SIM-NET: A View-Based Semantic Similarity Model for Ad Hoc Network of Geospatial Databases

Mohamed Bakillah*  
Centre de Recherche en Géomatique  
Université Laval, Canada

Yvan Bédard  
Centre de Recherche en Géomatique  
Université Laval, Canada

Mir Abolfazl Mostafavi  
Centre de Recherche en Géomatique  
Université Laval, Canada

Jean Brodeur  
Centre d’information topographique de Sherbrooke, Canada

Abstract
Semantic similarity is a fundamental notion in GIScience for achieving semantic interoperability among geospatial data. Until now, several semantic similarity models have been proposed; however, few of these models address the issues related to the assessment of semantic similarity in ad hoc network. Also, several models are based on a definition of concepts where features are independent, an assumption that reduces the richness of the geospatial concept representation. This paper presents the conceptual basis for Sim-Net, a novel semantic similarity model for ad hoc network based on Description Logics (DL). Sim-Net is based on the multi-view paradigm. This paradigm is used to include inferential knowledge in semantic similarity, that is, the knowledge about implicit dependencies between features of concepts. In Sim-Net, assessing semantic similarity relies on the notions of Semantic Reference Systems and Formal Concept Analysis (FCA), which are combined to establish a common semantic reference frame for ontologies of the ad hoc network called the view lattice. Sim-Net semantic similarity measure distinguishes concepts that belongs to different or similar domains and takes into account the neighbours of a concept in the network. An application example is used to show the positive impact of the properties of Sim-Net.

1 Introduction
Technological advances have allowed a paradigm shift from isolated information systems to ad hoc networks. The GIS community has also taken advantage of these developments, resulting in increasing availability of geospatial data and services. This was meant to fulfill the need of numerous applications to use data from several independent geographical information systems (Lutz et al. 2003, Lemmens 2006), for example in disaster management (Bakillah et al. 2007). However technical developments are not sufficient as we must also resolve semantic discrepancies, i.e. achieve semantic interoperability (Goodchild et al. 1998, Harvey, Kuhn et al. 1999, Kavouras et al. 2005, Bian and Hu 2007). Ontologies, which are explicit specifications of a conceptualization (Gruber 1993), are a key component to support semantic interoperability.

* Corresponding author: mohamed.bakillah.1@ulaval.ca
(Brodeur et al. 2003, Fonseca et al. 2005, Arpinar et al. 2006, Fallahi et al. 2008, Kavouras and Kokla 2008), since their role is to make explicit the semantics of data (Kuhn 2003, Agarwal 2005). Ontologies are often used to describe data resources, such as database schemas and contents (Brodaric et al. 2009). Nevertheless, the problem of achieving semantic interoperability is still not resolved since ontologies are semantically heterogeneous. From this point of view, semantic similarity plays a major role for achieving semantic interoperability. It is used to determine if geospatial concepts are close in meaning, so users of different geospatial data sets can exchange data in a meaningful way. However, while ad hoc network become widespread, the concern of assessing semantic similarity between concepts of ontologies in an ad hoc network as rarely been addressed but in the non-geospatial domain (Castano et al. 2006). Semantic similarity in ad hoc network is not the same as semantic similarity between two concepts because we have to consider that ontologies may describe different domains, and the neighbourhood of a concept may influence similarity. Also, several existing models for comparing concepts suppose that features of concepts are independent, an assumption that reduce the expressivity of geospatial concepts (e.g. temperature depends on altitude, geometrical representation depends on descriptive properties, etc.). As stated in Janowicz et al. (2008), semantic similarity depends on concepts representation: a poor description leads to inaccurate results because some factors are not taken into account. This is why we argue that to obtain more accurate semantic similarity results (in terms of what is being compared) we must integrate dependent properties into the definition of concepts.

This paper describes Sim-Net, a novel semantic similarity model for ad hoc networks of geospatial databases that addresses those issues. Sim-Net addresses the requirements posed by an ad hoc network on semantic similarity: first by using standard knowledge representation language, i.e. Description Logics (DL). The assessment of semantic similarity is supported by a common semantic reference frame which is established using the notion of Semantic Reference Systems proposed by Kuhn (2003) and Formal Concept Analysis. A contribution of Sim-Net is to explore the logic view paradigm as a mean to include inferential knowledge in semantic similarity. Inferential knowledge allows discovering implicit relationships among properties of concepts, rather than considering concepts as unstructured sets of independent properties.

This paper is organized as follows: section 2 is a state of art on semantic similarity, where we review the different types of semantic similarity models and compare some of their features with Sim-Net. Section 3 discusses the requirements for semantic similarity in ad hoc networks of geospatial databases. Section 4 presents the conceptual basis of Sim-Net: the view paradigm for representing dependencies between properties and inferential knowledge; the common semantic reference frame on which the assessment of semantic similarity is based; the reasoning method for discovering semantic relations between concepts of different ontologies and the Sim-Net semantic similarity measure. Section 5 presents an application example of Sim-Net that illustrates its main features, and Section 6 concludes this paper.

2 State of Art on Semantic Similarity

Numerous semantic similarity models have been proposed in the literature; models for the geospatial domain have been described in a recent review by Schwering (2008). The models for comparing concepts include geometric, feature and network models.

Geometric models are based on the concept of multidimensional vector space. Each dimension represents a property of concepts (for ex., size); the values of a property (for ex., thin,
large) are shown as values on the corresponding dimension. Concepts are represented by multidimensional regions in this vector space. Semantic similarity between concepts can be computed as a function of spatial distance (e.g. Minkowski distance) between vectors forming the boundaries of the regions representing concepts (Schwering and Raubal 2005). This model assumes that the compared concepts are defined with the same dimensions. This is not the case for concepts of different ontologies where a same real world phenomenon can be represented with different properties that are relevant to the application domain. In Schwering and Kuhn (2009), this model was extended to take into account relations between concepts. In this extended model, concepts be defined with different dimensions; however, dimensions either match or mismatch, but there is no partial match. In Schwering and Raubal (2005) and Schwering and Kuhn's models, properties are independent of each other; however, Raubal (2004) proposed that dependent properties may be modelled via non-orthogonal dimensions, but this idea was not further formalized.

In network models, concepts are nodes in a graph, and their semantics are given by their relative position in this graph (Raftopoulou and Petrakis 2005). Semantic similarity is a decreasing function of the distance between two concepts. Several network models have been proposed, which assign weights to the different types of relationships (Maquitman et al. 2005), combine the shortest path length with the depth first common ancestor concept (Li et al. 2003), compare neighbouring nodes of concepts (Do and Rahm 2002), or incorporate the notion of information content (Resnick 1999). The drawbacks of network models is that they often assume a representation of concepts with labels only, while geospatial concepts are more complex, having spatial, temporal and thematic properties. While Sim-Net uses notions of network models, it incorporate a complex representation of concepts adapted to the geospatial domain.

Feature models represent concepts as unstructured sets of features; they are based on set theory. The ratio model of Tversky (1977) evaluates the semantic similarity according to the ratio of common and exclusive features. Rodriguez and Egenhofer's Matching Distance model (2003) combines both the ratio model of Tversky with network distance. An example of feature model that determine qualitative relationships among concept is the geosemantic proximity model (Brodeur and Bédard 2001) which provides geosemantic proximity predicates based on Egenhofer’s topological predicates (Egenhofer 1993). Those feature models cannot provide partial matches between features since they either match or mismatch. However, the Matching Distance model has been extended to allow measuring such partial matches (Bakillah et al. 2006). However, those features models remain problematic since features are assumed to be independent.

The geometric, feature and network models are all inspired from the human perception of similarity (Schwering and Kuhn 2009). At another level, we must also mention the logic-based semantic similarity models, which define concepts with a logical language such as Description Logics (DL). For example, d’Amato et al. (2005) proposed a semantic similarity measure for ALC Description Logics. This measure uses instances of concepts. It should be noted that geometric, feature and network models can be represented with Description Logics (Borgida et al. 2005). Another example of logic-based model is Sim-DL by Janowicz (2006). Sim-DL, as Sim-Net, is based on Description Logics (DL). Sim-DL compares concepts described with ALCNR DL. It compares primitive concepts, roles, and cardinality restrictions on roles, and provides with a weighted sum of similarity with respect to these features. In comparison, Sim-Net also considers data type properties, which are required for expressing spatial and temporal properties. A difference between Sim-DL and Sim-Net is that we consider that properties are not independent from each other. In Sim-DL it is proposed that the weights for the different
similarity terms can be computed based on probabilistic methods, while we compute weights using domain similarity. As Sim-Net is specifically targeted at ad hoc networks, one difference between Sim-Net and other DL-based models is that it includes the notion of inter-ontology neighbourhood. While the notion of neighbourhood was exploited in the Matching Distance model of Rodriguez and Egenhofer (2003), the neighbourhood of a concept is restricted to close concepts in the same ontology but not to the network.

3 Semantic Similarity in Ad Hoc Network

3.1 Ad Hoc Networks of Geospatial Databases

An ad hoc network is a network where some of the databases (or “nodes”) are made available to a community of users for the duration of a punctual need (Fernandez 2007). We assume that each database of the ad hoc network commits to an ontology. Ontologies play a key role by capturing the shared conceptualization of a community of users. Consequently, they support interoperability between different databases (Smith and Mark 1998; Fonseca et al. 2005). Figure 1 shows how we represent the ad hoc network.

![Figure 1 Ad Hoc Network of Ontologies](image)

Each node of the network (labelled O_i) represents an ontology. The ontologies are gathered in subsets (represented with dotted circles) called punctual clusters. Ontologies contained in a punctual cluster are available to a community of users having common interests. They stand for the users’ conceptual representation. The ad hoc network is dynamically modified when the
users’ interest change or when nodes are added or dropped from the network. When such changes occur, punctual clusters have to be re-organized by a coordinator user. Semantic similarity among concepts of different ontologies can be affected by such changes, since we assume that the neighbourhood of a concept (the other concepts to which it is semantically linked) contribute to define its semantics.

3.2 What do we need for Semantic Similarity in Ad Hoc Network of Spatiotemporal Databases?

- **Anchoring semantic similarity in standard knowledge representation language.** According to Janowicz (2006), the Web Ontology Language (OWL), which is based on DL, is the most widely adopted ontology language for geo-ontologies as it is recommended by W3C (Baader et al. 2003). A semantic similarity model suitable for networks should be based on such commonly accepted language in order to avoid the problem of incompatibility between knowledge representation and comparison criteria (Janowicz 2006). The advantages of DL are its sound semantics and its reasoning capabilities. It also supports complex concept description (Borgida et al. 2005). However, while very expressive families of DL exist, they are not tractable because of their complexity.

- **Defining a common semantic reference frame for the ad hoc network.** Assessing semantic similarity among concepts from different ontologies in the ad hoc network requires that the concepts can be referenced in a common semantic reference system (Kuhn (2003), which is the analogous of spatial reference systems. Semantic reference systems are more than ontologies, but ontologies are a core component of them (Kuhn and Raubal 2003). The semantic datum's role it to ground the meaning of basic terms, the semantic reference frame is the formally defined framework to which terms can be related to obtain meaning (it could be a top-level ontology). Semantic referencing is the process of linking the terms in a local model to the semantic reference frame. Semantic translation and projection and transformation between models.

- **Determining semantic similarity among concepts of different domains.** An ad hoc network is populated with ontologies that describe different domains. Several semantic similarity models are based on the assumption that similar concepts have similar properties. In ad hoc network, we cannot always make this assumption since ontologies of different domains describe the same concept with different properties.

- **Representation of concepts suitable for geospatial concepts** Geospatial concepts are often related through logic rules that express dependencies between their properties. There are several reasons which demonstrate why logic rules are fundamental elements. First, geospatial concepts are often described with physical properties (e.g. temperature, altitude, size, density, etc.); by nature, physical properties tend to depends on each other (e.g. temperature depends on altitude). Also, the geometry and temporality used to represent geospatial objects depends on other spatial, temporal or thematic properties, for example, watercourses larger than 7,5 are represented by surfaces while those that are thinner than 7,5 m are represented by lines.

- **Propagation of semantic similarity in ad hoc network.** Semantic similarity can be deduced from existing relationships between concepts, through inference mechanisms based, for example, on the transitivity property of semantic relationships. Transitivity means that if a first concept C1 is similar to C2, and C2 is similar to C3, than C1 is similar to C3. However, this property has been criticized on the basis of the famous example of James: a lamp is
similar to the moon and the moon is similar to a ball; but a lamp is not similar to a ball. This is because similarity depends on what is being compared. Lamp and moon are similar with respect to their function (to provide with light), while moon and ball are similar with respect to their shape. Therefore, we should be careful when using transitivity and ensure that we compare concepts with respect to the same aspect. In this paper, however, we do not address this requirement and leave it for future work.

4 Sim-Net Semantic Similarity Model

The conceptual basis for Sim-Net consists of the following elements: first we present the logic view paradigm (section 4.1). We use the Formal Concept Analysis (FCA) theory to build a common semantic reference frame to which concepts and views of local ontologies can be referenced (section 4.2). In section 4.3, we give the DL-based reasoning rules used by Sim-Net to determine the semantic relationships among views and among concepts. Finally, we give the Sim-Net semantic similarity measure.

4.1 The Logic View Paradigm and View Extraction Method

Several researchers have shown interest in the view paradigm, both in the domain of database (Debrauwer 1998, Bédard and Bernier 2002, Benchikha et al. 2005, Parent et al. 2006) and in the domain of ontologies (Noy and Musen 2004, Bhatt et al. 2006, Stuckenschmidt 2006, Wouters et al. 2008). In the database domain, views are a mean of handling multi-representation (Bédard and Bernier 2002, Parent et al. 2006). Views can also represent the different states of an evolving object (Debrauwer 1998). In the ontology domain, the view paradigm supports ontology reuse by selecting only parts of an ontology that are relevant in a given context. In our approach, views are used as a mean to handle inferential knowledge obtained from logic rules.

Geospatial concepts are complex since they are described by spatial properties such as shape and position, spatial relations (Schwering 2008) and temporal relations. Furthermore, they are often described by logic rules that constraint their property values, for example when an industry has for property value type of product = toxic substance, it is situated at more than 3km from residential areas. The knowledge extracted from the conjunction of these logic rules is called inferential knowledge (Steffens 2005). For example, consider a concept “road” with properties road type = {street, boulevard} and “number of lanes”. The following logic rule expresses a relation between “road type” and “number of lane”: (road type(X) = boulevard)→(number of lanes(X) ≥ 2). If we have another rule (number of lanes(X) ≥ 2)→(road geometry = multi-lines), we can infer a new relationship between “road type” and “road geometry”, in the form of a new logic rule: (road type(X) = boulevard)→(road geometry = multi-lines). The general form of rules is:

\[ r : [p_i(X) = v_i] \rightarrow [p_j(Y) = v_j] \]  

(1)

where \( v_i \) is an element of the range of property \( p_i \), \( v_j \) is an element of the range of property \( p_j \), and X and Y are variables of instances of concepts. The first member of the logic rule is called the antecedent and the second member is the consequent.

The view paradigm consists in expressing inferential knowledge obtained from logic rules with logic views of a concept. From the above example, we see that a logic view of the concept road can be “boulevard” with properties road geometry = multi-lines and number of lanes(X) ≥ 2.
Because of the lack of space, we are only giving an overview of the view extraction method. We assume that a concept is defined by its name, a set of properties from the categories shown in Figure 2, a set of relationships and a set of logic rules. Figure 2 shows a classification of seven sub-types of properties related with “is-a-kind-of” relations. Spatial properties are properties which range is a spatial data type (point, line, polygon, etc.). The range of temporal properties is a temporal data type (instant, period), while thematic properties have string or numerical values as a range. The is-a-kind-of relation indicates that sub-properties inherit the characteristics of the super-properties. For example, spatiotemporal properties inherit the characteristics of spatial and temporal properties, so their range is a spatial data type associated to a temporal datatype.

![Figure 2](image)

**Figure 2** The different types of properties of concepts

The idea behind the view extraction method is to use each logic rule to extract a partial view of the concept, and then combine the partial views with compatible values of properties. It is summarized as follows:

1. Extracting existing logic rules from the definition of concepts. Two statements can be made in this regard: either rules are explicit ontology element defined at design time: then we can obtain them by browsing the ontology. Otherwise, it is possible that such rules are not directly available form the ontology but can be discovered by examining the instances of concepts with mining rules.

2. Applying the inference mechanism between existing rules to discover new rules. This mechanism states that if the consequent of a first logic rule $r1$ implies the antecedent of a second logic rule $r2$, then the antecedent of $r1$ also implies the consequent of $r2$:

   \[
   (r1: [p_j(Y) = v_j] \rightarrow [p_j(Y) = v_j]) \land \\
   (r2: [p_j(Y) = v_j] \rightarrow [p_k(Z) = v_k]) \land \\
   (\{ [p_j(Y) = v_j] \rightarrow [p_j(Y) = v_j] \rightarrow \\
   (r3: [p_j(X) = v_j] \rightarrow [p_k(Z) = v_k]) \land \\
   \]}

   (2)
3. Create a partial view from each logic rule, following the association of properties and property values stated by the rule.

4. Merge partial views to obtain the views of the concept: we merge all partial views that contain compatible values of properties, until the view specifies the range of each property of the concept. The validity of extracted views can be verified with consistency checking (verifying that views are instanciable).

Once views are extracted from concepts, we obtain a set of ontologies extended with views. The next step is to reference the concepts and these views in a common semantic reference frame.

4.2 Building the Common Semantic Reference Frame: The View Lattice

We use the notions of reference frame and referencing described in the theory of Semantic Reference Systems (Kuhn 2003). The view lattice plays the role of the semantic reference frame, and it is built using the Formal Concept Analysis (FCA) method. FCA provides a framework for placing concepts of different ontologies in a single hierarchy called the Concept Lattice (Ganter and Wille 1999). A view lattice is a set of formal concept and formal views that are linked by inheritance relationships. We summarize the method for building the view lattice:

Step 1) **Generation of the set of reference concepts** – reference concepts define the common vocabulary for a set of local ontologies. Our approach to generate the reference concepts is as follows: first, we gather all concepts of local ontologies. We group the concepts that are synonyms into subsets. These subsets are the reference concepts. Then, we find the subsets that contain concepts related with is-a relations and add this knowledge to the definition of reference concepts. Synonyms and is-a relations can be identified with WordNet (Miller 1995), a domain-independent thesaurus for the English language. Building this common vocabulary solves lexical heterogeneity across ontologies.

Step 2) **Projection of local concepts to reference concepts** – The next step is to project the concepts from different ontologies (local concepts) to the reference concepts in a projection matrix. Table 1 shows an example where local concepts are in lines and reference concepts formed from these local concepts are in columns.

<table>
<thead>
<tr>
<th>Local Concepts</th>
<th>Reference Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Stream/Watercourse)</td>
</tr>
<tr>
<td>O1: Stream</td>
<td>x</td>
</tr>
<tr>
<td>O2: Watercourse</td>
<td>x</td>
</tr>
<tr>
<td>O1: City</td>
<td>x</td>
</tr>
<tr>
<td>O2: Urban Area</td>
<td>x</td>
</tr>
<tr>
<td>O1: Land</td>
<td>x</td>
</tr>
<tr>
<td>O2: Lowland</td>
<td>x</td>
</tr>
<tr>
<td>O2: Upland</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Example of projection of local concept to reference concepts
Each local concept is identified by a prefix (in this example, O1 or O2) that identifies the local ontology it belongs to. An x sign indicates when a local concept contains a reference concept.

Step 3) **Identification of formal concepts from the projection matrix** – In the original FCA theory, a formal concept is a pair \((A = \text{set of object}, B = \text{set of attributes})\), where objects in \(A\) are described in terms of attributes of \(B\). In our context, it is local concepts that are described in terms of reference concepts. Therefore, we apply the FCA theory by establishing the correspondences object-local concept and attribute-reference concepts. A formal concept is a pair \(FC = <(\text{local concepts}), (\text{reference concepts})>\) that co-occur in the matrix, for example: \(<(\text{Stream, Watercourse}), (\text{Stream/Watercourse})>\) and \(<(\text{Land, Lowland}), (\text{Lowland})>\).

Step 4) **Identification of inheritance relationships and generation of the lattice** – We verify if a formal concept includes another formal concept in terms of reference concept, and built the upper part of the view lattice, which contains all formal concepts (Figure 3):

![Figure 3](image)

**Figure 3**  Upper part of the view lattice containing only formal concepts

Step 5) **Expansion of the upper part of the view lattice with formal views** – This is similar to steps 1 to 4, however the local views are projected to reference properties and values of properties. Table 2 illustrates an example of projection of the views of concepts Stream and Watercourse, where an x sign indicates that a view has a reference property and value.

We identify the formal views, which are pairs \(FC = <(\text{local views}), (\text{reference property: reference value})>\) that co-occur in the matrix. The formal views for Table 2 are the following, and the complete view lattice that expands the upper part of Figure 3 is shown in Figure 4.

\[
FV_1 = <(V_1(O1:Stream), V_1(O2:Watercourse)), (\text{Stream Geometry/Watercourse Geometry: Surface/Region})> \\
FV_2 = <(V_2(O1:Stream), V_1(O2: Watercourse)), (\text{Stream Class/Watercourse Category: River})> \\
FV_3 = <(V_1(O1:Stream), (\text{Stream Class/Watercourse Category: Canal; Stream Geometry/Watercourse Geometry: Surface/Region})>
\]
FV4 = \(<\text{V2(O1:Stream)}, \text{(Stream Class/Watercourse Category: River; Stream Geometry/Watercourse Geometry: Surface/Region → Time interval)}\>\)

FV5 = \(<\text{V1(O2:Watercourse)}, \text{(Stream Class/Watercourse Category: River; Stream Geometry/Watercourse Geometry: Surface/Region)}\>\)

FV6 = \(<\text{V2(O2:Watercourse)}, \text{(Stream Class/Watercourse Category: Ditch; Stream Geometry/Watercourse Geometry: Line)}\>\)

Table 2 Example of projection of local views to reference properties and values

<table>
<thead>
<tr>
<th>Local views</th>
<th>Reference properties and values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stream Class/Watercourse Category</td>
</tr>
<tr>
<td></td>
<td>Stream Geometry/Watercourse Geometry</td>
</tr>
<tr>
<td>V1 (O1:Stream)</td>
<td>x</td>
</tr>
<tr>
<td>V2 (O1:Stream)</td>
<td>x</td>
</tr>
<tr>
<td>V1 (O2: Watercourse)</td>
<td>x</td>
</tr>
<tr>
<td>V2 (O2: Watercourse)</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 4 Complete view lattice with formal concepts (white boxes) and formal views (grey boxes)
Step 6) **Referencing Views and Concepts to the View Lattice** – A concept $C$ is referenced to a formal concept $FC$ (denoted $C \xrightarrow{\text{RefTo}} FC$) if it verifies the following condition: the set of local concept of $FC$ is the smallest set that contains $C$ in the view lattice.

A view $V$ is referenced to a formal view $FV$ (denoted $V \xrightarrow{\text{RefTo}} FV$) if it verifies (V1) to (V3):

(V1) Properties of $FV$ are all included in properties of $V$;
(V2) For all properties owned both by $FV$ and $V$, the associated property values in $V$ are all included in the set of property values of $FV$;
(V3) $FV$ is the lowest formal view of the view lattice that satisfies conditions (V1) and (V2).

Once views and concepts of different ontologies are referenced to the common view lattice, the setting is ready for assessing semantic similarity with Sim-Net.

### 4.3 The Sim-Net Semantic Similarity Model

Sim-Net is a DL-based semantic similarity model for *ad hoc* network that determines semantic relationships among concepts and their semantic similarity value. Consider a set of $N$ ontologies $\{O_1, O_2, \ldots, O_N\}$ from the *ad hoc* network. Sim-Net takes this set of ontologies and returns a set of pairs of concepts, $O_i: C_j$ and $O_k: C_l$, their associated semantic relationship $R$ and semantic similarity value $SN$:

$$Sim\text{-Net} : O_1 \times O_2 \ldots O_N \rightarrow \{O_i : C_j; O_k : C_l; R(C_j, C_l); SN(C_j, C_l)\}$$

(3)

Sim-Net distinguishes from other semantic similarity model by the following assumptions:

- *Ad hoc* network’s ontologies describe different domains; therefore we cannot assume that concepts describing a same reality necessarily share common properties;
- When a concept from a first ontology is linked to concepts of other ontologies, these concepts also contribute in defining its semantics. Therefore assessing semantic similarities between concepts of two ontologies is different than between ontologies of a network.

In the following, we give fundamentals on DL (section 4.3.1) and we explain how semantic relationships (section 4.3.2) and semantic similarity values (section 4.3.3) are determined.

#### 4.3.1 Description Logics

Description Logics are a family of representation languages widely adopted for knowledge representation and reasoning (Baader *et al.* 2003, Lemmens, 2006, Fallahi *et al.* 2008). They are based on the notion of concepts and roles. Constructors (universal quantification, existential restriction, conjunction, etc.) allow defining complex concepts and complex roles from primitive ones. Common constructors are listed in Table 3.
**Table 3** Syntax and semantics of Common Description Logic Constructors

<table>
<thead>
<tr>
<th>Name</th>
<th>Syntax</th>
<th>Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top concept</td>
<td>$Δ^l$</td>
<td></td>
</tr>
<tr>
<td>Bottom concept</td>
<td>$\emptyset$</td>
<td></td>
</tr>
<tr>
<td>Atomic concept</td>
<td>$C$</td>
<td>$C^l \subseteq Δ^l$</td>
</tr>
<tr>
<td>Atomic role</td>
<td>$R$</td>
<td>$R^l \subseteq Δ^l \times Δ^l$</td>
</tr>
<tr>
<td>Full negation</td>
<td>$\neg C$</td>
<td>$Δ^l / C^l$</td>
</tr>
<tr>
<td>Concept equality</td>
<td>$C \equiv D$</td>
<td>$C^l = D^l$</td>
</tr>
<tr>
<td>Concept inclusion</td>
<td>$C \subseteq D$</td>
<td>$C^l \subseteq D^l$</td>
</tr>
<tr>
<td>Concept union</td>
<td>$C \cup D$</td>
<td>$C^l \cup D^l$</td>
</tr>
<tr>
<td>Concept intersection</td>
<td>$C \cap D$</td>
<td>$C^l \cap D^l$</td>
</tr>
<tr>
<td>Role equality</td>
<td>$R \equiv S$</td>
<td>$R^l = S^l$</td>
</tr>
<tr>
<td>Role inclusion</td>
<td>$R \subseteq S$</td>
<td>$R^l \subseteq S^l$</td>
</tr>
<tr>
<td>Existential quantification</td>
<td>$\exists R.C$</td>
<td>${a \in Δ^l \mid \exists b.(a,b) \in R^l \land y \in C^l}$</td>
</tr>
<tr>
<td>Value restriction</td>
<td>$\forall R.C$</td>
<td>${a \in Δ^l \forall b.(a,b) \in R^l \rightarrow y \in C^l}$</td>
</tr>
<tr>
<td>Maximum number restriction</td>
<td>$\leq NR.C$</td>
<td>${a \in Δ^l \mid {b \in Δ^l \mid (a,b) \in R^l \land b \in C^l} \leq n}$</td>
</tr>
<tr>
<td>Minimum number restriction</td>
<td>$\geq NR.C$</td>
<td>${a \in Δ^l \mid {b \in Δ^l \mid (a,b) \in R^l \land b \in C^l} \geq n}$</td>
</tr>
</tbody>
</table>

The semantics of those constructors are given by an interpretation $I=(Δ^l, ω)$, where $Δ^l$ is the set of instances and $\omega$ is the function that associate instances to their concepts. The different forms of description logics are determined by the constructors that are used, and give the expressive power of DL. Sim-Net is based on the $SHIQ(D)$ DL form, for we need to be able to express inverse roles ($I$) which are necessary to describe spatial relationships (e.g. IsIncludedIn is the inverse of Includes), qualified number restriction ($Q$) for expressing cardinality restrictions on relationships, and datatype properties ($D$), for example HasArea, that link a concept with a spatial data type such as polygon:

$$Lake \equiv Waterbody \cap \exists HasArea.Polygon.$$ 

### 4.3.2 Reasoning with DL for Determining Semantic Relationships in Ad Hoc Network

Sim-Net uses reasoning rules for determining semantic relationships at two levels: between views of different concepts, and between concepts. Consider the DL-expression of views $O_1: V_1$ and $O_2: V_2$ which are given by:

$$O_1: V_1 \equiv \bigcap_{A_i^l \in \{\text{Primitive views}\}} A_i^l \cap \bigcap_{R_j \in \{\text{Primitive Role}\}} \forall R_j.C_j^l \cap \bigcap_{T_k \in \{\text{Datatype Property}\}} \forall T_k.F_k$$  

(4)

with $i$ element of $[1, 2, i']$, $j$ element of $[1, 2, j']$ and $k$ element of $[1, 2, k']$.  

$$O_2: V_2 \equiv \bigcap_{B_m^l \in \{\text{Primitive views}\}} B_m^l \cap \bigcap_{S_n \in \{\text{Primitive Role}\}} \forall S_n.C_n^l \cap \bigcap_{U_p \in \{\text{Datatype Property}\}} \forall U_p.G_p$$  

(5)
with \(m\) element of \([1, 2, m']\), \(n\) element of \([1, 2, n']\) and \(p\) element of \([1, 2, p']\). Each view is defined by the conjunction of a set of primitive views \((A_i^j, B_n^k)\), a set of primitive roles \((R_j, S_m)\), and a set of datatype properties \((T_k, U_p)\). Each view is referenced to a formal view of the view lattice \(L\):

\[
O_1 : V_1 \xrightarrow{\text{RefTo}} L : FV_1 \quad \text{and} \quad O_2 : V_2 \xrightarrow{\text{RefTo}} L : FV_2 .
\]

Concepts are defined as sets of views, and their definition is given by:

\[
O_1 : C_1 \equiv V_1^1 \cup V_2^2 \cup \ldots \cup V_i^i \ldots \cup V_i^N \quad (6)
\]

\[
O_2 : C_2 \equiv V_1^1 \cup V_2^2 \cup \ldots \cup V_i^j \ldots \cup V_i^M \quad (7)
\]

Each concept is also referenced to a formal concept of the common view lattice \(L\):

\[
O_1 : C_1 \xrightarrow{\text{RefTo}} L : FC_1 \quad \text{and} \quad O_2 : C_2 \xrightarrow{\text{RefTo}} L : FC_2 .
\]

Sim-Net computes semantic relationships among views of concepts with view-reasoning rules that define the conditions for a semantic relationship to be verified (second row of Table 4). Then, Sim-Net deduces the relationship between two concepts by reasoning with the relationships between their views, using reasoning rules defined in the third row of Table 4. The principle for deducing relationship between concepts from relationships between views is that each concept is a set of views. The semantic relationships are based on classical set theory: equivalence, generalisation, overlap and disjointness. In addition, we consider semantic relationships that can be established between concepts (views) of different domains. Classically, equivalence between two concepts (views) is established only if they have exactly the same set of properties and relationships. In different domains, two concepts representing the same reality may have different properties. In this case, we allow two concepts (views) to be cross-domain equivalent if they have different properties but are related to the same formal concept (formal view) in the view lattice. We apply the same reasoning for defining the other cross-domain relationships.

### Table 4 Semantic relationships and reasoning rules

<table>
<thead>
<tr>
<th>Semantic Relationships</th>
<th>Views Reasoning Rules</th>
<th>Concept Reasoning Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strong Equivalence</strong></td>
<td><strong>Expression:</strong> ( O_1 : V_1 \xrightarrow{=} O_2 : V_2 )</td>
<td><strong>Expression:</strong> ( O_1 : C_1 \xrightarrow{=} O_2 : C_2 )</td>
</tr>
<tr>
<td><strong>Rules:</strong></td>
<td></td>
<td><strong>Rules:</strong></td>
</tr>
<tr>
<td>1) ( \forall A_i^j, i = 1 \ldots i', \exists B_j^k ) where ( A_i^j \equiv B_j^k ) and ( \forall B_j^k, j = 1 \ldots j', \exists A_i^j ) where ( A_i \equiv B_j )</td>
<td>1) ( \forall V_i^j, 1 \leq i \leq N, \exists V_j^i, 1 \leq j \leq M ) where ( V_i^j \equiv V_j^i )</td>
<td>2) ( \forall V_j^i, 1 \leq j \leq M, \exists V_i^j, 1 \leq i \leq N ) where ( V_j^i \equiv V_i^j )</td>
</tr>
</tbody>
</table>
\( R_j \equiv S_m \) and \( C_j^V \equiv C_m^V \) and
\[ \forall S_m, m = 1..m', \exists R_j \text{ where} \]
\[ S_m \equiv R_j \text{ and } C_m^V \equiv C_j^V \]
3) \[ \forall T_k, k = 1..k', \exists U_p \text{ where} \]
\[ T_k \equiv U_p \text{ and } F_k \equiv G_p \text{ and} \]
\[ \forall U_p, p = 1..p', \exists T_k \text{ where} \]
\[ U_p \equiv T_k \text{ and } G_p \equiv F_k \]

**Cross-Domain Equivalence**

**Expression:** \( O_1 : V_1 \xrightarrow{\sim} O_2 : V_2 \)

**Rules:**
\[ O_1 : V_1 \xrightarrow{\text{RefTo}} L : FV_1 \text{ and} \]
\[ O_2 : V_2 \xrightarrow{\text{RefTo}} L : FV_2 \text{ with} \]
\[ FV_1 \equiv FV_2 \]

**Strong Generalization**

(Inverse of Strong Specialization)

**Expression:** \( O_1 : V_1 \xrightarrow{\dashv} O_2 : V_2 \)

**Rules:**
1) \[ \forall A_i^V, i = 1..i', \exists B_i^V \text{ where} \]
\[ A_i^V \equiv B_i^V \]
2) \[ \forall R_j, j = 1..j', \exists S_m \text{ where} \]
\[ R_j \equiv S_m \text{ and } C_j^V \equiv C_m^V \text{ or } C_j^V \supset C_m^V \]
3) \[ \forall T_k, k = 1..k', \exists U_p \text{ where} \]
\[ T_k \equiv U_p \text{ and } F_k \equiv G_p \text{ or } G_p \subset F_k \]

**Cross-Domain Generalization**

(Inverse of cross-domain generalization)

**Expression:** \( O_1 : V_1 \xrightarrow{\Rightarrow} O_2 : V_2 \)

**Rules:**
\[ O_1 : V_1 \xrightarrow{\text{RefTo}} L : FV_1 \text{ and} \]
\[ O_2 : V_2 \xrightarrow{\text{RefTo}} L : FV_2 \text{ with} \]
\[ FV_1 \supset FV_2 \]

**Cross-Domain Equivalence**

**Expression:** \( O_1 : C_1 \xrightarrow{\sim} O_2 : C_2 \)

**Rules:**
\[ O_1 : C_1 \xrightarrow{\text{RefTo}} L : FC_1 \text{ and} \]
\[ O_2 : C_2 \xrightarrow{\text{RefTo}} L : FC_2 \text{ with} \]
\[ FC_1 \equiv FC_2, \text{ or} \]
1) \[ \forall V_1^j, 1 \leq i \leq N, \exists V_2^j, 1 \leq j \leq M \]
\[ \text{where } V_1^i \xrightarrow{\ast} V_2^j \]
2) \[ \forall V_2^j, 1 \leq j \leq M, \exists V_1^i, 1 \leq i \leq N \]
\[ \text{where } V_2^j \xrightarrow{\ast} V_1^i \]

**Strong Generalization**

(Inverse of Strong Specialization)

**Expression:** \( O_1 : C_1 \xrightarrow{\dashv} O_2 : C_2 \)

**Rules:**
1) \[ \forall V_2^j, 1 \leq j \leq M, \exists V_1^i, 1 \leq i \leq N \]
\[ \text{where } V_1^i \xrightarrow{\equiv} V_2^j \]
or \( V_1^i \xrightarrow{\equiv} V_2^j \)
2) \[ \forall V_1^i, 1 \leq i \leq N \text{ where} \]
\[ V_1^i \xrightarrow{\equiv} V_2^j \text{ and } V_1^j \xrightarrow{\equiv} V_2^j \]

**Cross-Domain Generalization**

(Inverse of cross-domain generalization)

**Expression:** \( O_1 : C_1 \xrightarrow{\Rightarrow} O_2 : C_2 \)

**Rules:**
\[ O_1 : C_1 \xrightarrow{\text{RefTo}} L : FC_1 \text{ and} \]
\[ O_2 : C_2 \xrightarrow{\text{RefTo}} L : FC_2 \text{ with} \]
\[ FC_1 \supset FC_2, \text{ or} \]
<table>
<thead>
<tr>
<th>Domain</th>
<th>Expression</th>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak Overlap</td>
<td>$O_1 : V_1 \xrightarrow{\ominus} O_2 : V_2$</td>
<td>1) $\exists A_i^V, i = 1...i'$ and $\exists B_i^V, l = 1...l'$ where $A_i \equiv B_i$ and/or $\exists S_m, m = 1...m'$ and $R_j, j = 1...j'$ where $R_j \equiv S_m$ and $C^V_j$ and $C^V_m$ are not disjoint and/or $\exists U_p, p = 1...p'$ and $\exists T_k, k = 1...k'$ where $T_k \equiv U_p$ and $F_k$ and $G_p$ are not disjoint</td>
</tr>
<tr>
<td>Strong Overlap</td>
<td>$O_1 : V_1 \xrightarrow{\ominus} O_2 : V_2$</td>
<td>1) $\exists A_i^V, i = 1...i'$ and $\exists B_i^V, l = 1...l'$ where $A_i \equiv B_i$ and/or $\exists S_m, m = 1...m'$ and $R_j, j = 1...j'$ where $R_j \equiv S_m$ and $C^V_j$ and $C^V_m$ are not disjoint and/or $\exists U_p, p = 1...p'$ and $\exists T_k, k = 1...k'$ where $T_k \equiv U_p$ and $F_k$ and $G_p$ are not disjoint</td>
</tr>
<tr>
<td>Cross-Domain Overlap</td>
<td>$O_1 : V_1 \xrightarrow{\ominus} O_2 : V_2$</td>
<td>1) $\exists A_i^V, i = 1...i'$ and $\exists B_i^V, l = 1...l'$ where $A_i \equiv B_i$ and/or $\exists S_m, m = 1...m'$ and $R_j, j = 1...j'$ where $R_j \equiv S_m$ and $C^V_j$ and $C^V_m$ are not disjoint and/or $\exists U_p, p = 1...p'$ and $\exists T_k, k = 1...k'$ where $T_k \equiv U_p$ and $F_k$ and $G_p$ are not disjoint</td>
</tr>
</tbody>
</table>

Cross-Domain Overlap

Expressions:

1. $O_1 : V_1 \xrightarrow{\ominus} O_2 : V_2$  
   $O_1 : V_1 \xrightarrow{\ominus} L : FV_1$ and $O_2 : V_2 \xrightarrow{\ominus} L : FV_2$ with $FV_1 \prec FV_2$ and $FV_2 \prec FV_1$

Rule:

$O_1 : C_1 \xrightarrow{\ominus} L : FC_1$ and $O_2 : C_2 \xrightarrow{\ominus} L : FC_2$ with $FC_1 \prec FC_2$, $FC_2 \prec FC_1$, or

$\exists V_1^i, i \leq M, \exists V_2^j, j \leq M$ where $V_1^i \xrightarrow{\ominus} V_2^j$.
The result obtained by applying the above reasoning rules is a single ontology graph which links views and concepts of a set of ontologies (Figure 5).
The semantic similarity among concepts is defined by a combination function that merges the results of three semantic similarity measures: similarity between views of concepts \( \text{Sim}_{\text{view}} \), cross-domain similarity \( \text{Sim}_{\text{cd}} \) and inter-ontology neighbourhood similarity \( \text{Sim}_{\text{ion}} \):

\[
SN(C_1, C_2) = \omega_{\text{view}} \text{Sim}_{\text{view}}(C_1, C_2) + \omega_{\text{cd}} \left[ \frac{\text{Sim}_{\text{cd}}(C_1, C_2) + \text{Sim}_{\text{ion}}(C_1, C_2)}{2} \right]
\]  

(8)

The \( \text{Sim}_{\text{view}} \) term accounts for the principle that concepts that have similar views are also similar. The \( \text{Sim}_{\text{cd}} \) term accounts for the principle that concepts are similar if they are referenced to similar formal concepts in the view lattice. This term allows us to find non-zero similarity even if concepts have no common properties, as it can produce when ontologies describe different domains. The \( \text{Sim}_{\text{ion}} \) term accounts for the fact that concepts are similar if their inter-ontology neighbourhoods contain similar concepts. For concepts of similar domains, we expect that \( \text{Sim}_{\text{view}} \) should be more important than \( \text{Sim}_{\text{cd}} \) and \( \text{Sim}_{\text{ion}} \). Therefore, more weight is given to \( \text{Sim}_{\text{view}} \) if domains of concepts are similar, while more weight is given to \( \text{Sim}_{\text{cd}} \) and \( \text{Sim}_{\text{ion}} \) when domains of concepts are dissimilar. We define in section 4.3.3.2 computation of weights; however we will first discuss the semantic similarity terms.

### 4.3.3.1 Adapting the Normalized Google Distance to the Ad Hoc Network

The Normalized Google Distance (NGD) was introduced by Cilibrasi and Vitanyi (2007) for assessing semantic distance between concepts in the Web. The semantics of a concept is given by the set of Web pages returned by the Google search engine when this concept is used as the query word. For concepts \( x \) and \( y \), the NGD is:

\[
\text{NGD}(x, y) = \frac{\max \{\log f(x), \log f(y)\} - \log f(x, y)}{\log M - \min \{\log f(x), \log f(y)\}}
\]  

(9)

\( f(x) \) is the number of web pages containing \( x \), \( f(x, y) \) is the number of pages containing \( x \) and \( y \) and \( M \) is the number of pages indexed by Google. NGD is a measure of the probability of co-occurrence of \( x \) and \( y \) in the Web. The more the concepts will co-occur in the same page, the smaller the semantic distance. The Web is considered as a giant ontology graph. The structure of an ad hoc network, where concepts of ontologies are related with semantic relationships discovered by Sim-Net, is similar to the ontology defined by the Web. For each similarity term \( \text{Sim}_{\text{view}}, \text{Sim}_{\text{cd}} \) and \( \text{Sim}_{\text{ion}} \), we define a similarity function based on a re-interpretation of NGD. Let \( a, b \) two variables (views, formal concepts or concepts). The network distance (ND) between \( a \) and \( b \) is given by:

\[
\text{Network Dist}(a, b) = \text{ND}(a, b) = \frac{\max \{\log f(a), \log f(b)\} - \log f(a, b)}{\log(\text{SizeNet}) - \min \{\log f(a), \log f(b)\}}
\]  

(10)

The following definitions give the interpretations of \( f(a), f(b), f(a, b) \) and \( \text{SizeNet} \). In this regard, it is worth mentioning that the probability of co-occurrence measured by the NDG when using the Web cannot be confounded with semantic similarity. For example, two words may co-occurs very frequently in the web but not represent the same thing (e.g. geospatial and data). However,
Sim-Net, while using the NDG formula, does not use the Web to assess co-occurrence of terms. Rather, it uses the NDG formula to measure the number of common and exclusive features of concepts with a formula that has proven to be adapted to networks.

**Definition 1 (View Similarity Sim\_view)** Consider two views \(a\) and \(b\), \(f(a)\) is the number of views that are directly related to \(a\) in the network, plus the ones that are linked to them with generalisation relationships:

\[
f(a) = \left| \bigcup_{i} V_i \right| (a \subseteq \exists R V_i) \lor (V_i \subseteq V_j, a \subseteq \exists R V_j) \lor (V_i \subseteq a)
\]

This set is interpreted as the number of occurrence of \(a\) because the definitions of all the views inside this set contain \(a\). \(f(a, b)\) is the number of views that are related to \(a\) and \(b\):

\[
f(a, b) = f(a) \cap f(b).
\]

SizeNet is the total number of views in the network. We incorporate this distance in a semantic similarity measure which will compare all views of a pair of concepts \(C_1\) and \(C_2\). The distance can be interpreted as a semantic dissimilarity: it is zero when compared elements are the same, and increases when compared elements are different. According to the literature, a network distance can be transformed into a semantic similarity value using an exponentially decaying function (Schwering 2008); following this principle, \(Sim_{\text{view}}(C_1, C_2)\) is given by

\[
Sim_{\text{view}}(C_1, C_2) = \frac{1}{\left| C_1 \right|} \sum_{i=1}^{\left| C_1 \right|} \max_j \left( e^{-\lambda(a, b)\cdot ND(a, b)} \right) \quad \text{with} \quad 1 \leq j \leq \left| C_2 \right|
\]

with \(|C_1|\) the number of views of \(C_1\), \(|C_2|\) the number of views of \(C_2\). We introduce \(\lambda(a, b)\) as a factor that determine the importance of the pair \(a, b\) in the similarity measurement. Each concept has several views, and each view has a set of instances which is a subset of the concept’s instances. We propose that the number of instances of a view with respect to the number of instance of a concept is representative of the importance of this view compared to other views in the semantic similarity assessment. This principle is employed to compute \(\lambda(a, b)\):

\[
\lambda(a, b) = \frac{|C_i^1| \cdot |C_i^2|}{|a^i| \cdot |b^i|}
\]

This formula is simply a ratio of the number of instances of the concepts, and the number of instances of their views.

**Definition 2 (Cross-domain Similarity Sim\_cd)**. Let \(a\) and \(b\) be formal concepts. Consider two concepts referenced to their respective formal concepts:

\[
O_1 : C_1 \xrightarrow{\text{RefTo}} L : a \quad \text{and} \quad O_2 : C_2 \xrightarrow{\text{RefTo}} L : b.
\]
\( f(a) \) is interpreted as the number of concepts of the network that are referenced to \( a \) and SizeNet the total number of concepts in the network. The more \( a \) and \( b \) are similar, the more \( C_j \) and \( C_2 \) are similar:

\[
Sim_{cd}(C_1, C_2) = e^{-ND(a,b)}
\]  

**Definition 3 (Inter-ontology neighbourhood Similarity \( Sim_{ion} \)).** The neighbourhood of concepts across different ontologies can help to identify similar concepts of different domains. Consider a concept \( C_1 \) of \( O_1 \) (Figure 6). \( C_1 \) is linked to concepts of \( O_3 \) and \( O_4 \). The set of concepts from other ontologies to which \( C_1 \) is linked constitute the inter-ontology neighbourhood of \( C_1 \), denoted \( ion(C_1) \). We consider that \( C_1 \) is similar to \( C_9 \) if \( C_9 \) is similar to some concept of \( ion(C_1) \). \( Sim_{ion}(C_1, C_2) \) is given by:

\[
Sim_{ion}(C_1, C_2) = \begin{cases} 
1 & \text{if } C_1 \equiv C_2 \\
\frac{1}{|ion(C_2)|} \sum_{i=1}^{\text{size}(ion(C_2))} e^{-ND(C_i, C_2)} & \text{otherwise}
\end{cases}
\]  

**Figure 6** The inter-ontology neighbourhood of \( C_1 \)

We take \( a \) and \( b \) to be concepts, and \( f(a) \) is the number of concepts of the network that directly related to \( a \), plus their sub-concepts:

\[
f(a) = \left| C_i \right| (a \subseteq \exists R.C_i) \lor (C_i \subseteq C_j, a \subseteq \exists R.C_j) \lor (C_i \subseteq a)
\]
SizeNet is the total number of concepts in the network. Some commonly discussed properties of similarity are the minimality (distance between from a concept to itself is zero), the symmetry (the distance from concept C1 to concept C2 is the same as vice-versa), and the triangle inequality (distance between two concepts C1 and C3 is always smaller than or equal to the distance between C1 and C2 plus the distance between C2 and C3). The NGD is symmetric, respect minimality but not triangle inequality (Ciibrasi and Vitanyi 2007). Sim-Net inherits the minimality and the non-respect of triangle inequality from the NGD. Sim-Net's respect of minimality is consistent with the fact that applying reasoning rules of Table 4 on two concepts having the same features will return the equivalence relation. Also, as discussed in section 3.2, triangle inequality cannot always be assumed for semantic distance. However, Sim-Net is not symmetric since in equation 10 and 13 we have introduced normalization factors that depend on the first concept C1 only (dividing by the number of views and the size of the neighbourhood of C1 respectively). This is consistent with the experiments of Tversky (1977) where it was shown that human do not perceive similarity as symmetric. Also, it is consistent with the fact that semantic relations (equivalence, generalization, specialisation, etc) are not necessarily symmetric.

4.3.3.2 Computing Weights using Domain Similarity

The weights in Sim-Net semantic similarity measure reflect the similarity among domains. If domains of the compared concepts are similar, more weight is given ... The domain is represented by the set of formal views and formal concepts to which a concept or view is referenced. We define the view domain, Dom(V), the concept domain, Dom(C), and the ontology domain, Dom(O). They are included in each other as follows:

\[
\text{Dom}(V) \subseteq \text{Dom}(C) \subseteq \text{Dom}(O) \tag{17}
\]

The ontology domain is the set of formal views and formal concepts to which all concepts and views of the ontology are referenced (Figure 7).

**Definition 4 (Concept Domain Dom(C)).** Consider the ontology sub-graph which starts from the root of the ontology, pass by the concept \(C_i\) and includes all sub-concepts of \(C_i\):

\[
\tau(O : C_i) = \{ C, V | C \subseteq C_i \land C_i \subseteq C, C \subseteq \exists \text{HasView} \} \tag{18}
\]

\(\text{Dom}(C_i)\) is the set of all formal concepts and formal views to which the concepts and views included in \(\tau(O : C_i)\) are referenced:

\[
\text{Dom}(O : C_i) = \left\{ FC, FV | \begin{array}{c}
C \xrightarrow{\text{ReferencedTo}} FC, C \in \tau(O : C_i), \\
V \xrightarrow{\text{ReferencedTo}} FV, V \in \tau(O : C_i)
\end{array} \right\} \tag{19}
\]

This means that we consider that the domain is the set of entities that surround a concept in the ontology. Considering the domain of a concept allows distinguishing two concepts that would have, for instance, the same name and features, but are representing aspects of different realities, for example, a bridge as a road network element or a bridge as a hazard to boat navigation.
**Definition 5 (View Domain $Dom(V_i)$).** $Dom(V_i)$ is similar to $Dom(C_i)$ but it contains only formal views:

$$Dom(O_{i};V_{i}) = \{FV\mid V \xrightarrow{\text{ReferencedTo}} FV, V \in \tau(O_{i};V_{i})\}$$  \hfill (20)\

where $\tau(O_{i};V_{i}) = \{V\mid V \subseteq V_{i} \lor V \subseteq V_{i}\}$. Figure 8 shows the sub-graphs $\tau(O_{i};C_{i})$ of a concept $C_{i}$ and $\tau(O_{i};V_{i})$ of a view $V_{i}$. Their respective domains are defined by the set of formal views and formal concept to which elements of their respective sub-graphs are referenced.

The domain similarity measure compares all formal views and formal concepts contained in the source and the target domains. Only pairs of formal concepts (or formal views) with maximal similarity are retained for the calculation of $Sim_{\text{dom}}$. The similarity between formal concepts $FC_{1}$ and $FC_{2}$ (or formal views) depends on the number of links that separates them from their most specific common subsumer MSCS. The latter is the most specific common parent of $FC_{1}$ and $FC_{2}$.

**Definition 6 (Domain Similarity $Sim_{\text{dom}}$).** To determine the similarity between two domains, we sum up the similarity of each pair of most similar formal concepts (or formal views) of the domain according to the following formula:

$$Sim_{\text{dom}}(Dom(C_{i}), Dom(C_{j})) = \frac{1}{|Dom(C_{i})|} \sum_{i} \max_{j} \left[ \frac{2D_{ij}}{\delta_{i} + \delta_{j} + 2D_{ij}} \right]$$ \hfill (21)
where \( i \) and \( j \) are subscripts for elements of \( \text{Dom}(C_1) \) and \( \text{Dom}(C_2) \) respectively, \( D_{ij} \) is the number of links from MSCS to the top of the view lattice, \( \delta_i \) and \( \delta_j \) are the number of links to MSCS. \( \text{Sim}_{\text{dom}} \) is used to compute weights in Sim-net, according to the next definition.

**Figure 8** Sub-graphs for the definition of \( \text{Dom}(C_4) \) and \( \text{Dom}(V_{42}) \)

**Definition 8 (Sim-Net Weights).** Global semantic similarity of equation 8 contains two weights: \( \omega_{\text{view}} \) for similarity terms expected to be important when comparing concepts of similar domains, and \( \omega_{\text{cd}} \) for the similarity terms that are expected to be important when comparing concepts of different domains. Therefore, we propose that \( \omega_{\text{view}} \) should increase with domain while \( \text{Sim}_{\text{dom}} \) should decrease when domain similarity is low. The formulas for the weights express this:

\[
\omega_{\text{view}}(C_1, C_2) = 1 - \frac{1 - \text{Sim}_{\text{dom}}(C_1, C_2)}{\alpha} 
\]

(22)

\[
\omega_{\text{cd}}(C_1, C_2) = \frac{1 - \text{Sim}_{\text{dom}}(C_1, C_2)}{\alpha} 
\]

(23)

where \( \alpha \) is a factor that can be used to reduce the importance of \( \omega_{\text{cd}} \). If we take \( \alpha = 1 \), we obtain that \( \omega_{\text{cd}} = 1 \) when domains are completely dissimilar. In this case, it could be judged that we don't want to completely remove the influence of the \( \text{Sim}_{\text{view}} \) term on the overall similarity. Therefore, we can increase the value of \( \alpha \). The weights computed with equations 22 and 23 are incorporated into global similarity equation (equation 8) to give the final similarity value. This is the result we get in a static network. In a dynamic, ad hoc, network, the adding of new ontologies may modify the similarities that were previously computed.

4.3.3.3 The Impact of “Ad Hoc” on Sim-Net Similarity Measure
In this section, we examine the impact of a change in the ad hoc network on Sim-Net semantic similarity. When a change is detected, such as the adding of a source to the network or to a punctual cluster of ontologies, it is possible that existing semantic similarity values must be modified. This is not only because we have to compute the new similarity values between concepts of the new ontology with concepts of existing ontologies, but also because semantic similarity depends on inter-ontology neighbourhood. Consequently, the definition of Sim-Net itself is recursive in time: as the network evolves, previous values of similarity must be recomputed because the inter-ontology neighbourhood of concepts is modified. However, frequent changes in the network may require too much computation. As stated by Janowicz et al. (2008), semantic similarity measures are complex and most of the time costly in computation time, so approximation of similarity when changes occur is a most promising approach. We propose an algorithm that account for the behaviour of Sim-Net in the dynamic network. This algorithm is based on the following change-management criterion: when a new ontology \( O_{NEW} \) is added, we use the ontology domain similarity (\( \text{Dom}(O) \)) developed in section 4.3.3.2 to determine if the adding of \( O_{NEW} \) should modify the semantic similarity values that were computed between concepts of two other existing ontologies \( O_i \) and \( O_j \). If \( \text{Dom}(O_{NEW}) \) is dissimilar to \( \text{Dom}(O_i) \), it is less probable that \( O_{NEW} \) will contain concepts that matches with those of \( O_i \). Consequently, the chances that the inter-ontology neighbourhoods of \( O_i \)'s concepts will be modified by the adding of \( O_{NEW} \) are low as well. We deduce that similarity values between \( O_i \) and any other ontology \( O_j \) would not be significantly modified by the adding of \( O_{NEW} \). In this case, we decide not to re-compute those similarities. However, if \( \text{Dom}(O_{NEW}) \) and \( \text{Dom}(O_i) \) were enough similar (more than a given threshold \( \text{ThDOM} \)), we expect that the adding of \( O_{NEW} \) will have a significant impact on the inter-ontology neighbourhoods of \( O_i \)'s concepts, so we decide to re-compute similarity values between \( O_i \) and other ontology \( O_j \), provided that \( \text{Dom}(O_j) \) is also similar to \( \text{Dom}(O_{NEW}) \).

**Dynamic Sim-Net Algorithm**

**Begin Algorithm**

1. Detection of new node \( O_{NEW} \) in the network or cluster
2. Regeneration of the View Lattice:
   - 2.1 Determine new reference concepts;
   - 2.2 Project local concept to new reference concepts;
   - 2.3 Identification of new formal concepts;
   - 2.4 Identification of new inheritance relationships and generation of the lattice;
   - 2.5 Determine new reference properties and values;
   - 2.6 Project local views to reference properties and values;
   - 2.7 Identification of new formal views;
   - 2.8 Identification of new inheritance relationships between formal views and expansion of the lattice
3. For each pair \( (O_{NEW}, O_i) \) of the network or cluster:
   - 3.1 Identify ontology domain \( \text{Dom}(O_i) \) in view lattice;
   - 3.2 Compute ontology domain similarity \( \text{Sim}_{DOM}(O_{NEW}, O_i) \);
   - 3.3 If \( \text{Sim}_{DOM}(O_{NEW}, O_i) > \text{ThDOM} \), compute semantic relationships between \( O_{NEW} \) and \( O_i \)
4. For each \( O_i \) of the network or cluster:
   - 4.1 For each \( O_j \), \( j \neq i \), of the network or cluster:
     - 4.1.1 If concept(s) of \( O_i \) and \( O_j \) have non-disjoint semantic relations with \( O_{NEW} \):
       - 4.1.1.1 Determine the new inter-ontology neighbourhood of those concepts;
       - 4.1.1.2 Re-compute the values in inter-ontology neighbourhood similarities, \( \text{sim}_{ion} \);
       - 4.1.1.3 Include new values of \( \text{sim}_{ion} \) into global semantic similarity

**End Algorithm**

**Figure 9** Dynamic Sim-Net Algorithm
The steps of the algorithm include also the regeneration of the view lattice, which is necessary in order to determine the new ontology domains (step 2). Step 3 verifies if ontology domains are similar enough while step 4 indicates the step for re-computing similarities. Note that we indicated that only the similarity term needs to be re-computed. In fact, the adding of new ontology may also have a slight impact on the values of the weights, since they depend on the domains, which may be slightly modified by the re-organization of the view lattice. The value of the chosen threshold $T_{DOM}$ is also a key factor that determines the efficiency of the algorithm. Experimental testing is required to fix the appropriated value.

5 Application Example

The following illustrates an example of the distinctive properties of Sim-Net. Consider the small ontologies in Figure 9. Ontologies O1 and O2 both describe hydrographic network. Ontologies O3 and O4 describe different domains. In the four ontologies we have concepts “watercourse”, “stream”, “water” and “flooding hazard” that represent different points of view on a watercourse.

![Ontology examples](image)

**Figure 10** Ontology examples

Watercourse and stream are constrained by the logic rules of Table 5, which allow extracting the views defined in second row. Table 6 gives the similarity of domain for the compared concepts, and Table 7 gives the semantic relationships and semantic similarity values obtained with Sim-Net.
<table>
<thead>
<tr>
<th>Logic Rules</th>
<th>Extracted Views</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>O1:</strong></td>
<td><strong>O1:</strong></td>
</tr>
<tr>
<td>r1: [depth(O1:watercourse) ≤ 20 m]</td>
<td>View1(O1:watercourse) :</td>
</tr>
<tr>
<td>→ [category(O1:watercourse) = intermittent]</td>
<td>category = intermittent</td>
</tr>
<tr>
<td>r2: [depth(O1:watercourse) &gt; 20 m]</td>
<td>depth ≤ 20 m</td>
</tr>
<tr>
<td>→ [category(O1:watercourse) = stable]</td>
<td>spatial extent = moving region</td>
</tr>
<tr>
<td>r3: [category(O1:watercourse) = intermittent]</td>
<td>connect: O1:waterbody</td>
</tr>
<tr>
<td>→ [spatial extent(O1:watercourse) = moving region]</td>
<td></td>
</tr>
<tr>
<td>r4: [category(O1:watercourse) = stable]</td>
<td>View2(O1:watercourse) :</td>
</tr>
<tr>
<td>→ [spatial extent(O1:watercourse) = region]</td>
<td>category = stable</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inferred rules:</strong></td>
<td></td>
</tr>
<tr>
<td>ir5: [depth(O1:watercourse) &gt; 20 m]</td>
<td></td>
</tr>
<tr>
<td>→ [spatial extent(O1:watercourse) = region]</td>
<td></td>
</tr>
<tr>
<td>ir6: [depth(O1:watercourse) ≤ 20 m]</td>
<td></td>
</tr>
<tr>
<td>→ [spatial extent(O1:watercourse) = moving region]</td>
<td></td>
</tr>
<tr>
<td><strong>O2:</strong></td>
<td><strong>O2:</strong></td>
</tr>
<tr>
<td>r1: [depth(O2:stream) ≤ 10 m] →</td>
<td>View1(O2:stream) :</td>
</tr>
<tr>
<td>[class(O2:stream) = disappearing stream]</td>
<td>class = disappearing stream</td>
</tr>
<tr>
<td>r2: [depth(O2:stream) &gt; 10 m] →</td>
<td>depth ≤ 10 m</td>
</tr>
<tr>
<td>[class(O2:stream) = river, rapids]</td>
<td>spatial extent = moving surface</td>
</tr>
<tr>
<td>r3: [class(O2:stream) = disappearing stream]</td>
<td>connect: O2: lake</td>
</tr>
<tr>
<td>→ [spatial extent(O2:stream) = moving surface]</td>
<td></td>
</tr>
<tr>
<td>r4: [class(O2:stream) ≠ disappearing stream]</td>
<td>[spatial extent(O2:stream) = surface]</td>
</tr>
<tr>
<td>→ [spatial extent(O2:stream) = surface]</td>
<td>View2(O2:stream) :</td>
</tr>
<tr>
<td><strong>Inferred rules:</strong></td>
<td></td>
</tr>
<tr>
<td>ir5: [depth(O2:stream) &gt; 10 m] →</td>
<td></td>
</tr>
<tr>
<td>[spatial extent(O2:stream) = surface]</td>
<td></td>
</tr>
<tr>
<td>ir6: [depth(O2:stream) ≤ 10 m] →</td>
<td></td>
</tr>
<tr>
<td>[spatial extent(O2:stream) = moving surface]</td>
<td></td>
</tr>
</tbody>
</table>

By demonstrating the comparison of “O1: watercourse” and “O2: stream” we are illustrating the contribution of the view paradigm. View1(O1: watercourse) defines that an intermittent watercourse is a watercourse which depth is less than 20 meters and which spatial extent is a moving region, and in the same way view1(O2: stream) gives the semantic of “disappearing stream”. Only when these views are extracted we can state that “O2: disappearing stream” is a kind of “O1: intermittent watercourse”, and doing the same for other views, find that “O2: stream” is more specific than “O1: watercourse”. Without extracting views, we find that “O1: watercourse” overlaps with “O2: stream” since we cannot compare “O1: intermittent watercourse” with “O2: disappearing stream”, and find them dissimilar. Without using views we have a lower semantic similarity value (0,48) than when using views (0,71). One important
property of Sim-Net is its ability to consider spatial relationships and temporal relationships, in
collection to feature and geometric models. A feature model would find no similarity between
ConnectWaterbody and ConnectLake since it allows no partial match. Sim-Net separate
“connect” from “waterboby” and “lake” in the view lattice, finds that lake is a kind of waterbody
and deduces that according to this property “O2: stream” is more specific than “O2: watercourse”.
Other examples show the specific properties of Sim-Net. The concepts “O1: watercourse” and
“O3: water” do not have any common properties, however with cross-domain reasoning rules we
find that “O1: watercourse” is a cross-domain specialisation of “O3: water” and their semantic
similarity value is computed relatively to the domain similarity. Finally, “O1: watercourse” and
“O4: flooding hazard” have no common properties and are not referenced to a common formal
concept since they are lexically different. Common existing semantic similarity models would
not find any similarity between these concepts. With similarity between inter-ontology
neighbourhoods we find that “O4: flooding hazard” overlaps with one concept in the inter-
ontology neighbourhood of “O1: watercourse”, i.e. “O3: water”. Therefore Sim-Net detects
semantic similarity between “O4: flooding hazard” and “O1: watercourse” with the help of a
third ontology O3 which acts as an intermediary ontology between O1 and O4. Furthermore
because domains similarity between “O4: flooding hazard” and “O1: watercourse” are dissimilar,
high weight is given to the similarity between inter-ontology neighbourhoods.

Table 6  Similarity between concept domains of Figure 9

<table>
<thead>
<tr>
<th>Concepts Domain Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SimDOM (O1:watercourse, O2: stream)</td>
</tr>
<tr>
<td>SimDOM (O1:watercourse, O3: water)</td>
</tr>
<tr>
<td>SimDOM (O3:water, O4: flooding hazard)</td>
</tr>
<tr>
<td>SimDOM (O1:watercourse, O4: flooding hazard)</td>
</tr>
</tbody>
</table>

Table 7 Similarity between concepts of Figure 9

<table>
<thead>
<tr>
<th>Sim-Net semantic relationships</th>
<th>Semantic similarity value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_1 : watercourse \rightarrow O_2 : stream$</td>
<td>$SN = 0,71$ (with views) $SN = 0,48$ (without views)</td>
</tr>
<tr>
<td>$O_1 : watercourse \subseteq O_3 : water$</td>
<td>$SN = 0,52$</td>
</tr>
<tr>
<td>$O_3 : water \cap O_4 : flooding hazard$</td>
<td>$SN = 0,38$</td>
</tr>
<tr>
<td>$O_1 : watercourse \cap O_4 : flooding hazard$</td>
<td>$SN = 0,42$</td>
</tr>
</tbody>
</table>

This demonstration shows some properties of Sim-Net in a static network. Now consider the
adding of a new ontology O_NEW (Figure 11) to show the properties of Sim-Net in a dynamic
network.
The ontology domain similarities between $O_{NEW}$ and the four other ontologies are shown in Table 8. We have to choose an appropriated threshold $Th_{DOM}$. For the purpose of the demonstration, we choose $Th_{DOM} = 0.30$ on the basis of previous results: concepts which domain similarity, according to Table 6, was at least 0.30 have a semantic similarity of more than 0.5 (Table 7). Note that this threshold is used only as an indication, and more investigation and extensive testing are required to choose an appropriated value. In this case, this would discard $O3$ and $O4$ from the re-computation of similarity values. Consequently, only the similarity between $O1$: watercourse and $O2$: stream is affected by the change. The new value of similarity (Table 9) increases compared to 0.72 because inter-ontology neighbourhood similarity was increase by the new ontology.

Table 8  Similarity between ontology domains

<table>
<thead>
<tr>
<th>Ontologies</th>
<th>Domain Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Sim_{DOM} (O1, O_{NEW})$</td>
<td>0.65</td>
</tr>
<tr>
<td>$Sim_{DOM} (O2, O_{NEW})$</td>
<td>0.48</td>
</tr>
<tr>
<td>$Sim_{DOM} (O3, O_{NEW})$</td>
<td>0.20</td>
</tr>
<tr>
<td>$Sim_{DOM} (O4, O_{NEW})$</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Table 9  New semantic similarity value

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Domain Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Sim_{DOM} (O1: watercourse, O2: stream)$</td>
<td>0.80</td>
</tr>
</tbody>
</table>

6 Conclusion and Future Work

While ad hoc networks of geospatial databases are becoming widespread, most existing semantic similarity models are targeted at comparing pairs of concepts in isolation. We have proposed a new semantic similarity model that contributes to two major issues. On the one hand, it addresses some requirements related to the assessment of semantic similarity in ad hoc network. In this regard, the new comparison criteria provided by Sim-Net include the comparison of inter-ontology neighbourhood, and the comparison of domains. We also study the behaviour of Sim-Net when changes in the network occur, and propose that domain similarity be used as a criterion for deciding whether the adding of new ontologies will modify the existing similarity values. On
the other hand, Sim-Net also copes with a rich representation of geospatial concepts were dependencies between properties can be represented and used to extract different views of a concept. The approach of assessing semantic similarity with Sim-Net is based on the establishment of a common semantic reference system built with the Formal Concept Analysis method. Sim-Net is also based on DL, which makes it readily adaptable to knowledge representation in existing ontologies as OWL is one of the most recommended ontology languages according to W3C.

Based on these findings, our research continues forwards as new research issues were raised during the development of Sim-Net. A first research issue that is specific to ad hoc network is the propagation of error in semantic similarity assessment. This is likely to occur since in Sim-Net the similarity between two concepts depends on their similarity with neighbourhood concepts. In general, propagation of error can occur because concepts were ill-defined (low quality input), or because the semantic similarity measure employed is not suitable to the concept representation, or finally because results are inconsistent. This is an issue we have already explored in our previous work on elements of semantic quality (Bakillah et al. 2008). Therefore, how this framework can be incorporated with Sim-Net is a promising research avenue towards the resolution of the issue of error propagation. A second issue we planned to explore in future work is the problem of selecting which concepts in the network have to be compared. As it is not required that all pairs of concepts of several ontologies be compared to answer a given query, the question is how to propagate this query to relevant concepts and in the appropriate order. In this regard, we believe that Sim-Net could play a major role by indicating the relevant concepts. Finally, future work will include the comprehensive testing of Sim-Net in ad hoc network.

Acknowledgments

The authors wish to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References


Bédard Y, Bernier E, 2002 Supporting Multiple Representations with Spatial View Management and the Concept of "VUEL". In Joint Workshop on Multi-Scale Representations of Spatial Data, ISPRS WG IV/3, ICA Com. on Map Generalization, July 7-8, Ottawa, Canada


Debrauwer L 1998 Des vues aux contextes pour la structuration fonctionnelle de bases de données à objets en CROME. Doctoral thesis, University of sciences and technologies, Lile, France


Rodriguez MA, Egenhofer MJ, 2003 Determining Semantic Similarity among Entity Classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering* 15: 442–56


Smith B, Mark D, 1998 Ontology and Geographic Kinds. *International Symposium on Spatial Data Handling*, Vancouver, Canada: 308-320


