

Introducing the new SIM-DLA Semantic Similarity Measurement Plug-in for the Protégé Ontology Editor

Christoph Mülligann
University of Münster,
Germany
cmuelligann@uni-muenster.de

Johannes Trame
University of Münster,
Germany
johannes.trame@uni-muenster.de

Krzysztof Janowicz
University of California, Santa
Barbara, USA
jano@geog.ucsb.edu

ABSTRACT

Semantic similarity measurement has been an active research area in GIScience and the Semantic Web for many years. However, implementations of these measures were largely missing, not publicly available, or tailored to specific application needs. To foster the application of similarity reasoning in information retrieval, ontology engineering, and spatial decision support, we implemented the SIM-DL semantic similarity server as well as a plug-in for the popular Protégé ontology editor. While SIM-DL has been successfully applied to several application areas, the implemented similarity theory was largely structural, could not handle concept and instance similarity within the same framework, and was based on a Protégé version and DIG interface that have been re-engineered over the last years. This paper introduces a new version, called SIM-DL_A, engineered from scratch to address these shortcomings. It is based on our new similarity theory, can handle inter-instance and inter-concept similarity using the same functions and alignments, and is available for the new Protégé version 4.1.

Categories and Subject Descriptors

H.2.8 [Database applications]: Spatial databases and GIS; H.3.3 [Information Search and Retrieval]: Retrieval models; I.2.4 [Knowledge Representation Formalisms and Methods]

General Terms

Measurement

Keywords

Semantic Similarity, Geospatial Semantics, Ontology Engineering, Context, Description Logics

1. INTRODUCTION

Similarity is of fundamental importance for information retrieval, e.g., for search engines on the Web. In contrast to syntactic or statistic measures, semantic similarity compares the meaning of concepts and has been studied to understand cognition and categorization in cognitive science and artificial intelligence research. In the case of the Geospatial Semantic Web, similarity is used for semantic-based geographic information retrieval [26, 13, 16], ontology matching and alignment [6, 28], concept learning [7], as well as during ontology engineering [10, 15]. The fact that most geographic information analysis, e.g. interpolation, kernel methods, or point pattern analysis, are based on spatial auto-correlation and spatial distance, shows why semantic similarity, as analogy to spatial distance, is considered crucial for making ontologies and semantics first class citizens of GIS and geo-statistics.

While numerous similarity theories have been defined over the last years, none of them was specifically tailored for reasoning over description logics-based ontologies, implemented as free and open source code, and integrated with major Semantic Web protocols and tools. For this reason, we developed the SIM-DL theory together with a similarity server and a plug-in to access the server's functionality from within the popular Protégé ontology editor. The server also implemented the DIG protocol [3] to communicate with other Semantic Web tools or to be integrated in Web-based applications. We have applied the SIM-DL server to several application areas such as information retrieval, spatial decision support, to address semantic heterogeneities between information communities, and ontology engineering. The server was also used by other researchers, e.g., to integrate geospatial information [19].

Nevertheless, SIM-DL had several shortcomings. As a structural measure, and similar to structural subsumption reasoning, the canonization of compared-to concepts played a crucial role and rewriting rules had to be applied before comparison. This approach can be used for various smaller description logics, but fails in case of logics such as *SHOIN*(\mathcal{D}) underlying OWL-DL. Additionally, SIM-DL was restricted to inter-concept similarity and singleton classes had to be defined to include instance similarity. Finally, the last version of SIM-DL was published in 2008 and was not compatible with new versions of Protégé; the last supported version was 3.3.

In 2009, we have introduced a new similarity theory, called SIM-DL_A [17], that can be used to reason on concepts defined using more expressive description logics. SIM-DL_A reduces the problem of inter-concept similarity to inter-instance similarity and is based on a modified tableau algorithm. In contrast to the structural SIM-DL measure, SIM-DL_A creates possible models of a concept, called proxy models, and compares them to the proxy models of other concepts. Those models can be thought of as representative individuals of the given concept, but, in fact, the same theory can handle any individuals stored in the ABox.

In this paper, we present the first implementation of SIM-DL_A as plug-in for the new Protégé 4.1 ontology editor. While SIM-DL_A is not restricted to geo-ontologies, we decided to use an extended version of the hydrology examples that were introduced to evaluate the cognitive plausibility of the original SIM-DL. Similarity also plays an increasing role for Linked Data and SIM-DL_A's ability to combine instance and concept similarity may be useful for work such as the similarity ontology proposed by Halpin et al. [12]. In the following, we briefly introduce related research to situate our work in the literature, highlight the major aspects of the SIM-DL_A theory, explain the implementation, and discuss application examples. While the theory and implementation are still under development, a fully operational version of the plug-in is available as free and open source software at <http://sim-dl.sourceforge.net>.

2. RELATED WORK

Due to its analogy to distance in space, semantic similarity has been an active research topic in GIScience for many years. Starting with the work of Bruns and Egenhofer [5], researchers have investigated the similarity of spatial scenes [20, 22], introduced measures to compare concepts specified in feature-based ontologies [26], or developed measures and applications that propose a geometric view on semantic similarity measurement [25, 1, 27], mostly following Gärdenfors' work on conceptual spaces [11]. ConceptVISTA was developed as a knowledge acquisition and management tool which used semantic similarity for knowledge integration [10].

Besides GIScience, semantic similarity measurement also plays an important role in bioinformatics, semantic-enabled information retrieval [18], and ontology matching [28]. A similarity library, called Simpack, was developed by Bernstein et al. [4], while Cruz and Sunna implemented a set of similarity functions for the alignment of geo-ontologies [6]. Among many others, similarity measures tailored to the needs of Web ontologies were also proposed by d'Amato et al. [7] and Stuckenschmidt [29].

3. SIMILARITY FRAMEWORK

When comparing classes or individuals represented using description logics (DL), and in contrast to the original SIM-DL (cp. [14]), SIM-DL_A avoids the problem of canonization by creating proxy models. Proxy models represent satisfiable instantiations of a DL class. For example, a concept *EarthSurface* equivalent to $WaterSurface \sqcup LandSurface$ would be mapped to the assertions $\{WaterSurface(p1), LandSurface(p2),$

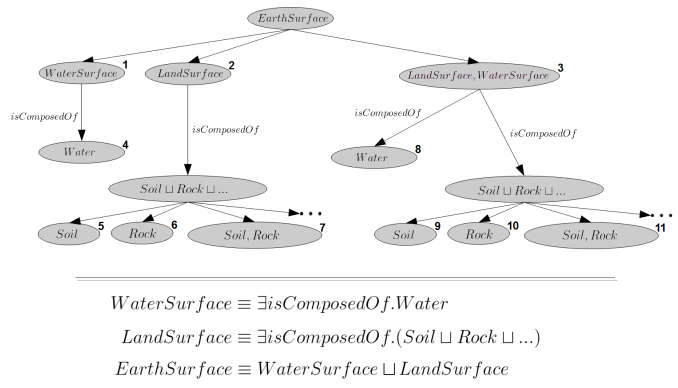


Figure 1: Completion graph for the simplified *EarthSurface* concept.

$Watersurface(p3), LandSurface(p3)\}$.¹ In other words, while most DL reasoners use a tableau algorithm to detect clashes during satisfiability checking, i.e., they try to create a contradiction-free model, SIM-DL_A uses a modified tableau algorithm to create a set of models that are representative for a given class.

Figure 1 shows the completion graph for a simplified *EarthSurface* class. The numbered nodes are proxy models of *EarthSurface*, while the other nodes are artifacts of the tableau algorithm. Proxy models may be connected via binary relations. For instance, model 2 has *isComposedOf* relations to the three proxy models 5, 6 and 7. These are, in fact, mere role fillers of model 2. In contrast, the proxy models 1, 2, and 3 would be used for comparison against proxy models derived from another class; in case of matching or similar roles, their fillers would be in turn compared to those of the *EarthSurface* proxies.

Each search and target concept generates different proxy models, computed by the tableau algorithm. Those proxy models are aligned by the similarity framework within a so called *Model Level Matrix*. Depending on the similarity application mode [16] the proxy models are selected for comparison and the average, minimal, or maximum similarity is computed. While the basic similarity functions from the SIM-DL theory [14] are used, the alignment differs. By comparing each proxy model resulting from the search concept with each proxy model of the target concept, an *Assertion Matrix* is created for computing the similarity between any pair combination of proxy models. All similarity values in the *Assertion Matrix* are computed by the similarity functions for primitives, i.e. base symbols, and (primitive) roles as all complex expressions have been unfolded during the tableau algorithm. The overall similarity between two proxy models depends on whether a non-symmetric (e.g., for information retrieval) or symmetric (e.g., for ontology alignment) measure is chosen [26, 16].

¹Depending on the similarity mode, e.g., *non-symmetric minimal similarity*, the assertions would have to be $\{WaterSurface(p1), \neg LandSurface(p1), LandSurface(p2), \neg WaterSurface(p2), Watersurface(p3), LandSurface(p3)\}$. A detailed discussion of the semantics of similarity and differences between modes is given in [16].

4. IMPLEMENTATION

The SIM-DL_A framework has been implemented in Java and extends the OWLReasoner interface, specified by the OWL-API (for OWL 2.0)². We extended the interface by a method for computing similarity among concepts and/or individuals. For reasons of performance standard methods, for instance, such as satisfiability checking are delegated by default to the the Pellet OWL2 reasoner³ or any other delegated OWL-Reasoner of choice. The similarity method provides access to the novel implementation of the SIM-DL_A algorithm which mainly consists of the modified tableau algorithm and the alignment framework (section 4.3). While implementing the OWLReasoner interface, the SIM-DL_A framework can easily be invoked as a OWLReasoner in any application built upon the OWL API. Furthermore, the OWL-API is capable of integrating description logics and Linked Data through RDF-OWL mappings that are performed in both directions.

As in a previous release, the SIM-DL_A framework has been integrated and deployed as a plug-in for the Protégé ontology editor and knowledge acquisition system⁴ (section 4.4). Hence, it can easily be used and integrated in to tasks such as ontology engineering [15] as well as for populating and maintaining instance-level data.

In this section we highlight selected implementation details that are relevant for the understanding of the similarity engine and functionality offered by the plug-in. The SIM-DL_A theory has been explained in [17]; however, we do introduce some minor modifications here to address previous shortcomings.

4.1 Search and Target Entities

While SIM-DL was limited to inter-concept similarity, SIM-DL_A can also handle inter-instance similarity. This requires some changes to the underlying context theory and the user interface. In case a concept is selected for comparison, it is internally rendered (section 4.2) into a string representation. During the rendering process complex concept definitions are rewritten and unfolded to primitives and logical connectives. The rendered expressions serve as a starting points for the tableau algorithm which produces proxy models. Instead of explicitly specifying target concepts to be compared, a context concept C_C can serve as a input to restrict the application domain; see [16, 26] for details. In this case, the search concept will be compared against all concepts subsumed by the context concept C_C . In case of inter-instance similarity, instances (as sets of assertions) are directly mapped to a proxy model representation, due to the fact that an individual already embodies one possible instantiation of a known or unknown class. During the alignment (section 4.3) the proxy models are selected for comparison and different similarity functions are applied.

4.2 Rendering

Axioms in the OWL API appear as nested Java objects. The original SIM-DL server was also implemented following rigid principles of object-oriented design. However, due to the high number of nested lists (and deep cloning) necessary

for the tableau implementation, this resulted in a high number of objects and nested loops. For instance, intersection was modeled as a Java class connecting two other expression (as objects), which could be complex again. Hence, each individual intersection object had to be unfolded, i.e., the first and second object had to be accessed and potentially unfolded again. Therefore, we decided to take a radically different solution for the re-implementation and render those objects as strings of terms with a syntax that allows sub-term access without any kind of explicit references. For example, a concept *Channel* that is equivalent to $Navigable \sqcap Watercourse \sqcap \neg Artificial \sqcap \exists meets.Landmass$ would be rendered to `!I Navigable !I Watercourse !I %N Artificial !E meets Landmass`. The internal structure of the *Channel* class is determined by counting one- ('%') and two-valued ('!') operators. By recursively unfolding all concepts down to primitives without subclass and equivalence relations, there is no need for OWL API access during the execution of the tableau algorithm and the number of Java objects is greatly reduced. The only connection to the original OWL representations is a reasoner that is employed by the tableau to query sub- and inverse property axioms, that are not part of the concept definitions as such. Since recursive definitions, i.e., a concept appears in its own (nested) definition, are caught during the rendering process, that case has not to be considered by the blocking mechanism of SIM-DL_A anymore.

When rendering individuals no tableau has to be employed and consequently no string representation has to be created. Instead we directly map to the same proxy model representation that is also the result of the tableau. By doing so, it is possible to compare individuals with concepts and compute what might be considered the *prototypicality* of an individual. For individuals, we render the class it instantiates (and the super classes) and property assertions. In addition we record the individual's URI to check for identity.

4.3 Alignment

The alignment phase compares proxy models, i.e., their assertions, from the source concept to those of the target concepts; or, in case of inter-instance similarity, the compared-to source and target instances. Based on the selected similarity mode, the alignment phase also determines which compared pairs will be used for the final similarity computation and how normalization will be handled. For instance, non-alignable [21] assertions from the target concept do not reduce a maximum, non-symmetric similarity as used in semantic-based information retrieval. More details on the alignment process are given in [17].

4.3.1 Model Level Matrix - Selecting proxy models for overall similarity comparison

The alignment process is twofold. In a first step an empty *Model Level Matrix* M_M is built, where all proxy models of the search entity (columns) are compared against each proxy model of the target entity (rows). In a second step, for each proxy model comparison a *Assertion Matrix* M_A is created to compute the similarity between proxy models. The resulting proxy model similarity will be entered into the *Model Level Matrix*. Depending on the application context, the user can choose whether the average, minimum, or

²<http://owlapi.sourceforge.net/>

³<http://clarkparsia.com/pellet/>

⁴<http://protege.stanford.edu/>

the maximum similarity should be selected for comparison, as the matrix contains all possible similarity combinations of proxy models. For average similarity the mean of each column is used for computing the standardized sum, while for minimum or maximum similarity, the highest (or lowest, respectively) value out of all columns is used.

4.3.2 Assertion Matrix - Computing similarity between two proxy models

Each proxy model consists of a list of primitive concepts and roles. As we map individuals directly to proxy models in the rendering process, we keep their URI for identity checking. During the alignment process, the *Assertion Matrix* is built by comparing all concepts and roles (plus their fillers) of one proxy model P_A against those from proxy model P_B . In case two primitive concepts are compared, a modified version of the Jaccard Similarity Coefficient is applied. It measures the degree of overlap between two sets S_1 and S_2 as ratio of the cardinality of shared members in all complex concept definitions from $S_1 \cup S_2$ to the cardinality retrieved from $S_1 \cap S_2$. In other words, the more frequently two primitives co-occur in complex concept definitions, the more similar they are. For formal details we refer to [17]. When comparing roles, also their role fillers are compared. Therefore, two different similarity functions are applied one for roles as well as their fillers. Fillers, such as px in $meets(p1, px)$ are recursively defined and represented again as proxy models, e.g., $Landmass(px)$. For each filler comparison a new *Assertion Matrix* is created. Since a multiplicative approach has been adjudged to be cognitive plausible [16], the overall similarity of two roles-filler pairs is defined as the product of the role similarity and the filler similarity.

Because *SHI* supports role hierarchies but does not support role intersections or compositions, the Jaccard Coefficient cannot be used as a measure of similarity between roles.⁵ As proposed in previous work [17], a network-based approach is used instead. Therefore, the similarity between two roles is defined as the ratio between the shortest path from role R_1 to R_2 and the maximum path within the graph representation of the role hierarchy; if the path leads over the universal role U , similarity is set to 0 [16]. *SIM-DL_A* uses a weighted version of Rada’s network similarity [24] by which roles that are on a deeper level of the role hierarchy are more similar to each other; details are given in [17]. In case of roles from temporal [9] or conceptual neighborhood graphs [5], the hierarchy of the respectively graph is used (without a weighting).

4.4 SIM-DL_A Protégé Plug-in

Protégé is the most common graphical user interface for the construction of Web ontologies and can be extended by various plug-ins. The old *SIM-DL* Protégé plug-in has been re-engineered to support the latest version of Protégé. While Protégé 3.x did not provide native support for OWL (only via a plug-in), Protégé 4.1 is fully built upon the OWL-API following the OWL2 specification. Besides some minor changes, most of the interface elements and options of the

⁵The current *SIM-DL_A* theory and implementation have been developed for the *SHI* description logics but can be extended by qualified number restrictions and other language features in the future.

initial *SIM-DL* plug-in have been transferred to the new infrastructure. In Protégé 4.1, all UI elements are fully customizable and different views can be combined in arbitrary ways.

Therefore, the *SIM-DL_A* plug-in has been implemented as a view and can be combined with the class hierarchy view (Fig.2-1) or an individual view (Fig.2-2). Since the plug-in adapts the new look & feel provided by Protégé 4.x, the user can simply define the search entity as well as the target entities (or the context concept) using drag and drop. To do so, they can be selected from the class or individual view and dropped in to the according fields (Fig.2-3). The different similarity modes can be selected in the application context panel and the user can choose how the results should be visualized, e.g., as ordered list (Fig.2-4), using a font-size scaling visualization known from tag clouds, or by categorization into n categories. The user can also provide a prediction of expected similarity values. This feature can be used to evaluate the ontology against human similarity ratings and has been applied as a proxy for *fitness for purpose* in ontology engineering [15].

5. EXEMPLARY APPLICATION

At its core *SIM-DL_A* always computes inter-instance similarity. When inter-concept similarity is requested, proxy models are created by the modified tableau algorithm as a prerequisite for similarity measurement. Consequently, the new plug-in supports inter-concept, inter-instance, and even concept-to-instance similarity in different similarity modes and context settings; see [16] for different kinds of contexts and their impact on semantic similarity. We provide a simplified example of inter-concept similarity and its application in subsection 5.1. Subsection 5.2 describes inter-instance similarity and deals with those applications that are enabled by similarity measurement on the assertion level. Note that inter-instance similarity is heavily domain specific. *SIM-DL_A* provides the basic framework for measuring inter-instance similarity but cannot provide specific similarity functions for all different kinds of applications. These additional functions can be added by others. For instance, *edit distance* functions may be included to compare names of water bodies, while other measures may be used to compare pH-levels, and so forth.

5.1 Inter-Concept Similarity

Related work has shown that inter-concept similarity enhances search for spatial entities as well as communication about spatial concepts. Search results cannot only be ranked by spatial but also by semantic distance. For instance, the search for an instance of type *River* might not only return nearby instances of exactly the same type but also instances of type *Stream*, *Channel* or *Canal* (cp. figure 2); see [16] for implementation in to digital gazetteer user interfaces. Inter-concept similarity may also serve as an indicator for agreement or disagreement between domain experts [15]. For the example given in figure 2, it can be assessed from the font sizes of concept names (or directly via the similarity values) which concepts in the *Navigable* context⁶ are similar to

⁶In the presented example, setting the context concept to *Navigable* only compares those concepts to *River* that are subclasses of *Navigable*.

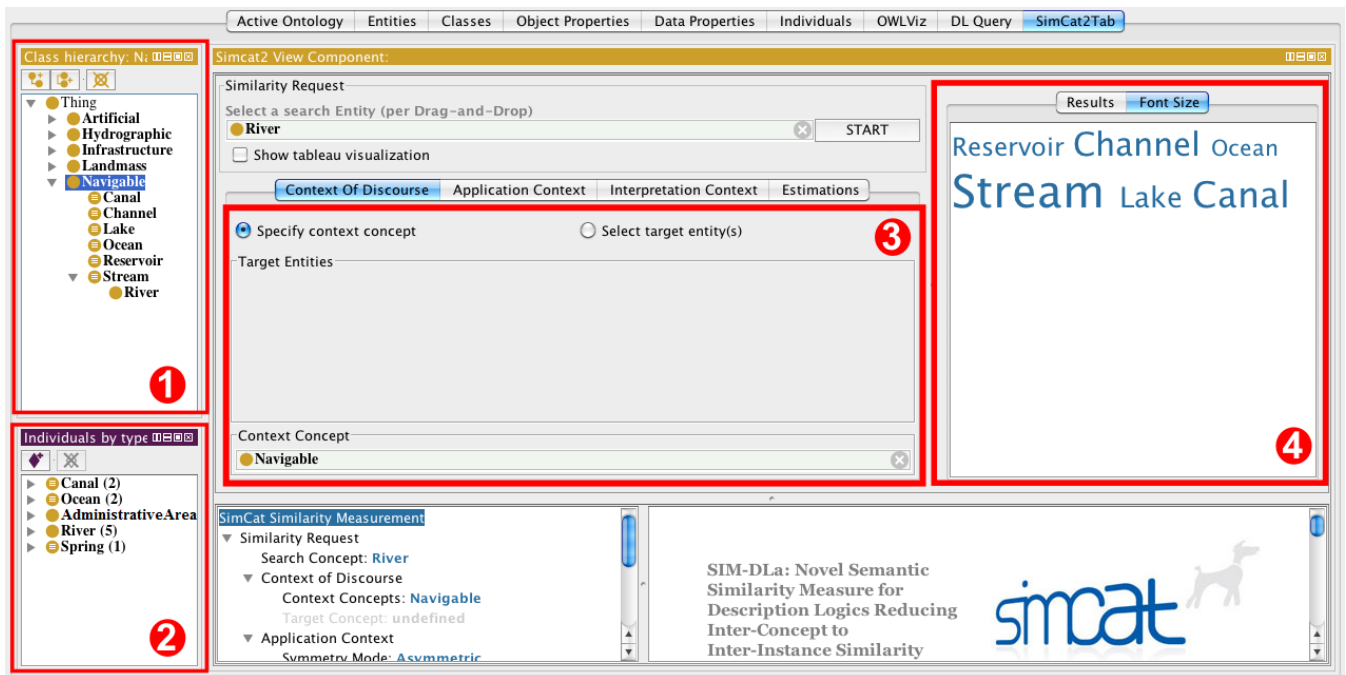


Figure 2: A similarity request from the Protégé SIM-DL_A plug-in for the concept *River* in the *Navigable* context. Results are rendered as a font-size scaling cloud in the upper right corner. The respective similarity values are: *Stream* 88.89, *Canal* 70.24, *Channel* 63.89, *Lake* 50.51, *Reservoir* 41.72, *Ocean* 40.4 and can be accessed using the *Results* tab.

River, that is, *Stream*, *Canal*, *Channel*, and which are rather dissimilar to *River*, such as *Lake*, *Reservoir*, and *Ocean*. Instead of browsing through the ontology to search for the most influential commonality[26] factors, these can be assessed from the visualized similarity ranking. For instance, in figure 2 it is linearity and not the artificial vs. natural distinction that separates *River* from other concepts, e.g., *Lake*. The similarity mode (non-symmetric vs. symmetric) as well as minimum, maximum, and average similarity) can be accessed using the *Application Context* view.

Users interested in the technical aspects of SIM-DL_A might want to examine how and which proxy models were created by the modified tableau algorithm. Figure 3 shows the completion graph for the concept *River* with six proxy models in total. This visualization is also essential for debugging the implementation, testing the theory, and understanding the alignment process. The graph visualization can be activated using the checkbox in the Protégé plug-in; see figure 2.

5.2 Inter-Instance Similarity

Measuring similarity between instances can be done for a multitude of reasons. For instance, it could be used in the Linked Data cloud to relate individuals, or to aggregate Linked Sensor Data [23]. Methods such as string matching have strong limitations due to their disregard of semantics. Finding correspondents for individuals that lack certain properties or identifiers may help to semi-automatically interlink data.

In the simplified example depicted in table 1, an unknown individual (“UnknownRiver”) is compared to a set of existing

Name/ID	Assertions
Ems	River(Ems) inside(Ems,Germany) hasDestination(Ems,NorthSea)
Mosel	River(Mosel) hasDestination(Mosel,Rhine) inside(Mosel,France) inside(Mosel,Germany)
Rhine	River(Rhine) hasDestination(Rhine,NorthSea) inside(Rhine,Netherlands) inside(Rhine,Switzerland) inside(Rhine,Germany)
DortmundEmsCanal	Canal(DortmundEmsCanal) connectedTo(DortmundEmsCanal,Ems) connectedTo(DortmundEmsCanal,RhineHerneCanal)
RhineHerneCanal	Canal(RhineHerneCanal) connectedTo(RhineHerneCanal,Rhine) connectedTo(RhineHerneCanal,DortmundEmsCanal)
NorthSea	Ocean(NorthSea)
France	AdministrativeArea(France)
Netherlands	AdministrativeArea(Netherlands)
Switzerland	AdministrativeArea(Switzerland)
Germany	AdministrativeArea(Germany)
UnknownRiver	River(UnknownRiver) hasDestination(UnknownRiver,UnknownRiver2) inside(UnknownRiver,France) inside(UnknownRiver,Germany)
UnknownRiver2	River(UnknownRiver2) hasDestination(UnknownRiver2,UnknownOcean) inside(UnknownRiver2,Germany)
UnknownOcean	Ocean(UnknownOcean)

Table 1: Assertions in the hydrography use case.

individuals that use a shared vocabulary. To demonstrate how SIM-DL_A can be applied in such a setting, “Unknown-

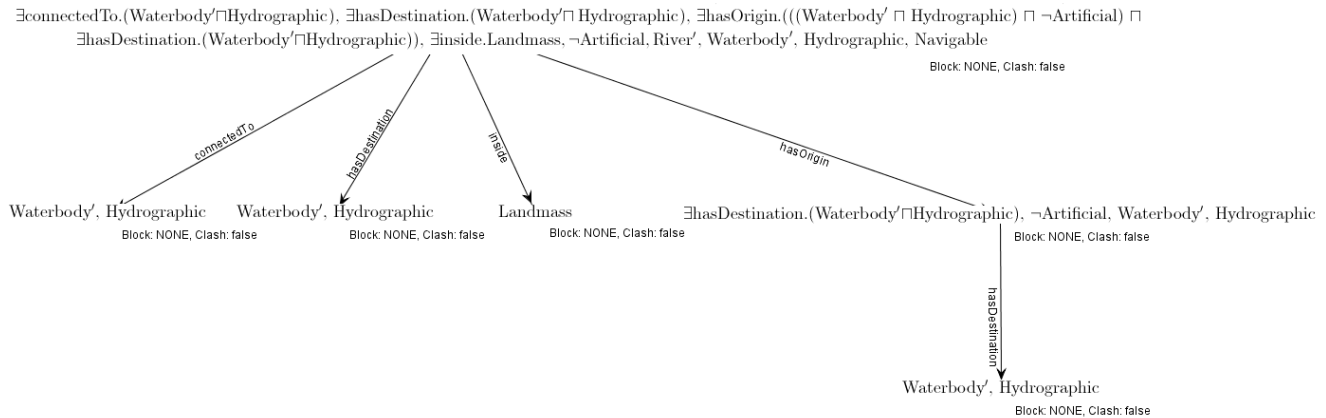


Figure 3: Completion graph of the concept *River*. All equivalent and superclass(es) of *River* are rendered in the top proxy model. Outgoing from this one, four first-level and one second-level proxy model(s) are expanded through existential quantification. Apart from the common DL notation we use the apostrophe to indicate *primitiveness*, i.e., $C \sqsubseteq A$ can be rewritten to $C \equiv A \sqcap C'$.

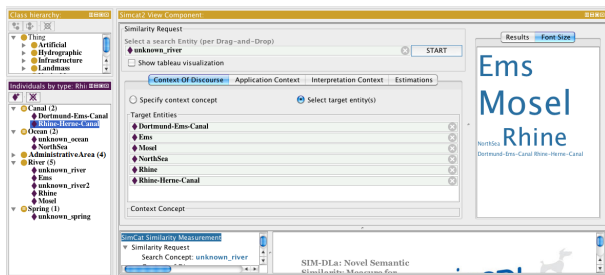


Figure 4: Visualized result set from the hydrography use case in the Protégé SIM-DL_A plug-in. The (symmetric) similarity values corresponding to the font sizes for individual “UnknownRiver” are: “Mosel” 93.33, “Ems” 80, “Rhine” 66.67, “Dortmund-Ems-Canal” 10, “Rhine-Herne-Canal” 6.67, “NorthSea” 0.

“River” is simply a partial copy of the individual “Mosel”. This reflects the situation where two Linked Data sets should be interlinked and may store partial information about the same entities. The results (cp. figure 4) exhibit a strong influence of the geographic feature type on similarity, since *River* instances are ranked the highest among all hydrographic features in scope. Still SIM-DL_A manages to distinguish between instances of the same type (“Mosel”, “Ems”, “Rhine”) by accounting for roles, role filler types, and URIs. As described before, the current implementation does not provide any application specific similarity functions as these have to be added by application developers. Therefore, SIM-DL_A would not be able to compute partial matches for the names of waterbodies, e.g., using edit-distance, nor use spatial distance based on the geometric representations of the features.

6. CONCLUSIONS AND FURTHER WORK

In this paper, we motivated the need for a new semantic similarity reasoner, introduced the SIM-DL_A plug-in for the Protégé ontology editor, explained relevant implementation

details, and highlighted application areas by providing examples. While the server and similarity theory are still under development, they are already more flexible and offer new possibilities compared to the initial SIM-DL similarity server. The presented implementation is operational and can be downloaded and used within Protégé.

SIM-DL_A implements a rudimentary syntactic similarity fallback mode for unknown language constructs. However, this mode can only distinguish between a full match and no match and is used to handle ontologies which use more expressive description logics than *SHL*. For instance, two qualified number restrictions such as $(\geq 1 \text{ hasDestination.Waterbody})$ can be matched but $(\geq 1 \text{ hasDestination.Waterbody})$ and $(\geq 2 \text{ hasDestination.Waterbody})$ would produce no match. The required similarity functions were already introduced for SIM-DL but have not been implemented for SIM-DL_A so far. This would require further modifications to the tableau. For instance, the algorithm should not generate infinite proxy models for a minimum qualified number restrictions, and so forth. While recent research indicates [2] that expressivity is not necessarily the most important criterion and many successful large-scale ontologies rely on languages such as \mathcal{EL} (which can be handled by SIM-DL_A), support for OWL-DL would be a very important step for the acceptance of our work by a larger community.

While similarity is at the core of most work on ontology alignment [8] and also implemented in semantic-based information retrieval interfaces [18], it is still not very popular in the broader Semantic Web community. We believe that Linked Data and an increasing number of small, application-driven ontologies will boost the acceptance and need for similarity reasoning and the provided relevance ranking. By using reification or thresholds, it could be applied to compute the similarity relations proposed by Halpin et al. [12] and, hence, go beyond (mis)using `owl:SameAs`.

The DIG protocol[3] for DL reasoners used by the initial

SIM-DL has been deprecated. Therefore, we plan to release an implementation of the new SIM-DL_A reasoner as a OWLLink server⁷ on top of the OWL API and, thereby, make it usable outside of Protégé.

7. ACKNOWLEDGMENTS

This research has been partially funded by the SimCat II project (DFG JA1709/2-2). The free and open source code for the plug-in, derived applications, documentations, and technical details are available <http://sim-dl.sourceforge.net>.

8. REFERENCES

- [1] O. Ahlqvist and A. Shortridge. Characterizing land cover structure with semantic variograms. In A. Riedl, W. Kainz, and G. Elmes, editors, *Progress in Spatial Data Handling -12th International Symposium on Spatial Data Handling*, pages 401–415. Springer, 2006.
- [2] F. Baader, C. Lutz, and A.-Y. Turhan. Small is again beautiful in description logics. *KI*, 24(1):25–33, 2010.
- [3] S. Bechhofer. The dig description logic interface. In *International Workshop on Description Logics (DL2003)*, Rome, Italy, September 2003.
- [4] A. Bernstein, E. Kaufmann, C. Kiefer, and C. Bürki. Simpact: A generic java library for similarity measures in ontologies. Technical report, Department of Informatics, University of Zurich, 2005.
- [5] H. Bruns and M. Egenhofer. Similarity of spatial scenes. In J. M. Kraak and M. Moleenar, editors, *Proceedings of the 7th International Symposium on Spatial Data Handling (SDH'96)*, pages 173–184, Delft, The Netherlands, August 1996. Taylor and Francis.
- [6] I. Cruz and W. Sunna. Structural alignment methods with applications to geospatial ontologies. *Transactions in GIS*, 12(6):683–711, 2008.
- [7] C. d’Amato, N. Fanizzi, and F. Esposito. A semantic similarity measure for expressive description logics. In *CILC 2005, Convegno Italiano di Logica Computazionale*, Rome, Italy, June 2005.
- [8] J. David, J. Euzenat, and O. Sváb-Zamazal. Ontology similarity in the alignment space. In P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Z. 0007, J. Z. Pan, I. Horrocks, and B. Glimm, editors, *International Semantic Web Conference (1)*, pages 129–144, 2010.
- [9] C. Freksa. Temporal reasoning based on semi-intervals. *Artificial intelligence*, 54(1-2):199–227, 1992.
- [10] M. Gahegan, R. Agrawal, A. R. Jaiswal, J. Luo, and K.-H. Soon. A platform for visualizing and experimenting with measures of semantic similarity in ontologies and concept maps. *Transactions in GIS*, 12(6):713–732, 2008.
- [11] P. Gärdenfors. *Conceptual Spaces - The Geometry of Thought*. Bradford Books, MIT Press, Cambridge, MA, 2000.
- [12] H. Halpin, P. J. Hayes, J. P. McCusker, D. L. McGuinness, and H. S. Thompson. When owl: sameas isn’t the same: An analysis of identity in linked data. In P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika,
- [13] L. Z. 0007, J. Z. Pan, I. Horrocks, and B. Glimm, editors, *International Semantic Web Conference (1)*, pages 305–320, 2010.
- [14] M. Hildebrand. Search-based user interaction on the semantic web, a survey of existing systems. Technical report, Centrum Wiskunde & Informatica, The Netherlands, 2007.
- [15] K. Janowicz. Sim-DL: Towards a semantic similarity measurement theory for the description logic \mathcal{ALCN} in geographic information retrieval. In R. Meersman, Z. Tari, and P. Herrero, editors, *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, OTM Confederated International Workshops and Posters. Proceedings, Part II*, volume 4278 of *Lecture Notes in Computer Science*, pages 1681 – 1692. Springer, Montpellier, France, October 2006.
- [16] K. Janowicz, P. Maué, M. Wilkes, S. Schade, F. Scherer, M. Braun, S. Dupke, and W. Kuhn. Similarity as a quality indicator in ontology engineering. In C. Eschenbach and M. Grüninger, editors, *5th International Conference on Formal Ontology in Information Systems*, volume 183, pages 92–105. IOS Press, October 2008.
- [17] K. Janowicz, M. Raubal, and W. Kuhn. The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science*, (2):29–57, 2011.
- [18] K. Janowicz and M. Wilkes. SIM-DLA: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-concept to Inter-instance Similarity. In L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvoenen, R. Mizoguchi, E. Oren, M. Sabou, and E. P. B. Simperl, editors, *6th Annual European Semantic Web Conference (ESWC2009)*, volume 5554 of *Lecture Notes in Computer Science*, pages 353–367. Springer, 2009.
- [19] K. Janowicz, M. Wilkes, and M. Lutz. Similarity-based information retrieval and its role within spatial data infrastructures. In *GIScience '08: Proceedings of the 5th international conference on Geographic Information Science*, pages 151–167. Springer, 2008.
- [20] W. Koros. Ontology mapping for geoinformation integration. Technical report, International Institute For Geo - Information Science And Earth Observation (ITC), The Netherlands, 2010.
- [21] B. Li and F. Fonseca. Tdd - a comprehensive model for qualitative spatial similarity assessment. *Spatial Cognition and Computation*, 6(1):31–62, 2006.
- [22] A. B. Markman. *Similarity and Categorization*, chapter Structural alignment, similarity, and the internal structure of category representations., pages 109–130. Oxford University Press., Oxford, UK, 2001.
- [23] K. Nedas and M. Egenhofer. Spatial-scene similarity queries. *Transactions in GIS*, 12(6):661–681, 2008.
- [24] K. Page, D. De Roure, K. Martinez, J. Sadler, and O. Kit. Linked sensor data: Restfully serving rdf and gml. In T. K., A. Ayyagari, and D. De Roure, editors, *Proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09)*, volume Vol-522, pages 49–63. CEUR, 2009.
- [25] R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and application of a metric on semantic

⁷<http://owllink-owlapi.sourceforge.net/>

nets. *Systems, Man and Cybernetics, IEEE Transactions on*, 19(1):17–30, 1989.

- [25] M. Raubal. Formalizing conceptual spaces. In A. Varzi and L. Vieu, editors, *Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004)*, volume 114 of *Frontiers in Artificial Intelligence and Applications*, pages 153–164. IOS Press, Torino, Italy, November 2004.
- [26] A. Rodríguez and M. Egenhofer. Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science*, 18(3):229–256, 2004.
- [27] A. Schwering. *Semantic Similarity Measurement including Spatial Relations for Semantic Information Retrieval of Geo-Spatial Data*. Ph.D. Thesis, Institute for Geoinformatics, University of Münster, 2007.
- [28] P. Shvaiko and J. Euzenat. Ten challenges for ontology matching. In R. Meersman and Z. Tari, editors, *On the Move to Meaningful Internet Systems: OTM 2008*, volume 5332 of *Lecture Notes in Computer Science*, pages 1164–1182. Springer, 2008.
- [29] H. Stuckenschmidt. A semantic similarity measure for ontology-based information. In *Proceedings of the 8th International Conference on Flexible Query Answering Systems, FQAS '09*, pages 406–417, Berlin, Heidelberg, 2009. Springer-Verlag.