheri, A.: Pay-As-You-Go Multi-User Feedback Model for Ontology Matching. In: International Conference on Knowledge Engineering and Knowledge Management (EKAW), pp. 80–96. Springer (2014)
4. Cruz, I.F., Loprete, F., Palmonari, M., Stroe, C., Taheri, A.: Quality-Based Model for Effective and Robust Multi-User Pay-As-You-Go Ontology Matching. In: Semantic Web Journal. IOS Press (2015)
5. Cruz, I.F., Palandri Antonelli, F., Stroe, C.: AgreementMaker: Efficient Matching for Large Real-World Schemas and Ontologies. PVLDB 2(2), 1586–1589 (2009)
6. Cruz, I.F., Palandri Antonelli, F., Stroe, C.: Efficient Selection of Mappings and Automatic Quality-driven Combination of Matching Methods. In: ISWC Interna-

tional Workshop on Ontology Matching (OM). CEUR Workshop Proceedings, vol. 551, pp. 49–60 (2009)
7. Cruz, I.F., Stroe, C., Palmonari, M.: Interactive User Feedback in Ontology Matching Using Signature Vectors. In: IEEE International Conference on Data Engineering (ICDE). pp. 1321–1324 (2012)
8. Cruz, I.F., Sunna, W.: Structural Alignment Methods with Applications to Geospatial Ontologies. Transactions in GIS, Special Issue on Semantic Similarity Measurement and Geospatial Applications 12(6), 683–711 (2008)
9. Cruz, I.F., Tamassia, R.: How to Visualize a Graph: Specification and Algorithms. In: IEEE Symposium on Visual Languages (VL) (1994), `http://www.cs.brown.edu/people/rt/gd-tutorial.html`
10. Cruz, I.F., Xiao, H.: The Role of Ontologies in Data Integration. Journal of Engineering Intelligent Systems 13(4), 245–252 (December 2005)
11. Di Battista, G., Eades, P., Tamassia, R., Tollis, I.G.: Graph Drawing: Algorithms for Geometric Representation of Graphs. Prentice Hall (1999)
12. El Jerroudi, Z., Ziegler, J.: iMERGE: Interactive Ontology Merging (poster). In: International Conference on Knowledge Engineering and Knowledge Management Knowledge Patterns (EKAW) (2008)
13. Euzenat, J., Ferrara, A., Hollink, L., Isaac, A., Joslyn, C., Malaisé, V., Meilicke, C., Nikolov, A., Pane, J., Sabou, M., Scharffe, F., Shvaiko, P., Spiliopoulos, V., Stuckenschmidt, H., Sváb-Zamazal, O., Svátek, V., dos Santos, C.T., Vouros, G.A., Wang, S.: Results of the Ontology Alignment Evaluation Initiative 2009. In: ISWC International Workshop on Ontology Matching (OM). CEUR Workshop Proceedings, vol. 551, pp. 73–126 (2009)
14. Falconer, S.M., Storey, M.A.D.: A Cognitive Support Framework for Ontology Mapping. In: International Semantic Web Conference/Asian Semantic Web Conference (ISWC/ASWC). Lecture Notes in Computer Science, vol. 4825, pp. 114–127. Springer (2007)
15. Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M.: The AgreementMakerLight Ontology Matching System. In: International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE). pp. 527–541. Springer (2013)
16. Ivanova, V., Lambrix, P.: User Involvement for Large-Scale Ontology Alignment. In: International Workshop on Visualizations and User Interfaces for Knowledge Engineering and Linked Data Analytics (2014)
17. Joslyn, C., Donaldson, A., Paulson, P.: Evaluating the Structural Quality of Semantic Hierarchy Alignments. In: International Semantic Web Conference (Posters & Demos) (2008)
18. Lanzenberger, M., Sampson, J.: AlViz - A Tool for Visual Ontology Alignment. In: Conference on Information Visualization (IV). pp. 430–440. IEEE Computer Society (2006)
19. Li, Y., Stroe, C., Cruz, I.F.: Interactive Visualization of Large Ontology Matching Results. In: ISWC International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data (Voila!) (2015)
20. North, C., Shneiderman, B.: Snap-together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata. In: Advanced Visual Interfaces (AVI). pp. 128–135 (2000)
21. Pesquita, C., Faria, D., Santos, E., Neefs, J.M., Couto, F.M.: Towards Visualizing the Alignment of Large Biomedical Ontologies. In: Data Integration in the Life Sciences (DILS). pp. 104–111 (2014)