

# Web-based Ontology Alignment with the GeneTegra Alignment tool

Nemanja Stojanovic, Ray M. Bradley, Sean Wilkinson, Mansur Kabuka,  
and E. Patrick Shironoshita

INFOTECH Soft, Inc.

1201 Brickell Avenue, Suite 220

Miami, Florida 33131, USA

*[nemanja,rbradley,sean,kabuka,patrick]@infotechsoft.com*

## Abstract

Ontologies are increasingly gaining practical usage for semantic data in various ways and across multiple domains. From this growing applicability arises an ever-greater need to manage large datasets, reduce analytical complexity and efficiently as well as accurately integrate different heterogeneous ontologies into or within existing systems, all while minimizing data corruption and maintaining existing semantics. In this paper, we present the GeneTegra Alignment Tool (GT-Align), a practical implementation of the ASMOV ontology alignment algorithm within a Web-based interface, focusing on biomedical data and using Unified Medical Language System (UMLS) for the background knowledge. GT-Align allows iterative alignment of multiple ontologies as well as active user involvement throughout the process.

## 1 Introduction

Ontologies have been increasingly acknowledged as an appropriate abstraction instruments for representing entities and their relationships within various domains (Euzenat and Shvaiko, 2007). Due to this abstract expressiveness, ontologies have been proven to have a highly extensible applicability spectrum, allowing a greater variety of systems to incorporate them in their modeling (Kalfoglou and Schorlemmer, 2003; Noy, 2004;). Because of this increasing development, the need for flexible tools enabling semantic matching of heterogeneous ontologies is becoming much more apparent (Shvaiko and Euzenat, 2013).

In this paper, we present a demonstration of an ontology alignment Web interface called GT-Align, consisting of a server implementation that wraps an ontology alignment algorithm and exposes REST API endpoints which the client side

user interface employs to enable iterative alignment of biomedical ontologies.

Our solution to the problem of ontology alignment is twofold. First, we use an ontology alignment algorithm to identify shared relationships between heterogeneous entities and generate a set of suggested mappings. Second, we allow the user to engage with the results on each iteration by accepting, rejecting or clearing (reverting an acceptance or a rejection) mappings or mapping groups. The interface also allows the user to upload a set of equivalence mappings as an input partial alignment to bootstrap a new alignment process. All mappings must be positively accepted; in other words, no mappings are deemed accepted until positively indicated as such by the user. Rejection of a mapping is an indication that such a mapping should never happen. Clearing of a mapping, on the other hand, indicates that it is not accepted but still possible. The GT-Align Web Interface supports manual evaluation of results through visual inspection, including inspection of parents and children of elements as well as inspection of labels and other textual information. The tool also provides information on the confidence of a mapping as calculated by the underlying ASMOV algorithm (Jean-Mary et al., 2009; Jean-Mary and Kabuka, 2014), and where applicable it also provides reference to codes in the Unified Medical Language System (UMLS) to which concepts are tagged. For algorithm details and the explanation on UMLS usage, see Algorithm section.

The main goal of GT-Align is to enable easier ontology alignment of biomedical data and allow domain experts to validate the results and thus ensure high quality alignments. Put succinctly, GT-Align enables the production of an alignment between any two biomedical ontologies, allowing users to review and revise mappings interactively.

## 2 Algorithm

The underlying alignment algorithm for GT-Align is called ASMOV, which was developed for use in the integration of data and ontologies in the biomedical and life sciences domain within the GeneTegra Information System ([www.genetegra.com](http://www.genetegra.com)). The algorithm makes use of an iterative approach with similarity calculations along multiple dimensions coupled with a process of semantic verification that seeks to remove mapping inconsistencies. ASMOV uses a combination of string-, constraint-, formal-resource-, graph-, model-, and instance-based matching mechanisms. ASMOV has participated in several rounds of the evaluations performed by the Ontology Alignment Evaluation Initiative (OAEI), placing as one of the top three performers in the benchmark tests of the contests in which it has participated (Jean-Mary and Kabuka, 2007; Jean-Mary and Kabuka, 2008; Jean-Mary et al., 2009; Jean-Mary et al., 2009). The ASMOV algorithm uses UMLS as an underlying vocabulary aimed at improving lexical matching between source and target entities. The interface enables the user to turn off this feature, in which case lexical matching is based on Levenshtein edit distance. Prior evaluations of the algorithm showed that the use of an underlying vocabulary significantly improves the quality of mapping, while also reducing the time needed for completion of the alignment process (Jean-Mary and Kabuka, 2007).

## 3 User Interface

Visualization types that have been shown as most effective at enabling user involvement in an alignment process are tree and graph structures, with both having specific benefits to the user (Bo Fu, 2013). Furthermore, Granitzer et al. (2010) have shown that an intelligent combination of both structures is present in many advanced alignment visualization tools. By combining list, tree and graph visualizations to present alignment data to the user at distinct levels of abstraction, GT-Align can yield a more productive alignment through user feedback. The user can explore detailed information on individual concepts as well as parameters of mapping candidates such as estimated confidence and status. Furthermore, the user can filter this data by ontological sub regions or by individual mapping features.

Additionally, the GT-Align user interface is built on modern web technologies including JavaScript/HTML/CSS as well as SVGs for data visualizations, enabling GT-Align to stay on the cutting edge of UI tools (Li et al., 2015). In addition to a wide platform support (including mobile), web technologies maintain consistent high quality of UI capabilities via frequent improvements. This enables GT-Align to be easily deployed into any environment as well as quickly updated with latest technological advances at a minimum expense to the user.

The following sections focus on the individual visualizations and capabilities of the different views present in the GT-Align Web Interface.

### 3.1 Ontology Import View

This view allows the user to upload ontologies into the system, which they can further inspect in the Hierarchical Tree View. The ontology import process uses an extensible set of rules to normalize lexical labels used within it, marking one as the preferred label and others as alternative labels. These labels are then annotated to concepts within the UMLS Metathesaurus. Having normalized labels provides a consistent visual identification scheme that is more easily recognized and thus friendlier to the user.

### 3.2 Hierarchical Tree View

This part of the system displays ontologies as a hierarchical tree of concepts. The ontology tree visualization serves as the fundamental visualization in GT-Align. The view is shown in Figure 1.

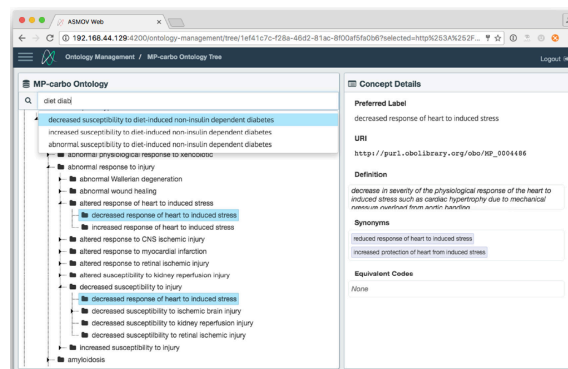


Figure 1: Hierarchical Tree View

When concepts are asserted as children of multiple parents, they are displayed within each parent, as is standard practice. Metadata about the ontology, including its ID and any annotations such as

textual descriptions, are displayed on an information pane. The system also utilizes an auto-complete search for individual concepts within the ontology. The indented tree visualization presents information in a commonly used abstraction allowing users to explore ontologies without a specialized knowledge of the visualization itself. This enables anybody familiar with the concept of an ontology to get started with the software very quickly. The usefulness of an a tree visualization in displaying hierarchical relationships has been demonstrated by its long-term usage in many areas. From visualizing file systems or HTML structures to visualizing ontologies in tools such as WebProtégé (Tudorache et al., 2013), an indented tree visualization is familiar to most, enabling quicker onboarding into the GT-Align system.

### 3.3 Alignment Execution View

Through this view, the user can execute an alignment with specific parameters. The user starts by selecting two ontologies that will be aligned. Each of the ontologies can either be selected from the set of ontologies that were previously added or may be uploaded by the system. Additionally, the user can upload a partial alignment as an input parameter to the alignment process.

### 3.4 Alignment Selection View

This view provides a tabular summary of all the alignments executed in the system. It contains the historical overview the alignment processes ran by the user along with details about each process. An example of this view is shown in Figure 2.

Source #	Target #	Alignment #	State #	Created #	Execution Time #
Bibliographic references	Bibliographic references	RI	Finished	8 days ago	00:00:07
Bibliographic references	Bibliographic references	Not finished	Submitted	8 days ago	-
HP-carbo	MP-carbo	HMP6-3	Finished	12 days ago	00:07:02
HP-carbo	MP-carbo	HMP6-2	Finished	12 days ago	00:07:01
HP-carbo	MP-carbo	HMP6	Finished	13 days ago	00:09:12
HP-carbo	MP-carbo	HMP5	Finished	13 days ago	00:05:24
HP-carbo	MP-carbo	-	Failed	13 days ago	-
HP-carbo	MP-carbo	HMP4	Finished	13 days ago	00:14:51
HP-carbo	MP-carbo	-	Failed	13 days ago	-
HP-carbo	MP-carbo	HMP3	Finished	13 days ago	00:12:02
HP-carbo	MP-carbo	HMP3	Finished	13 days ago	00:09:07
HP-carbo	MP-carbo	HMP2	Finished	13 days ago	00:12:32
Bibliographic	Bibliographic	TS.2	Finished	13 days ago	00:00:21

Figure 2: Alignment Selection View

Clicking on a specific column heading will sort the table according to the corresponding parameter. The displayed parameters include links to the source and target ontologies used in the alignment, the custom name supplied by the user along with the date and

time for when the alignment was created or last modified and how long it took to execute. The view also displays the current execution status which gets updated as an alignment progresses through different stages. Alignments can execute asynchronously within the GT-Align platform, allowing the user to perform tasks in parallel. The extensive computational work is offloaded to the server and doesn't hinder the user experience. Once an alignment is completed, the user receives a notification of its final status. Selection of an alignment transitions the user to the Alignment Overview.

### 3.5 Alignment Overview

After an alignment is obtained, the mappings are presented in an overview pane with a circular graph layout. Due to the structure of ontologies, a graph-based visualization is a natural fit for displaying their alignment. Unlike indented trees, graphs are more suitable to display multiple inheritance without any visual redundancy. This prevents the user of potentially needing to make additional efforts when understanding the data at hand or being confused by concept repetition. Tree visualization is particularly less adequate when displaying large ontologies because the expansion of nodes to greater depths can quickly become overwhelming. Large trees also make it difficult to access the overall structure of an ontology. Using a graph visualization allows us to handle large amounts of data in a way that is more customizable and flexible. The graph visualization is shown in Figure 3.

Concepts from both ontologies are distinguished in the graph by color and positioning. They are separated based on their originating ontology where the concepts from the source ontology are placed on one side and the concepts from the target ontology on another. The user can rotate the graph as they please. Further clustering of concepts is performed based on their hierarchical position in the ontology. The closer a concept is to the root, the closer it is to the center of the graph. Conversely, the outer section of the graph represents the leaf nodes. This design allows the user to easily assess the structure of both ontologies while performing minimal work.

Two concepts connected with a line represent a single mapping. Thickness of this line corresponds to the confidence value, i.e. the level of confidence in the mapping being correct, according to the underlying algorithm. The thicker the li-

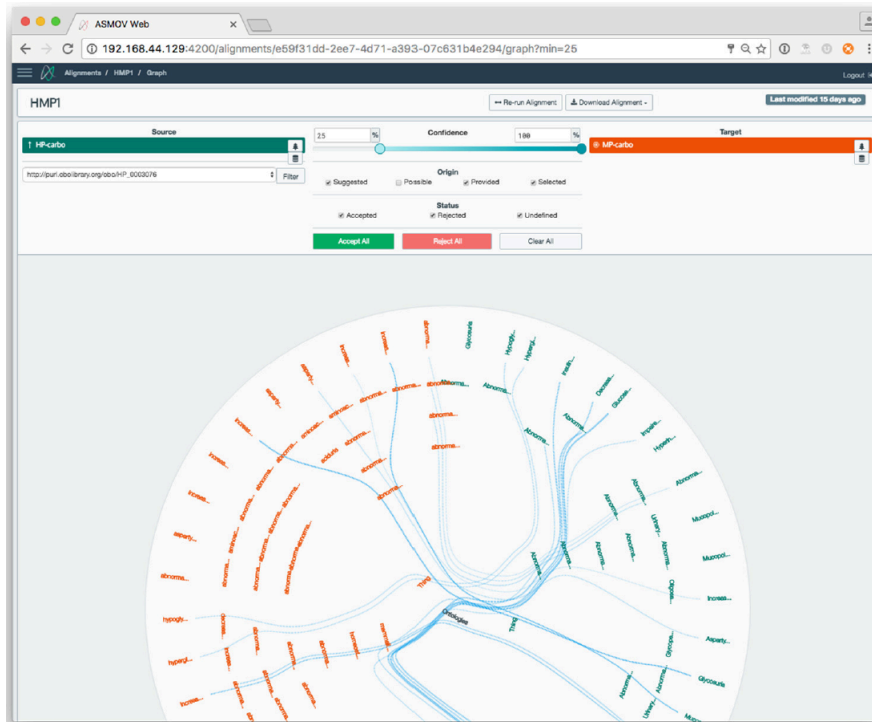


Figure 3: Alignment Overview

ne the higher the generated mapping confidence. This lets the user evaluate a section of the alignment by the amount of high- or low- confidence mappings it contains.

The user can click on a specific mapping or group of mappings under a common parent. Selection of a group of mappings transitions to the Mapping Group View, while selection of any individual mappings transitions to the Mapping View. At the top is a toolbar that provides control to filter the mappings by the confidence value. Additionally, the mappings can be filtered by the mapping origin (suggested by algorithm, found by algorithm, provided by partial alignment), mapping state (accepted by user, rejected by user, undefined), and a branch in the ontology. This filtering is automatically reflected in the graph, allowing users to quickly see the alignment overview at different scales. Besides filtering, the toolbar allows for bulk editing of mappings, enabling the user to accept, reject or clear mappings for large sections of the alignment. The toolbar additionally allows the user to export the mappings through the Alignment RDF format and the EDOAL format, as well as a merged OWL ontology.

### 3.6 Mapping Group View

This view, shown in Figure 4, displays the suggested mappings for a concept and its children. The main purpose of this view is to allow the user

to examine a group of mappings separately from the alignment as a whole. The mappings are shown as a vertical list of concept pairs, giving the user an alternative presentation to the graph that is familiar and straightforward.

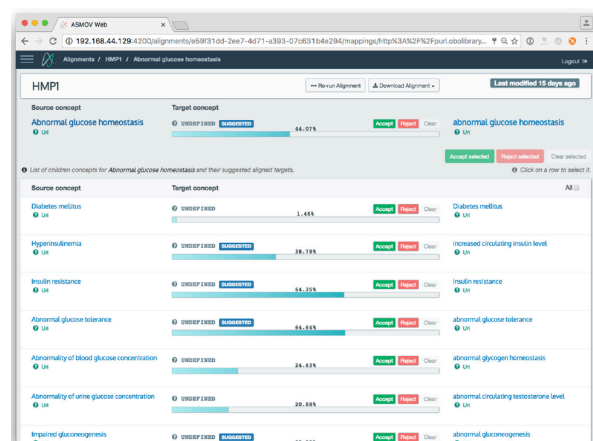


Figure 4: Mapping Group View

Each concept pair contains a rectangular visualization of their mapping confidence on a scale of 0-100. The mapping state is show above each of the confidence visualizations. Control buttons are provided allowing the user to alter the mapping state of the whole group as well as of individual mappings within it. Selection of an individual mapping transitions to the Mapping View section.



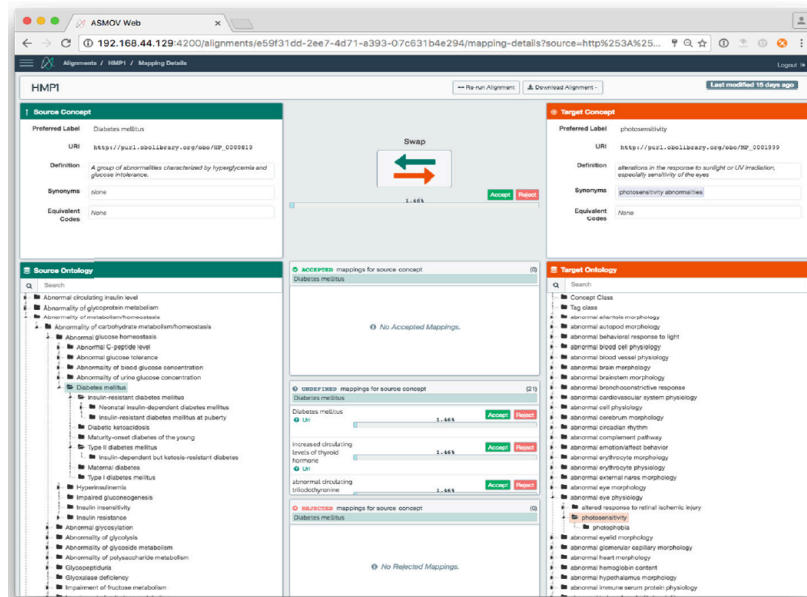


Figure 5: Mapping View

### 3.7 Mapping View

This multi-pane view, shown in Figure 5, provides an in-depth visualization of a single mapping, allowing the user to review or modify the mapping.

**Top left pane:** This pane contains information about the currently focused source concept selected by the user, including a preferred label, alternative labels, and annotated UMLS codes if available. All other panes display information in relation to this currently focused concept.

**Top right pane:** This pane contains a selected target concept for mapping to the currently focused concept. In the center between the top right and left panes are controls allowing the user to accept or reject the mapping between the focused and selected concept, or create a new one if none yet exists. The controls also include an option for the user to swap the focused and selected concepts, causing other panes to adjust accordingly. Three central panes under the controls show one or more mappings in different states.

**Bottom center panes:** The top pane shows the accepted mapping for the focused source concept, if one exists. The middle central pane shows a list of possible mappings according to ASMOV. The top concept in this list is the mapping suggested by ASMOV, other mappings are alternative possibilities. The bottom central pane shows a list of rejected mappings, if any exist. In all three panes, mappings show their preferred label and mapping confidence value. Clicking on a concept in any of

these panes places it on the selected mapping pane, making it the new focused concept.

**Bottom side panes:** Two bottom panes on each side contain the hierarchical tree view of both ontologies (see Hierarchical Tree View). On the bottom left is the ontology tree corresponding to the focused source concept, and on the bottom right is the ontology tree corresponding to the selected target concept. Each ontology tree highlights the selection of the corresponding concepts. All ancestors of both concepts are shown in the respective tree view along with all siblings of each ancestor, but children of ancestor siblings are initially hidden although the user can explore them if desired. Clicking on a concept within the left ontology tree will set it as the currently focused concept, updating all other panes accordingly. Clicking on any concept from the right ontology tree places it on the selected concept pane, choosing it as the mapping for the source concept.

## 4 Future Work

Previous sections highlight the current status and main features of the GT-Align Web Interface. There are several planned features expected to be produced in the future. Subsequent releases will allow users to specify subsumption relationships in addition to equivalence relationships. GT-Align will include the capability of performing an automated evaluation of precision and recall against a reference alignment, and to display the results of such evaluation. Additionally, it will incorporate

functionality to dynamically change the set of fixed weights for the various similarity values calculated by ASMOV. It will also present the separate confidence scores for the different measures of similarity generated by ASMOV. GT-Align will support pivot systems such as EDOAL to allow importing of alignments from other systems. Finally, GT-Align will support interactive collaboration by multiple users.

## 5 Conclusion

In this paper, we presented GT-Align, a versatile implementation of an ontology alignment Web interface, backed by an efficient, robust and field-tested algorithm called ASMOV. Besides the algorithmic alignment, the interface enables iterative user involvement, allowing domain experts to improve and validate the results thus contributing to the quality of the alignment. We showed features to evaluate relationships between entities in different biomedical ontologies as well as explore ontologies on their own through hierarchical trees. The interface enables users to analyze aligned mappings from a high-level perspective through groups refined by optional user filters and combined with algorithm results. It further provides capabilities for fine grain inspection of individual concepts through their mappings as well as relations to other concepts. An assortment of visualizations provided by the user-interface enables multiple perspectives on the data itself along with the alignment results. Additionally, we presented our plans for future development. In summary, GT-Align is a robust and easy to use solution for ontology alignment of biomedical data.

## Acknowledgements

This work is supported by grant R44GM097851 from the National Institute of General Medical Sciences (NIGMS), part of the U.S. National Institutes of Health (NIH).

## References

Euzenat, Jérôme, and Pavel Shvaiko. *Ontology matching*. Vol. 18. Heidelberg: Springer, 2007.

Fu, Bo, Natalya F. Noy, and Margaret-Anne Storey. "Indented tree or graph? A usability study of ontology visualization techniques in the context of class mapping evaluation." In *International Semantic Web Conference*, pp. 117-134. Springer, Berlin, Heidelberg, 2013.

Granitzer, Michael, Vedran Sabol, Kow Weng Onn, Dickson Lukose, and Klaus Tochtermann. "Ontology alignment—a survey with focus on visually supported semi-automatic techniques." *Future Internet* 2, no. 3 (2010): 238-258.

Jean-Mary, Yves R., and Mansur R. Kabuka. "ASMOV results for OAEI 2007." In *Proceedings of the 2nd International Conference on Ontology Matching-Volume 304*, pp. 150-159. CEUR-WS.org, 2007.

Jean-Mary, Yves R., and Mansur R. Kabuka. "ASMOV: results for OAEI 2008" In *Proceedings of the 3rd International Workshop on Ontology Matching-Volume 431*, pp. 132-139. CEUR-WS.org, 2008.

Jean-Mary, Yves R., E. Patrick Shironoshita, and Mansur R. Kabuka. "ASMOV: Results for OAEI 2009" In *Proceedings of the 4th International Workshop on Ontology Matching-Volume 551*, pp. 152-159. CEUR-WS.org, 2009.

Jean-Mary, Yves R., E. Patrick Shironoshita, and Mansur R. Kabuka. "ASMOV: Results for OAEI 2010" In *Proceedings of the 5th International Workshop on Ontology Matching-Volume 689*, pp. 126-133. CEUR-WS.org, 2010.

Jean-Mary, Yves R., E. Patrick Shironoshita, and Mansur R. Kabuka. "Ontology matching with semantic verification." *Web Semantics: Science, Services and Agents on the World Wide Web* 7, no. 3 (2009): 235-251.

Jean-Mary, Yves R., and Kabuka, Mansur. "Ontology alignment with semantic validation." U.S. Patent 8,738,636, issued May 27, 2014.

Kalfoglou, Yannis, and Marco Schorlemmer. "Ontology mapping: the state of the art." *The knowledge engineering review* 18, no. 01 (2003): 1-31.

Li, Yiting, Cosmin Stroe, and Isabel F. Cruz. "Interactive Visualization of Large Ontology Matching Results." In *VOILA@ ISWC*, p. 37. 2015.

Noy, Natalya F. "Semantic integration: a survey of ontology-based approaches." *ACM Sigmod Record* 33, no. 4 (2004): 65-70.

Shvaiko, Pavel, and Jérôme Euzenat. "Ontology matching: state of the art and future challenges." *IEEE Transactions on knowledge and data engineering* 25, no. 1 (2013): 158-176.

Tudorache, Tania, Csongor Nyulas, Natalya F. Noy, and Mark A. Musen. "WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web." *Semantic web* 4, no. 1 (2013): 89-99.