

Locally Constrained Ontologies

Karlis PODNIEKS

Institute of Mathematics and Computer Science, University of Latvia
Raina bulvaris 29, Riga, LV-1459, Latvia

`karlis.podnieks@mii.lu.lv`

Abstract. In 2014, Barzdins, Rencis and Sostaks introduced *granular ontologies* as a specific organization of databases allowing for extremely fast processing of *ad hoc* queries, and proved the following Granularity Theorem: Consider an ontology represented by graphical means of a UML class diagram. Then, under certain restrictions on association multiplicity constraints, this ontology is granular, if and only if it is a tree ontology. (In a *tree ontology*, associations and classes form a tree, and have the multiplicity 1..1 in the direction to root class.) The possibility of removing the restrictions was formulated as an open problem.

The present paper solves this problem. It appears that the principal cause of the “tree phenomenon” is the *local character* of ontology constraints expressed by graphical means of UML class diagrams (roughly, each of such constraints involves at most one association). In the paper, properties of locally constrained ontologies (“locality phenomena”) are explored, and Generalized Granularity Theorem is proved, showing that in the Granularity Theorem, all restrictions to multiplicity constraints can be removed.

Keywords: ontology, UML class diagram, first order language, locally constrained ontology, granular ontology, tree ontology.

1 Introduction

The inspiration of the present paper comes from the so-called *granular ontologies* that were introduced in Barzdins et al. (2014) as a specific organization of databases allowing for extremely fast processing of *ad hoc* queries.

The idea is as follows: in a granular database, instances of the entities are distributed among disjoint and disconnected subsets called *granules*. Not every database can be converted into a granular database, but in cases when this is possible the processing of *ad hoc* queries can be implemented very efficiently. A hospital database in which one can easily “disconnect” the data of different patients, thus obtaining natural granules, is a typical example. The necessary transformation procedure is described in Barzdins et al. (2014).

The granularity property is meant not as a contingent property of some particular database states. Granularity must follow from the database organization, or, in terms of the present paper, the *granularity property must follow from the constraints of the ontology*. Only in this way can we ensure that this property uniformly holds in all possible database states.

Thus, granules must somehow be already labeled in the ontology. This is achieved by declaring one of the ontology classes as the *master class*, and by requiring that each granule contains exactly one object from the master class and that all the other objects belonging to the granule are connected to this master object via association chains of the ontology. The second requirement: any two granules must be totally disconnected, meaning that no two objects belonging to different granules are connected via association chains of the ontology (more precisely: via those association chains that do not cross the master class).

How to verify, is an ontology granular, or not? The following kind of ontologies is easily verified as granular:

Definition 1. A restricted UML ontology (Definition 2 below) is called a **tree ontology** (star ontology – in Barzdins et al. 2014) iff in O :

- a) no more than one association connects two classes;
- b) associations form a *tree*; and
- c) all associations are *functional and mandatory* (multiplicity 1..1) in the direction to the root class.

The authors of Barzdins et al. (2014) have proven the following

Theorem 1. (Granularity Theorem) Consider a restricted UML ontology which does not contain any other constraints than the standard multiplicities (0..1, 1..1, 0..*, 1..*). Then, this ontology is granular iff it is a tree ontology.

The possibility of removing the restriction involving the standard multiplicities was formulated as an open problem.

The inspiration of the present paper was to improve the Granularity Theorem as far as possible (and, as a consequence, to solve the above-mentioned open problem). It appears that the principal cause of the “tree phenomenon” is the *local character* of most ontology constraints that can be expressed by the graphical means of UML class diagrams (see OMG 2016).

So, let us consider a certain subtype of the ontologies that can be defined by using only the graphical means of UML class diagrams. To make our explanations precise, let us introduce the following

Definition 2. An ontology is called a **restricted UML ontology** iff it consists of classes, attributes and associations, in which:

- a) association inclusion (super- and sub-associations) is not allowed;
- b) associations can be attached only to bottom-level classes;
- c) only binary associations are allowed, self-loop associations are not allowed;
- d) each object belongs to some bottom-level class, and bottom-level classes are pairwise disjoint;
- e) no other constraints are allowed.

Remark 1. Thus, in restricted UML ontologies: f) class inclusion, i.e. super- and sub-classes are allowed; g) any multiplicity constraints $m : n$ are allowed at the association ends, where m, n are natural numbers $0, 1, 2, \dots$, and n can be also $*$ (infinity).

The restriction (c), in fact, “does not restrict”. Indeed, to eliminate n -ary associations and self-loop associations, one can introduce intermediate classes (see Berardi et al. 2005, Section 7.1.2).

But (a, b, d) are serious restrictions. They were introduced because the methods used in Barzdins et al. (2014) and in the present paper do not work for ontologies violating these restrictions. They make the structure of the corresponding UML class diagrams extremely clear:

Definition 3. UML class diagram representing a restricted UML ontology, can be expressed as the *union* of two independent parts: a) **class inclusion diagram** containing only classes (and no associations); and b) **association diagram** containing only bottom-level classes and associations.

An ontology represented by a UML class diagram can be converted, in a natural and straightforward way, into a set of axioms (representing *ontology constraints*) in some appropriate first order language. The axioms obtained in this way from a *restricted* UML ontology, roughly, appear to be *local* in the following sense: each involves *at most one binary association* predicate.

So, let us switch to the ontologies represented as sets of axioms in first order languages (first order ontologies). And let us call such ontologies *locally constrained*, if (roughly) each of their axioms involves at most one binary association predicate (for a precise definition – see Definition 13 below).

Intuitively, when defining a locally constrained ontology, one cannot coordinate two or more different associations by setting common conditions on them. This implies certain “locality phenomena” established in the paper.

For example (Theorem 2 below), assume that the axioms of a locally constrained ontology imply that some association chain does not connect any objects. Then these axioms imply that, in this chain, *at least one association is empty*.

Or (Theorem 4), assume that the axioms of a locally constrained ontology imply that some association chain is functional (i.e, it connects each left-hand side object to at most one right-hand side object). Then these axioms imply that, in this chain, *all associations are functional*.

Formalization and establishing of the existence of *locality phenomena* in locally constrained ontologies represents one of the novelties of the present paper.

The Generalized Granularity Theorem (Theorem 5) proved in Section 11 also represents one of such locality phenomena. As an application of it, the following improved version of Theorem 1 was obtained:

(Granularity Theorem 2, Theorem 6 below) *Consider any restricted UML ontology O that is consistent in the sense that it allows for a model, in which all classes and associations are non-empty. Then, O is granular iff it is a tree ontology.*

Thus, in the Granularity Theorem, all restrictions to multiplicity constraints can be removed. This solves the above-mentioned open problem formulated in Barzdins et al. (2014).

The present paper is purely *model-theoretic*. Computational complexity issues of processing queries to databases described by locally constrained schemas will be considered in a separate paper.

The method used in proofs below represents a piece of “standalone mathematics”: a specific problem is solved by a method designed especially for this purpose. Proofs are lengthy, but largely straightforward, the only non-trivial points being:

- a) Permutation Lemma (Section 5);
- b) class-diagram-independent definitions of association chains, nets, and ontology diagrams (Section 6);
- c) the introduction of the notion of locally constrained ontologies (Section 7).

Sections 3, 4 contain the definition of first order ontologies, an explanation of their formal relationship to UML ontologies, and the necessary facts from the first order model theory.

In Section 5, Permutation Lemma, the main technical tool of the paper, is proved. Section 6 contains generalized, class-diagram-independent formal definitions of associations nets, association chains, and ontology diagrams.

In Section 7, locally constrained ontologies are defined and the first locality phenomenon (Theorem 2 about connecting association chains) is established. In Section 8, the theoretical possibility of association singularity phenomenon is considered and the second locality phenomenon (Theorem 3 about singular association chains) is established. In Section 9, the third locality phenomenon (Theorem 4 about functional association chains) is established.

In Section 10 a generalized formal definition of granular ontologies is discussed, and in Section 11 the main result of the paper – the Generalized Granularity Theorem (Theorem 5) is proved. Theorems 2, 3, 4 were used in the proof of this theorem.

In Section 12, the finite model versions of the results are obtained.

2 Related Work

The translation procedure of UML class diagrams to sets of first order formulas, called in Section 3 “the natural first order translation”, represents a subset of the procedure introduced in Berardi et al. (2005).

The singularity phenomenon considered in Section 8 extends the concept of strong satisfiability introduced for UML class diagrams in Lenzerini and Nobili (1990).

The finite model setting considered in Section 12, was explored for UML class diagrams in Balaban and Maraee (2013).

Granular ontologies, considered in Sections 1, 10 were introduced in Barzdins et al. (2014).

Very specific kinds of local ontology constraints – “local value constraints”, “local cardinality constraints”, etc. were considered in Ding and Peng (2004) and Klein (2004).

The term “Local Constraint Based Analysis (LCBA)” (Ozdamar and Ulusoy 1994), used in operational research, has a different meaning from the one used in the present paper.

3 First Order Ontologies

The overall picture is as follows. Let us consider an ontology O that refers to classes (or, concepts) OBJ, C_1, C_2, \dots ; to binary associations R_1, R_2, \dots ; to attributes (or, properties) A_1, A_2, \dots ; and to data types (or, value domains) VAL, D_1, D_2, \dots .

To formalize such an ontology, we start by defining a first order language $L[O]$ consisting of:

- a) unary predicate constants $OBJ(x), C_1(x), C_2(x), \dots, VAL(x), D_1(d), D_2(d), \dots$;
- b) binary predicate constants $R_1(x, y), R_2(x, y), \dots, A_1(x, d), A_2(x, d), \dots$; and
- c) the equality predicate $x = y$.

The formula $C_i(x)$ is intended to mean “object x belongs to class C_i ”, the formula $D_i(d)$ means “object d is a data value of type D_i ”, the formula $R_i(x, y)$ means “object x is connected to object y by association R_i ”, and the formula $A_i(x, d)$ means “object x possesses d as a value of attribute A_i ”.

Predicates OBJ, VAL play a specific role: $OBJ(x)$ is intended to mean “ x is an object”, and $VAL(d)$ means “ d is an attribute value”. Thus, all our ontologies must contain the following axioms:

- $\forall x(\neg(OBJ(x) \wedge VAL(x)))$ (objects are disjoint from values);
- $\forall x(C_i(x) \rightarrow OBJ(x))$ (classes consist of objects only);
- $\forall x(D_i(x) \rightarrow VAL(x))$ (data types consist of values only);
- $\forall x, y(R_i(x, y) \rightarrow OBJ(x) \wedge OBJ(y))$ (associations can connect only objects);
- $\forall x, y(A_i(x, y) \rightarrow OBJ(x) \wedge VAL(y))$ (attributes can connect only objects and only to values).

Let us call this set of axioms SET_1 . All of them are local in the sense of Section 7 below.

Constraints of the ontology O (usually expressed, for example, by the graphical means of UML class diagrams, as OCL formulas etc.) are now formulated as axioms in the language $L[O]$. For example, to define a restricted UML ontology, the following kinds of axioms are used:

- $\forall x(\neg(C_i(x) \wedge C_j(x)))$ (disjoint classes);
 - $\forall x(C_i(x) \rightarrow C_j(x))$ (subclass of);
 - $\forall x(C_i(x) \rightarrow C_j(x) \vee \dots \vee C_k(x))$ (complete coverage);
 - $\forall x, y(R_i(x, y) \rightarrow C_j(x) \wedge C_k(y))$ (association typing);
 - $\forall x(C_i(x) \rightarrow \exists y R_j(x, y))$;
 - $\forall y(C_i(y) \rightarrow \exists x R_j(x, y))$ (mandatory participation);
 - $\forall x, y_1, y_2(R_i(x, y_1) \wedge R_i(x, y_2) \rightarrow y_1 = y_2)$ (association functionality);
 - $\forall x, d(A_i(x, d) \rightarrow C_j(x) \wedge D_k(d))$ (attribute typing);
 - $\forall x(C_i(x) \rightarrow \exists d A_j(x, d))$ (mandatory attribute);
 - $\forall x, d_1, d_2(A_i(x, d_1) \wedge A_i(x, d_2) \rightarrow d_1 = d_2)$ (attribute functionality).
- Multiplicity* constraints can be easily expressed as well, for example, $\forall x(A_m(x) \rightarrow \exists y_1, y_2(R_i(x, y_1) \wedge \forall y(R_i(x, y) \rightarrow y = y_1 \vee y = y_2)))$ (multiplicity 1..2).

Let us call this type of axioms SET_2 . All of them are local in the sense of Section 7 below.

Description logics allow to define, in particular, ontologies containing transitive associations. The transitivity property can be expressed as the following local axiom:

$$\forall x, y, z (R_i(x, y) \wedge R_i(y, z) \rightarrow R_i(x, z)).$$

Association inclusion (prohibited in restricted UML ontologies) would require a non-local axiom:

$$\forall x, y (R_i(x, y) \rightarrow R_j(x, y)).$$

Definition 4. In the ontology O , the association R' is called the **inverse** of the association R iff O contains the axiom $\forall x, y (R'(x, y) \leftrightarrow R(y, x))$. Let us call these **inversion axioms**.

In practice, if some ontology contains an association, then one can naturally use its inverse association as well. So, let us assume that all ontologies contain, with each association, its inverse association, and the corresponding inversion axiom.

Definition 5. Let us define a **first order ontology** as a *first order theory* in some first order language of the kind described by the above items (a, b, c). In particular, the axioms of the first order ontology include the axioms of the classical first order logic with equality, the axioms of SET_1 , and inversion axioms for all associations.

Remark 2. Throughout the present paper, we call first order ontologies just *ontologies*, and “association” always means “binary association”.

Definition 6. Consider a restricted UML ontology O . In an obvious way (see the axiom examples above), one can transform the constraints represented in the UML class diagram of O into first order formulas. The obtained first order ontology O' is called the **natural first order translation** of O .

The translation procedure of UML class diagrams to sets of first order formulas, defined above, represents a subset of the procedure introduced in Berardi et al. (2005).

Remark 3. The axioms of a natural first order translation (of a restricted UML ontology) all belong to SET_2 , except the axioms of logic, the axioms of SET_1 , and inversion axioms.

4 Model Theory

Now, we can apply to ontologies the usual first order model theory (see any textbook in mathematical logic, for example, Mendelson 2009). We can consider *interpretations* of ontology languages and *models* of ontologies (i.e., interpretations in which the axioms of the ontology are true). We will consider only the so-called *normal interpretations* and *normal models* in which the equality predicate $x = y$ is interpreted as the literal equality of domain objects (and not as some equivalence relation between them).

For example, according to *Model Existence Theorem* (Mendelson 2009, Section 2.8, Proposition 2.26), if some closed formula F does not follow from the axioms of the ontology O , then there exists a (finite, or countable infinite) normal model of O in which F is false.

Remark 4. Throughout the present paper, the terms “interpretation” and “model” always mean *normal* interpretations and models.

Remark 5. *Infinite models* must be allowed here (as they were allowed in Kaneiwa and Satoh 2010), or else the first order model theory would not be applicable.

Less elegant *finite model* versions of most of the results are also possible. To formulate these versions, one must replace, for example, “a closed formula F does not follow from the axioms of the ontology O ” with “a closed formula F is false in some finite model of the ontology O ”. For details see Section 12.

If we think of ontologies as *database schemas*, then we can think of ontology models as the corresponding databases – more precisely – as *database states*. According to Model Existence Theorem, if some ontology O , as the first order theory, is consistent, then there is a (finite, or countable infinite) normal model of O ; i.e., there is a (finite, or countable infinite) database state that satisfies the axioms of O .

5 Permutation Lemma

For the proofs below we will need the possibility to change ontology models by re-interpreting some of the associations in order to obtain models with some desired additional properties. The main tool serving this purpose will be the following very technical lemma.

The idea of the lemma is as follows. Imagine, we have a model J of an ontology O , and $D[J]$ is the domain of J (i.e., the set of all its objects). Let us consider a *permutation* h of $D[J]$ (i.e., a bijection of $D[J]$ onto itself) that does not affect interpretations of classes and associations, with the only exception - the association R is the only one that is somewhat re-interpreted (to add some desired new properties to the interpretation J). Permutation Lemma states that in this way the truth values of closed formulas will not be affected, and so, the “permuted” interpretation will remain a model of O .

Note that this lemma, when transforming a model, does not add new objects to it; so, a finite model would be transformed into a finite one.

More precisely, permutation h does not affect the interpretation of class C , or association R iff for any objects $x, y \in D[J]$:

$$C_J(h(x)) \leftrightarrow C_J(x);$$

$$R_J(h(x), h(y)) \leftrightarrow R_J(x, y).$$

Lemma 1. (Permutation Lemma) *Let us consider a first order ontology O , and some normal model J of it. Let us denote by $D[J]$ the domain of J (i.e., the set of all its objects).*

Let us consider any permutation h of $D[J]$ (i.e., a bijection of $D[J]$ onto itself) that does not affect class interpretations.

And finally, let us select, among the associations of O , some subset \mathbf{R} . Let us consider a modified interpretation J_1 that differs from the model J only by “permuted” interpretations of associations belonging to \mathbf{R} and their inverses: if $R \in \mathbf{R}$ then for all objects $x, y \in D[J]$, $R_{J_1}(x, y)$ is defined as true in J_1 iff $R_J(h(x), h(y))$ is true in J . The inverse of R is re-interpreted correspondingly.

Then:

a) If some formula F does not contain associations of the set \mathbf{R} , or their inverses, then F is true in J_1 iff F is true in J .

b) If some closed formula F consists only of equalities, class predicates, and association predicates from the set \mathbf{R} only and their inverses, then F is true in J_1 iff F is true in J .

Proof. (a) is obvious.

(b) follows immediately from a somewhat more general assertion:

b') If some formula $F(x, y, z, \dots)$ consists only of equalities, class predicates, and associations from the set \mathbf{R} only and their inverses, then for all objects $x, y, z, \dots \in D[J]$, $F_{J_1}(x, y, z, \dots)$ is true in J_1 iff $F_J(h(x), h(y), h(z), \dots)$ is true in J .

Let us prove (b') by induction on the structure of F .

1. Atomic formulas.

Equality. Since h is an injection, and J, J_1 are normal interpretations, then for any objects, $x = y$ is true in J_1 (and J) iff $h(x) = h(y)$ is true in J .

Class predicates. Since h does not affect class interpretations (i.e., h may permute objects only within class interpretations), for any class predicate C and any object $x \in D[J]$: $C_{J_1}(x)$ is true in J_1 iff $C_J(h(x))$ is true in J , iff $C_J(x)$ is true in J .

Attribute predicates. Formula F does not contain attribute predicates.

Association predicates from the set \mathbf{R} . If $R \in \mathbf{R}$, then for all objects $x, y \in D[J]$, $R_{J_1}(x, y)$ is defined as true in J_1 iff $R_{J_1}(h(x), h(y))$ is true in J .

Inverses of the association predicates from the set \mathbf{R} . If R' is inverse to some $R \in \mathbf{R}$, then for all objects $x, y \in D[J]$, $R'_{J_1}(x, y)$ is defined as true in J_1 iff $R'_{J_1}(h(x), h(y))$ is true in J .

2. Logical connectors.

If (b') is true for F, G , then, obviously, (b') is true for $\neg F; F \wedge G; F \vee G; F \rightarrow G$ as well.

3. Quantifiers (we consider only $\forall x$, since $\exists x$ is covered by $\neg \forall x \neg$).

Consider the formula $\forall x G(x, y, z, \dots)$. By the induction assumption, for all objects $x, y, z, \dots \in D[J]$, $G_{J_1}(x, y, z, \dots)$ is true in J_1 iff $G_J(h(x), h(y), h(z), \dots)$ is true in J . Since h is a surjection, $h(x)$ runs over the entire domain $D[J]$. Hence, for any $y, z, \dots \in D[J]$, $\forall x G(x, y, z, \dots)$ is true in J_1 iff $\forall x G(x, h(y), h(z), \dots)$ is true in J . \square

6 Proper Association Nets and Ontology Diagrams

In an ordinary situation, we would say that the associations R, R' form a *proper chain*, if R connects the objects of some class C to the objects of some class C' , and R' connects the objects of C' to the objects of some C'' , and the three classes are pairwise disjoint. Similarly, each UML class diagram represents a proper association *net*.

We propose generalized versions of these notions. The corresponding definitions do not refer to class diagrams, so, they look somewhat "abstract". However, such a level of generality seems appropriate for conveniently formulating lemmas and theorems in the present paper.

For the rest of this section, let us fix some first order ontology O , and a model M of it. We will speak about "objects" meaning members of the domain $D[M]$. We will

speak about “classes” and “associations” meaning, correspondingly, *interpretations* of classes and associations of O in the model M .

Let us say that (in the model M) an association R is *bipartite* iff (in M) its left-hand-side projection (domain: $dom(R) = \{x|\exists yR(x,y)\}$) is disjoint from its right-hand side projection (range: $rng(R) = \{y|\exists xR(x,y)\}$). In other words, $dom(R) \cap rng(R) = \emptyset$ is true (in M). For example, in UML class diagrams, “normal” binary associations that connect two disjoint classes, are bipartite (in all models) in the sense of this definition.

(In the model M) two objects x, y are called *compatible* with respect to classes of O iff x, y belong to (the interpretations of) the same classes of O . Formally: iff $C(x) \leftrightarrow C(y)$ is true (in M) for all classes C of O . Thus, the condition “ x, y are compatible with respect to classes of O ” can be expressed as a formula in the language of O .

To simplify the language, from now on, let us stop to indicate the model M in our definitions.

Let us say that some *set* S of objects is compatible with respect to classes of O iff any two elements of S are compatible with respect to classes of O . This means, in fact, that for any class C of O , either $S \subseteq C$, or $S \cap C = \emptyset$ is true (in M , of course).

Similarly, let us say that a *collection* Σ of sets of objects is compatible with respect to classes of O iff the union of Σ is compatible with respect to classes of O .

Let us say that an *association* R is compatible with respect to classes of O iff $dom(R)$ is compatible with respect to classes of O , and $rng(R)$ is compatible with respect to classes of O (but, of course, $dom(R)$ and $rng(R)$ need not to be compatible with each other). The condition “ R is compatible with respect to classes of O ” can be expressed as a formula in the language of O .

Note. In the sense of this definition, in a UML class diagram, an association attached to a super-class would not be compatible with respect to lower level classes. This represents one of the reasons, why, in restricted UML ontologies, such associations are not allowed.

In the above definitions, compatibility with respect to classes of O is allowed to be a *contingent property* of particular models of O . For example, it may happen that for a particular compatible association R and a particular class C , in one of the models, $dom(R) \subseteq C$, but in another model, $dom(R) \cap C = \emptyset$.

In the lemmas and theorems below, we will need no more than this notion of contingent compatibility. Only in some special situations, will we need the following more strict definition in which the exact position of the set S with respect to each particular class C must be determined by the axioms of the ontology.

Definition 7. In a first order ontology O , a set S defined by some formula $S(x)$, is called **determined** with respect to classes of O iff for every class C , the axioms of O imply either the formula $S \subseteq C$, or the formula $S \cap C = \emptyset$.

In particular, an association R is called *determined* with respect to classes of O iff for every class C , the axioms of O imply either the formula $dom(R) \subseteq C$, or the formula $dom(R) \cap C = \emptyset$, and either the formula $rng(R) \subseteq C$, or the formula $rng(R) \cap C = \emptyset$.

In a UML class diagram, *association net* means any subgraph of the diagram. Let us define the corresponding notion in a diagram-independent setting.

In a “non-proper” association net, associations are simply placed on arrows of some finite directed graph:

Definition 8. In a first order ontology O , an **associations net** is a finite directed graph *without isolated nodes and self-loop arrows*, with at least one association of O attached to each arrow, if the following conditions hold:

- a) the directions of arrows and associations attached to them coincide;
- b) if an association is attached to some arrow, then neither it nor its inverse association is attached to another arrow.

Remark 6. Thus, association nets do not contain isolated nodes and self-loop arrows.

In a *proper* association net, associations must be coordinated, like as they would be coordinated in a UML class diagram. The following definition will be justified by Propositions 2, 3 and Lemma 2 below:

Definition 9. The association net is called a **proper associations net** iff the following conditions hold:

- c) the attached associations all are bipartite and compatible with respect to classes of O ;
- d) if the associations R, R' are attached to different arrows, then the sets $dom(R), rng(R), dom(R'), rng(R')$ are pairwise disjoint, except the cases described in the following item (e);
- e) at each node of the graph, *ranges of incoming associations and domains of outgoing associations* are compatible with respect to classes of O .

Remark 7. Thus, proper association nets cannot contain self-loop arrows, but *can contain parallel arrows*; i.e., arrows that have the same starting and ending nodes. And remember that this definition depends of the model M .

Obviously, a single association is proper iff it is bipartite and compatible with respect to classes of O .

If W is a proper association net, then so is any subnet of W .

If we would revert some arrows of a proper association net W , simultaneously replacing the associations attached to them by the corresponding inverses, then we would obtain a proper association net that is, in the obvious sense, “equivalent” to W .

For each particular association net W , the assertion “ W is a *proper* association net” can be expressed as a closed formula in the language of the ontology O . For some nets, this formula may follow from the axioms of O . Then we will say that “the axioms of O imply that the W is a proper association net”.

Definition 10. A sequence of nodes connected by arrows is called a *chain graph*.

Remark 8. The arrows of a chain graph are not supposed to be uni-directed. And a chain graph does not contain loops.

Definition 11. An association net based on a chain graph is called an **association chain**. An association chain is called a **proper association chain** in ontology O iff it is proper as an association net in O .

We will say that an *association chain belongs to the association net* W iff it can be obtained from a non-directed path in W by reverting zero or more arrows and simultaneously replacing the associations attached to them by the corresponding inverses.

Remark 9. An association chain does not contain repetitions, and does not contain simultaneously an association and its inverse.

If an association chain Q is proper, then so is the inverse chain Q^{-1} .

If an association chain Q belongs to some association net W , then Q^{-1} belongs to W as well.

For each particular chain Q , the assertion “ Q connects x to y ” can be expressed by an appropriate formula $Q(x, y)$ in the language of the ontology. For example, if the chain Q consists of 3 associations R^1, R^2, R^3 , then $Q(x, y)$ is the following formula:

$$\exists z_1, z_2 [R^1(x, z_1) \wedge R^2(z_1, z_2) \wedge R^3(z_2, y)].$$

In an interpretation of O , the sequence of objects x, z_1, z_2, y can be regarded as an *instance of the chain* Q .

The following is obvious:

Proposition 1. Consider a proper association chain Q of the ontology O , and a model of O . Then, in any of the instances of Q the objects are pairwise different.

We will say that the chain Q *does not connect* iff $\forall x, y \neg Q(x, y)$. Of course, a single association does not connect iff it is empty.

If the axioms of the ontology O imply the formula $\forall x, y \neg Q(x, y)$, then Q does not connect in all models of O . But, if the axioms of O *do not imply* this formula, then, by Model Existence Theorem, there are models of O in which Q connects some objects.

Now we can introduce a generalized notion of ontology diagrams “in a diagram-independent setting” (note that an ontology contains, with each association, the inverse of it):

Definition 12. An association net D is called an **association diagram** of the ontology O iff,

- a) each association R of O is attached to some arrow of D ;
- b) if R is attached to some arrow of D , then neither R , nor the inverse of R are attached to other arrows of D ;
- c) each arrow of D has an association of O attached to it.

Correspondingly, a proper association net with this property is called a **proper association diagram** of O .

Remark 10. Thus, note again that association diagrams do not contain isolated nodes and self-loop arrows.

Note that if an ontology O allows for a proper association diagram, then all its associations are bipartite and compatible with respect to classes of O .

For each particular association diagram D of O , the assertion “ D is a proper association diagram of O ” can be expressed by a formula in the language of O .

The following two propositions and lemma show that the definitions introduced above are natural. The first proposition can be easily verified. (In fact, worthy of verification are only conditions (c, d, e) in the Definition 9 of proper association nets.)

Proposition 2. *Consider a restricted UML ontology O , defined by a UML class diagram. Let us denote by D the association diagram of O . Obtain the natural first order translation O' of O . Then the axioms of O' imply that D is a proper association diagram of O' .*

Lemma 2. *Let us consider an ontology O , the axioms of which do not imply any class inclusions, i.e. O consists of bottom-level classes only. Assume that the axioms of O imply that:*

- a) each object belongs to some class;*
- b) classes are pairwise disjoint;*
- c) all associations are bipartite and determined with respect to classes of O ;*
- d) there are no self-loop associations (i.e., domain and range of an association cannot be contained in the same class).*

Assume also that the axioms of O do not imply that some of the associations are empty.

Then, the axioms of O imply that the domains and ranges of all associations are each contained in a definite unique class.

Proof. Consider a model of O in which the association R is non-empty (such a model exists, according to Model Existence Theorem). According to conditions (a, b, c, d), in this model $dom(R)$ and $rng(R)$ are contained in exactly two different classes C_1, C_2 respectively. And, since R is determined, the axioms of O imply the formulas $dom(R) \subseteq C_1$ and $rng(R) \subseteq C_2$. According to condition (b), classes C_1, C_2 are the only ones having this property with respect to R . \square

Proposition 3. *In the conditions of Lemma 2, assume that D is a proper association diagram of O . Draw a UML class diagram D_{UML} by depicting:*

- a) all classes of O – as nodes of a directed graph;*
- b) all the associations R of D – as labels attached to the arrows connecting the class nodes in which $dom(R)$ and $rng(R)$ are contained.*

After this, remove from D_{UML} all the isolated nodes; i.e., the nodes to which no arrows are attached.

Then the diagrams D and D_{UML} are isomorphic.

Proof. Consider an association R in D . It is attached to an arrow A that connects two nodes, let us denote these nodes by N_1, N_2 . According to Lemma 2, $dom(R)$ and $rng(R)$ are contained in exactly two different bottom-level classes C_1, C_2 respectively. In D_{UML} , R is attached as a label to the arrow A' connecting the nodes C_1 and C_2 . So, let us relate, in the required isomorphism, the node N_1 – to the node C_1 , the node N_2 – to the node C_2 , and the arrow A – to the arrow A' .

By repeating this step for each association of D , we obtain the required isomorphism between the entire diagrams D and D_{UML} . \square

7 Locally Constrained Ontologies

Which kind of first order ontologies corresponds best to the restricted UML ontologies? In other words, which kind of first order ontologies corresponds best to the ontologies

that are defined by using the graphical means of UML class diagrams, if we exclude association inclusions and associations attached to super-classes? One of the restrictions of such a graphical means is their *local character*: one cannot define ontology constraints that involve two or more associations, or associations and attributes simultaneously.

Definition 13. A first order ontology is called **locally constrained** iff each of its non-logical axioms:

- a) either does not contain association predicates; or
- b) contains *exactly one association predicate* and/or its inverse predicate, and zero or more class predicates (but does not contain attribute predicates).

This definition is justified by the following

Proposition 4. Consider a restricted UML ontology O , defined by a UML class diagram D . Obtain the natural first order translation O' of O . Then O' is a locally constrained ontology.

Proof. See Definition 2. Constraints, used to define restricted UML ontologies, are listed as SET_1 and SET_2 of Section 3, hence, they all are local in the above-defined sense. Instances of logical axioms can be non-local, but this cannot cause problems: these axioms are true in all interpretations. So are the inversion axioms. \square

Note. Locally constrained ontologies cannot contain association inclusion axioms – such axioms would be non-local, for example: $\forall x, y(R_i(x, y) \rightarrow R_j(x, y))$.

Intuitively, when defining a locally constrained ontology, one cannot coordinate different associations by setting common conditions on them. And one cannot coordinate associations with attributes as well. As we will see, this implies certain *locality phenomena*.

The conditions (d, e) of the Definition 9 of proper association nets seem to involve coordination between two associations attached to different nodes. One cannot coordinate two associations in a single local axiom. But the coordination may follow from several local axioms that involve *classes*. For example, the associations R, R' form a proper chain, if R connects some class C to some C' , and R' connects C' to some C'' , and these three classes are pairwise disjoint. This situation can be specified by using five local axioms:

$$\begin{aligned} &\forall x, y(R(x, y) \rightarrow C(x) \wedge C'(y)); \forall x, y(R'(x, y) \rightarrow C'(x) \wedge C''(y)); \\ &\forall x \neg(C(x) \wedge C'(x)); \forall x \neg(C(x) \wedge C''(x)); \forall x \neg(C'(x) \wedge C''(x)). \end{aligned}$$

Another example: we cannot formulate a single local axiom asserting that some particular association chain containing two or more associations connects some objects (or does not connect any objects). However, could this not be done indirectly; i.e., by using more than one local axiom? In this case, the answer is negative, and this represents the first of the locality phenomena promised above.

Theorem 2. (First Locality Phenomenon) Assume O is a locally constrained ontology. If the axioms of O imply that Q is a proper association chain that does not connect any objects, then these axioms also imply that at least one of the associations of Q is empty.

This theorem follows easily from the following

Lemma 3. *Assume O is a locally constrained ontology. Consider any model M of O in which Q is a proper association chain, and all associations of Q are non-empty. Then M can be modified by changing the interpretations of associations of Q only, obtaining a model of O in which Q is proper and connects some objects.*

Proof. Consider any model M of O in which the association chain $Q = (R^1, \dots, R^n)$ is proper, and all associations R^i are non-empty. Let us denote by a_i, b_i the objects in M that are connected by R^i ; i.e., for which, $R^i(a_i, b_i)$ is true in M .

Let us modify M obtaining a model of O in which Q is connecting (namely, in which Q connects a_1 to b_n , while remaining proper).

To obtain a sequence of objects connected by the chain Q , we will, as the first step, re-define the interpretation of R^1 in order to connect a_1 to a_2 (instead of b_1); i.e., in order to obtain $R^1(a_1, a_2)$ instead of $R^1(a_1, b_1)$.

If $R^1(a_1, a_2)$ is already true in M , then no action is needed.

But if $R^1(a_1, a_2)$ is false in M , then $a_2 \neq b_1$, and we can try swapping b_1 with a_2 , and apply the Permutation Lemma.

Namely, let us introduce the permutation h of M that only permutes b_1 and a_2 . Since Q is a proper chain, because of the condition (b), $rng(R^1)$ and $dom(R^2)$ are compatible; hence, b_1 and a_2 belong to the same classes. So, h does not affect the membership of objects in class interpretations. And, because of the conditions (a, c), $a_1 \neq b_1$ and $a_1 \neq a_2$, thus $h(a_1) = a_1$.

To apply the Permutation Lemma, let us take $\mathbf{R} = \{R^1\}$, and let us consider a modified interpretation J_1 that differs from the model M only by a permuted interpretation of R^1 : $R^1_{J_1}(x, y)$ is defined as true in J_1 iff $R^1_M(h(x), h(y))$ is true in M .

In fact, since $h(a_1) = a_1$, the only changes are: $R^1_M(a_1, b_1)$ was true, and $R^1_M(a_1, a_2)$ was false, now, $R^1_{J_1}(a_1, b_1)$ is false, and $R^1_{J_1}(a_1, a_2)$ is true. This re-definition of R^1 affects only R^1 and its inverse association. It does not affect neither other associations of Q , nor the ones outside of Q .

By Permutation Lemma, those axioms of O that do not contain R^1 or its inverse, remain true in J_1 , as well as the axioms that contain R^1 and its inverse only, and zero or more class predicates (but do not contain attribute predicates). Since O is locally constrained, it contains no other kinds of axioms. Thus, J_1 is a model of M in which the chain (R^1, R^2) connects a_1 to b_2 .

Now, let us apply the above construction to J_1 and the associations R^2, R^3 , re-defining the interpretation of R^2 , and obtaining a model J_2 of O in which the chain (R^2, R^3) connects a_2 to b_3 , and hence, the chain (R^1, R^2, R^3) connects a_1 to b_3 . Etc., in the end, we will obtain a model of O in which the chain Q connects a_1 to b_n (while remaining proper). \square

Now, let us prove Theorem 2.

Proof. Let us assume the contrary: that the axioms of O do not imply that one of the associations of $Q = (R^1, \dots, R^n)$ is empty. Then, by Model Existence Theorem, there is a model M of O in which Q is proper, and all its associations are non-empty. By Lemma 3 we obtain from M another model M' of O in which Q remains proper, but

is connecting. This contradicts the assumption of Theorem that the axioms of O imply that Q does not connect. \square

8 Singular and Non-singular Association Chains

The singularity phenomenon considered in this Section extends the concept of strong satisfiability introduced for UML class diagrams in Lenzerini and Nobili (1990).

Theoretically, in a first order ontology the situation described in the following definition is not excluded:

Definition 14. A proper association chain Q is called a **singular chain** iff

$$\forall x_1, x_2, y_1, y_2 [Q(x_1, y_1) \wedge Q(x_2, y_2) \rightarrow y_1 = y_2].$$

Namely, a singular chain connects *all* its left-hand-side objects to a *single* right-hand-side object. This may happen, for example, if the range of the last association of Q is a class C that consists of a single object: $C = \{c\}$. Such situations can be easily created in descriptions logics with nominals (Tobies 2000).

An empty association is singular. An association chain that does not connect any objects is also singular.

However, it is hard to imagine such a situation occurring in “practical” ontologies. So, the axioms of a “realistic” ontology should not allow proving that some of its proper association chains are singular. By Model Existence Theorem, this means that “realistic” ontologies should allow for models in which all proper association chains are non-singular, and each class possesses at least two instances, thus extending the concept of strong satisfiability introduced in Lenzerini and Nobili (1990).

For example, by using only the graphical means of *UML class diagrams*, it seems impossible to constrain a non-empty association to be singular, or a class – to possess a single instance only. For consistent ontologies, this is confirmed by the following

Proposition 5. *Assume O is a restricted UML ontology that is consistent in the sense that it allows for a model, in which all classes and associations are non-empty. Then O also allows for a model in which each class possesses at least two instances and all associations are non-singular.*

Proof. Let us build the natural first order translation O' of O , and let us consider any model M of O' , in which all classes and associations are non-empty. Let us create a copy M' of M based on completely different objects and values, and let us consider the union $M \cup M'$, where the operation is applied pairwise to all classes, associations, data types and attributes.

It is easy to verify that, in $M \cup M'$, the axioms of O' remain true. (Consider, one by one, the axioms of SET_1 , SET_2 and inversion axioms.) Thus, $M \cup M'$ is a model of O' in which each class possesses at least two instances and all associations are non-singular (i.e. they connect their left-hand-side objects to at least two different right-hand-side objects). \square

Intuitively, one could expect that in a locally constrained ontology, singularity of an association chain can be caused only by singularity of one of its associations. This is true, as proved in the following

Theorem 3. (Second Locality Phenomenon) *Assume O is a locally constrained ontology. If the axioms of O imply that Q is a singular proper association chain, then these axioms also imply that at least one of the associations of Q is singular.*

This theorem follows easily from the following

Lemma 4. *Assume O is a locally constrained ontology. Consider any model M of O in which Q is a proper association chain, and all associations of Q are non-singular. Then M can be modified by changing the interpretations of associations of Q only, obtaining a model of O in which Q is proper and non-singular.*

Proof. Consider any model M of O in which the association chain $Q = (R^1, \dots, R^n)$ is proper, and all associations R^i are non-singular. Accordingly, let us denote by a_i, b_i, c_i, d_i the objects in M such that

$$R^i(a_i, c_i) \wedge R^i(b_i, d_i) \wedge c_i \neq d_i$$

is true in M . I.e., R^i connects the entire set $\{a_i, b_i\}$ with the entire set $\{c_i, d_i\}$ (two situations are possible: $a_i = b_i$ and $a_i \neq b_i$).

Let us modify M to obtain a new model of O in which the entire chain Q is non-singular (namely, in which Q connects a_1 or the entire set $\{a_1, b_1\}$ with the entire set $\{c_n, d_n\}$).

R^1 already connects the entire set $\{a_1, b_1\}$ with the entire set $\{c_1, d_1\}$ (two situations are possible: $a_1 = b_1$ and $a_1 \neq b_1$).

As the next step, let us re-define the interpretation of R^1 in order to obtain a model in which the chain (R^1, R^2) connects a_1 or the entire set $\{a_1, b_1\}$ with the entire set $\{c_2, d_2\}$.

Namely, let us introduce the permutation h of M that only permutes part of or the entire set $\{c_1, d_1\}$ with part of or the entire set $\{a_2, b_2\}$. Since Q is a proper chain, because of the condition (b), $\text{rng}(R^1)$ and $\text{dom}(R^2)$ are compatible; hence, the permuted objects belong to the same classes. So, h does not affect the membership of objects in class interpretations. And, because of the conditions (a, c), the set $\{a_1, b_1\}$ is disjoint from $\{c_1, d_1\}$ and from $\{a_2, b_2\}$. Thus, $h(a_1) = a_1$; $h(b_1) = b_1$.

Several different situations are possible:

a) If $a_2 = b_2$, then let h to swap c_1 with a_2 only. This will force R^1 to connect a_1 with a_2 . In this case, the chain (R^1, R^2) will connect a_1 to c_2 and d_2 .

b) If $a_2 \neq b_2$, then:

b1) If $\{c_1, d_1\} = \{a_2, b_2\}$ then let h to do nothing. In this case, the chain (R^1, R^2) already connects the entire set $\{a_1, b_1\}$ with the entire set $\{c_2, d_2\}$.

b2) If these sets are disjoint, then (since $c_1 \neq d_1$) let h to swap c_1 with a_2 , and d_1 with b_2 . This will force R^1 to connect a_1 with a_2 and b_1 with b_2 . In this case, the chain (R^1, R^2) will connect the entire set $\{a_1, b_1\}$ with the entire set $\{c_2, d_2\}$.

b3) If these sets have exactly one common element, then let h to swap the remaining ones. This will force R^1 to connect the elements of $\{c_1, d_1\}$ with the elements of

$\{a_2, b_2\}$ (in this, or in the reverse order). In this case, the chain (R^1, R^2) will connect the entire set $\{a_1, b_1\}$ with the entire set $\{c_2, d_2\}$.

To apply the Permutation Lemma, let us take $\mathbf{R} = \{R^1\}$, and let us consider a modified interpretation J_1 that differs from the model M only by a “permuted” interpretation of R^1 : $R_{J_1}^1(x, y)$ is defined as true in J_1 iff $R_M^1(h(x), h(y))$ is true in M .

In fact, since $h(a_1) = a_1$; $h(b_1) = b_1$, only the right-hand-side of $R^1(x, y)$ has been changed, and in the desired way. This re-definition of R^1 affects only R^1 and its inverse association. It does not affect neither other associations of Q , nor the ones outside of Q .

By Permutation Lemma, those axioms of O that do not contain R^1 , remain true in J_1 , as well as the axioms that contain R^1 and its inverse only, and zero or more class predicates (but do not contain attribute predicates). Since O is locally constrained, it contains no other kinds of axioms. Thus, J_1 is a model of O in which the chain (R^1, R^2) connects a_1 or the entire set $\{a_1, b_1\}$ with the entire set $\{c_2, d_2\}$.

Now, let us apply the above construction to the model J_1 and the associations R^2, R^3 , re-defining the interpretation of R^2 , and obtaining a model J_2 of O in which the chain (R^2, R^3) connects a_2 or the entire set $\{a_2, b_2\}$ with the entire set $\{c_3, d_3\}$, and hence, the chain (R^1, R^2, R^3) connects a_1 or the entire set $\{a_1, b_1\}$ with the entire set $\{c_3, d_3\}$.

Etc., in the end, we will obtain a model of O in which the chain Q connects a_1 or the entire set $\{a_1, b_1\}$ with the entire set $\{c_n, d_n\}$. Thus, in this model, the chain Q is non-singular (while remaining proper). \square

Now, let us prove Theorem 3.

Proof. Let us assume the contrary – that the axioms of O do not imply that one of the associations of $Q = (R^1, \dots, R^n)$ is singular. Then, by Model Existence Theorem, there is a model M of O in which Q is proper, but all its associations are non-singular. By Lemma 4 we obtain a model M' of O in which Q remains proper, but is non-singular. This contradicts the assumption of Theorem that the axioms of O imply that Q is singular. \square

To prove the Generalized Granularity Theorem below (Theorem 5), we will also need the following

Lemma 5. *Assume O is a locally constrained ontology. Consider any model M of O in which there is a proper association net $Q_1 + Q_2$ consisting of two non-singular association chains Q_1, Q_2 having a common starting node, but different first association names. Then M can be modified by changing only the interpretation of the first association of Q_1 , obtaining a model of O in which Q_1, Q_2 are proper chains connecting some object b to different objects c, d .*

Proof. Consider any model M of O in which $Q_1 + Q_2$ is a proper association net, Q_1, Q_2 are non-singular association chains having a common starting node, but different first associations. Accordingly, let us denote by $a_i, b_i, c_i, d_i (i = 1, 2)$ the objects in M such that

$$Q_i(a_i, c_i) \wedge Q_i(b_i, d_i) \wedge c_i \neq d_i$$

is true in M .

Since $Q_1 + Q_2$ is a proper net, the sets $\{a_1, b_1, a_2, b_2\}; \{c_1, d_1, c_2, d_2\}$ are disjoint, and the elements of the set $\{a_1, b_1, a_2, b_2\}$ are compatible with respect to classes of O . The sets $\{c_1, d_1\}; \{c_2, d_2\}$ need not be disjoint if the first associations of Q_1, Q_2 have a common end node.

But, of course, $c_1 \neq c_2$ or $c_1 \neq d_2$. It will suffice to consider only the case $c_1 \neq c_2$.

If $a_1 = a_2$, then we already have an object a_1 that is connected to two different objects c_1, c_2 : $Q_1(a_1, c_1) \wedge Q_2(a_1, c_2)$.

If $a_1 \neq a_2$, let us modify M to obtain a new model of O in which Q_1 will connect a_2 to c_1 (Q_2 already connects a_2 to c_2).

Assume R is the first association of Q_1 . Let us swap a_1 with a_2 in the interpretation of R .

Namely, let us introduce the permutation h of M that only permutes a_1 with a_2 . The permuted objects belong to the same classes, so, h does not affect the membership of objects in class interpretations. Secondly, the sets $\{a_1, b_1, a_2, b_2\}; \{c_1, d_1, c_2, d_2\}$ are disjoint; hence, $h(c_1) = c_1; h(d_1) = d_1$.

To apply the Permutation Lemma, let us take $\mathbf{R} = \{R\}$, and let us consider a modified interpretation J_1 that differs from the model M only by a “permuted” interpretation of R : $R_{J_1}(x, y)$ is defined as true in J_1 iff $R_M(h(x), h(y))$ is true in M .

In fact, since $h(c_1) = c_1; h(d_1) = d_1$, only the left-hand-side of $R(x, y)$ has been changed, and in the desired way. This re-definition of R affects only R and its inverse association. It does not affect neither other associations of $Q_1 + Q_2$, nor the ones outside of $Q_1 + Q_2$.

By Permutation Lemma, those axioms of O that do not contain R , remain true in J_1 , as well as the axioms that do contain R and its inverse only, and zero or more class predicates (but do not contain attribute predicates). Since O is locally constrained, it contains no other kinds of axioms. Thus, J_1 is a model of O in which Q_1 connects a_2 to c_1 , and Q_2 connects a_2 to $c_2 \neq c_1$. \square

9 Functional Association Chains

Similar to the multiplicity constraint 0..1 used in UML class diagrams and functional roles used in description logics, we introduce the following

Definition 15. An association chain Q is called **functional** iff it connects each of its left-hand-side (domain) objects to zero or one right-hand-side (range) object:

$$\forall x, y_1, y_2 [Q(x, y_1) \wedge Q(x, y_2) \rightarrow y_1 = y_2].$$

Singular association chains are functional.

Intuitively, one could expect that in a locally constrained ontology a functional association chain consists of functional associations only. This is true, as proved in the following

Theorem 4. (Third Locality Phenomenon) Assume O is a locally constrained ontology, the axioms of which imply that Q is a proper association chain, but do not imply that some of the associations of Q are singular. Then, if the axioms of O imply that Q

is a functional association chain, these axioms also imply that all associations of Q are functional.

Note that the condition “the axioms ... do not imply that some of the associations of Q are singular” is necessary for this theorem to be true. This observation extends the concept of strong satisfiability introduced for UML class diagrams in Lenzerini and Nobili (1990).

Theorem 4 follows easily from the following

Lemma 6. *Assume O is a locally constrained ontology. Consider any model M of O in which Q is a proper association chain, all associations of Q are non-singular, and one of the associations of Q is non-functional. Then M can be modified by changing the interpretations of associations of Q only, obtaining a model of O in which Q is proper and non-functional.*

Proof. Consider any model M of O in which $Q = (R^1, \dots, R^n)$ is a proper association chain, all associations of Q are non-singular, and the association R^i is non-functional. Let us denote by a, b_1, b_2 the objects for which

$$R^i(a, b_1) \wedge R^i(a, b_2) \wedge b_1 \neq b_2.$$

is true in M .

Step 1. First, let us obtain a model M' in which the initial segment (R^1, \dots, R^i) connects some object c with both of objects b_1, b_2 . If $i = 1$, then no action is needed: $c = a$.

If $i > 1$, then consider the initial segment (R^1, \dots, R^{i-1}) . All its associations are non-singular; hence, non-empty as well. Then, by Lemma 3, we can modify the model M obtaining a model M' in which this chain connects some object c with some object d . Of course, $d \in \text{rng}(R^{i-1})$ and $a \in \text{dom}(R^i)$ are compatible. If $d = a$, then no action is needed.

If $d \neq a$, then let us apply Permutation Lemma to swap d and a in the interpretation of the left side of R^i , obtaining a model M'' in which the following formula is true:

$$R^i(d, b_1) \wedge R^i(d, b_2) \wedge b_1 \neq b_2.$$

In M'' , (R^1, \dots, R^i) connects c with both of objects b_1, b_2 .

Step 2. Now, consider the tail segment $Q_T = (R^{i+1}, \dots, R^n)$ in order to obtain a model in which the entire chain $Q = (R^1, \dots, R^n)$ connects c with two different objects e_1, e_2 . If $i = n$, then no action is needed: $e_1 = b_1; e_2 = b_2$.

If $i < n$, then, all associations of Q_T are non-singular; hence, by Lemma 4, we can modify M'' obtaining a model M''' in which Q_T is non-singular. Denote by d_1, d_2, e_1, e_2 the objects for which

$$Q_T(d_1, e_1) \wedge Q_T(d_2, e_2) \wedge e_1 \neq e_2$$

is true in M''' .

The objects $d_1, d_2 \in \text{dom}(R^{i+1})$ and $b_1, b_2 \in \text{rng}(R^i)$ are compatible.

Let us apply Permutation Lemma to modify the interpretation of the right side of R^i , obtaining a model M'''' in which the chain Q connects c with e_1, e_2 . Several different situations are possible:

a) If $d_1 = d_2$, then let us swap b_1 with d_1 only. This will force $R^i + Q_T$ to connect b_1 with e_1 and e_2 , and the entire chain Q – to connect c with e_1 and e_2 .

b) If $d_1 \neq d_2$, then:

b1) If $\{d_1, d_2\} = \{b_1, b_2\}$ then let us do nothing. In this case, c is already connected to e_1 and e_2 .

b2) If these sets are disjoint, then let us swap d_1 with b_1 , and d_2 with b_2 .

b3) If these sets have exactly one common element, then let us to swap the remaining ones.

In the model M'''' , the entire chain Q connects the object c with two different objects e_1, e_2 ; i.e., Q is non-functional (while remaining proper). \square

Now, let us prove Theorem 4.

Proof. The axioms of O imply that Q is a functional association chain; hence, Q is functional in all models of O . But these axioms do not imply that some of the associations of the chain Q are singular.

Assume the contrary to the assertion of Theorem: the axioms of O do not imply that in $Q = (R^1, \dots, R^n)$ a particular association R^i is functional. Then, by Model Existence Theorem, there is a model M of O in which all associations of Q are non-singular, and the association R^i is non-functional. By Lemma 6, M can be modified by changing the interpretations of associations of Q only, obtaining a model of O in which Q is proper and non-functional. This contradicts the functionality of Q in all models of O . \square

10 Granular Ontologies

This kind of ontology was introduced in Barzdins et al. (2014) as a specific organization of databases allowing for extremely fast processing of *ad hoc* queries. See Section 1 for a detailed explanation.

However, in the Generalized Granularity Theorem (Theorem 5 below) a more general definition of granularity will be used that does not refer explicitly to classes of the ontology.

Definition 16. Assume that the axioms of the ontology O imply that D is a proper association diagram of O , and A is a node in it.

In a model of O , consider the union $(\cup \text{rng}(R)) \cup (\cup \text{dom}(R'))$ over all ingoing associations R and all outgoing associations R' of the node A . This union is called the **master class** of O , and its elements – master objects of O .

Let us say that O is a **granular ontology** with A in D as its *master node* iff the axioms of O imply that:

a) each non-master object is connected to a *single master object* via some proper association chain that belongs to D and ends in A ; and

b) no two objects connected to different master objects are connected to each other via proper association chains that belong to D and do not cross A .

For the ultimate justification of this definition see Proposition 6 below.

In a model of O , a *granule* is the set of all objects connected to a particular master object (the master object itself included).

Remark 11. Recall that “some chain of D ” also means chains obtained by reverting arrows and associations contained explicitly in D .

It is well known that, in general, in a first order language the transitive closure of a predicate cannot be expressed as a single formula (see, for example, Areces 2000, Theorem 1.2, and Baader 1991). Fortunately, this does not preclude expression of the granularity property of some ontology O in the language of O itself.

Indeed, as noted above, for each particular D , the assertion “ D is a proper association diagram of O ” can be expressed as a closed formula in the language of O . Hence, for example, a formula expressing

“ x is connected to y via some chain belonging to D ”

can be obtained as a disjunction of the following formulas:

$C^Q =$ “the chain Q connects x to y ”,

where Q runs over all chains belonging to D .

In this way, the granularity property of O with A in D as the master node can be expressed by the following formulas in the language of O :

$F_A(x, a) =$ “there is a chain Q that belongs to D and ends in A , such that $C^Q(x, a)$ ”;

$G_A(x, y) =$ “there is a chain Q that belongs to D and does not cross A , such that $C^Q(x, y)$ ”;

$F =$ “for all x , there is exactly one a such that $F_A(x, a)$ ”;

$G =$ “for all $x, y, a, b, F_A(x, a) \wedge F_A(y, b) \wedge a \neq b \rightarrow \neg G_A(x, y)$ ”;

$Gran(O, D, A) = F \wedge G$.

So, by saying that “the axioms of O imply that O is a granular ontology with A in D as the master node”, we mean the provability of the formula $Gran(O, D, A)$.

As the ultimate justification of Definition 16 let us prove the following

Proposition 6. *Consider a restricted UML ontology O , defined by a UML class diagram D . Obtain the natural first order translation O' of O . According to Proposition 2, the axioms of O' imply that D is a proper association diagram of O' . The following is equivalent:*

a) O is a granular ontology in the sense of Section 1 with A being its master class.

b) O' is a granular ontology in the sense of the above Definition 16 with the node of class A in D being its master node.

Proof. Once more, according to Proposition 2, the axioms of O' imply that D is a proper association diagram of O' . The rest of the proof is straightforward. \square

Motivating example: how complicated could be “topology” of a granular ontology?

Let us consider an ontology O_3 represented as a *triangle* ABC consisting of three pairwise disjoint classes A, B , and C , and three associations R_{BA}, R_{CA} , and R_{BC} connecting the respective classes. The associations R_{BA} and R_{CA} are assumed to be mandatory and functional in the direction to A . Thus, the axioms of O_3 include, among others:

$$\forall x \neg (A(x) \wedge B(x)); \forall x, y (R_{BA}(x, y) \rightarrow B(x) \wedge A(y));$$

$$\forall x(B(x) \rightarrow \exists y R_{BA}(x, y)); \forall a_1, a_2, b[R_{BA}(b, a_1) \wedge R_{BA}(b, a_2) \rightarrow a_1 = a_2].$$

These are all local axioms. By including a *non-local* axiom

$$\forall a_1, a_2, b, c[R_{BA}(b, a_1) \wedge R_{CA}(c, a_2) \wedge R_{BC}(b, c) \rightarrow a_1 = a_2]$$

we could ensure that any two objects b, c connected by R_{BC} are connected to the same object of the class A . Thus, in this way, we could ensure that O_3 is a granular ontology with A as its master class. In any model of O_3 , each granule would consist of all objects of classes B, C connected to a particular object of the master class A .

Obviously, in this way, by introducing non-local axioms, one could obtain granular ontologies of arbitrary “topological” complexity (in fact, any directed graph could be used as the starting point).

However, could the granularity property be ensured by using *local axioms* only? Sometimes, it could, indeed. For example, *tree ontologies* (see Definition 1 above) are easily verified as granular. And, obviously, any tree ontology can be represented alternatively, as a set of local axioms of the following types:

disjointness: $\forall x \neg(A(x) \wedge B(x))$;

typing: $\forall x, y(R_{BA}(x, y) \rightarrow B(x) \wedge A(y))$;

mandatory participation: $\forall x(B(x) \rightarrow \exists y R_{BA}(x, y))$;

functionality: $\forall x, y_1, y_2[R(x, y_1) \wedge R(x, y_2) \rightarrow y_1 = y_2]$.

However, *are the tree ontologies the only possible kind of locally constrained granular ontologies?* The answer is: yes, in some generalized sense, they are the only kind, as will be proved in Section 11 below.

For the proof, we will need the following lemma (establishing, in fact, another locality phenomenon).

Lemma 7. (“No triangles”) Assume O is a locally constrained ontology the axioms of which imply that:

a) D is a proper association diagram of O , and

b) O is a granular ontology with A in D being its master node.

Consider in D a “triangle” consisting of three different nodes A (the master node), B , and C and three association chains R^{BA} , R^{CA} , and R^{BC} belonging to D , starting and ending in the respective nodes. Assume, R^{BA} does not cross C , R^{CA} does not cross B , and R^{BC} does not cross A .

Then the axioms of O imply that one of the chains R^{BA} , R^{CA} is singular, or that the chain R^{BC} does not connect any objects.

Proof. Assume the contrary: that the axioms of O do not imply that one of the chains R^{BA} , R^{CA} is singular, or that the chain R^{BC} does not connect. Then, by Model Existence Theorem, there is a model M of O in which the chains R^{BA} , R^{CA} are non-singular, and the chain R^{BC} is connecting. Accordingly, let us denote by

$$a_1, a_2, \alpha_1, \alpha_2, b_1, b_2, c_1, c_2$$

the objects in M such that the formulas

$$R^{BA}(b_1, a_1) \wedge R^{BA}(b_2, a_2) \wedge a_1 \neq a_2;$$

$$R^{CA}(c_1, \alpha_1) \wedge R^{CA}(c_2, \alpha_2) \wedge \alpha_1 \neq \alpha_2$$

are true in M . And let us denote by β, γ the objects for which $R^{BC}(\beta, \gamma)$ is true in M .

Of course, the objects $a_1, a_2, \alpha_1, \alpha_2$ all are master objects.

If $b_1 = b_2$, then b_2 is connected to two different master objects – a_1 and a_2 . This contradicts the granularity of O . Thus, $b_1 \neq b_2$.

Similarly, we can conclude that $c_1 \neq c_2$.

Since R^{BC} is a proper chain, by Proposition 1, $\beta \neq \gamma$.

To apply Permutation Lemma, let us swap β with b_1 or b_2 , and γ with c_1 or c_2 in such a way that β and γ become connected to different master objects. Such a permutation h exists, because $b_1 \neq b_2$ and $c_1 \neq c_2$.

Let us denote by $R_1^{BA}, R_L^{BA}, R_1^{CA}, R_L^{CA}, R_1^{BC}, R_L^{BC}$ the first and last associations of R^{BA}, R^{CA}, R^{BC} respectively. Then:

$$a_1, a_2 \in \text{rng}(R_L^{BA}); \alpha_1, \alpha_2 \in \text{rng}(R_L^{CA});$$

$$b_1, b_2 \in \text{dom}(R_1^{BA}); c_1, c_2 \in \text{dom}(R_1^{CA}); \beta \in \text{dom}(R_1^{BC}); \gamma \in \text{rng}(R_L^{BC}).$$

Since D is a proper association net, and R^{BA} does not cross C , and R^{CA} does not cross B , and R^{BC} does not cross A , we can conclude that:

a) The sets $\text{rng}(R_L^{BA}), \text{rng}(R_L^{CA})$ are disjoint from the last four of the above sets. Hence, $a_1, a_2, \alpha_1, \alpha_2$ are not affected by the permutation h .

b) The sets $\text{rng}(R_1^{BA}), \text{rng}(R_1^{CA})$ are disjoint from the last four of the above sets. Hence, the permutation h affects only the left sides of R_1^{BA}, R_1^{CA} .

In this way, by changing the interpretations of R_1^{BA}, R_1^{CA} , we obtain (by Permutation Lemma) a model of O in which β, γ are connected to two different master objects. Since β, γ are connected to each other by the chain R^{BC} which does not cross A , this contradicts the granularity of O . \square

11 Generalized Granularity Theorem

Now we are ready to establish our main locality phenomenon:

Theorem 5. (Generalized Granularity Theorem). *Assume O is a locally constrained ontology the axioms of which imply that:*

- a) D is a proper association diagram of O ; and
- b) O is a granular ontology with A in D being its master node.

Assume also that these axioms do not imply that some of the associations of O are singular.

Then the association diagram D represents a tree with A as the root node, in which no more than one association connects two nodes, and the axioms of O imply that all associations of D are functional in the direction to A .

Again, note that the condition “the axioms ... do not imply that some of the associations of O are singular” is necessary for this theorem to be true.

Proof. First, let us note that, as an association diagram of an ontology, D does not contain isolated nodes and self-loop arrows (see Definition 12).

According to Model Existence Theorem, there is a model M of O in which all associations of O are non-singular and hence, non-empty.

Hence, D cannot contain arrows isolated from the master node A . Indeed, in M , the associations attached to such arrows, would be non-empty; hence, the objects connected by these associations would not be connected to the master node by association chains of D . This would contradict the granularity of O .

Let us prove, that D cannot contain triangles (see Lemma 7 for the definition of a triangle). Assume the contrary – that D contains a triangle A', B, C . If none of the three nodes is the master node A , let us add to this picture a chain Q connecting A' to A . And let us replace by A the nearest to A node of the triangle that is located on Q . We have obtained a triangle A, X, Y that contains the master node. Now we can apply Lemma 7: the axioms of O imply that X or Y is connected to A by a singular association chain, or X and Y is connected by a chain that does not connect any objects. But the first clause is impossible because of Theorem 3 (the axioms of O do not imply the existence of singular associations). And the second clause is impossible because of Theorem 2 (the axioms of O do not imply the existence of empty associations).

Now, consider two association chains of D , attached to A and located on the same chain of edges. By Lemma 5, both association chains are equal, or one of them is a segment of the other. Indeed, if, to some edge, these chains would attach different association names, then, by Lemma 5, from the model M we could obtain a model M' in which these chains would connect some object to different master objects. This would contradict the granularity of O . Thus, in D , no more than one association connects two nodes.

And finally, if the axioms of O would not imply that all association chains of D attached to A are functional in the direction to A , then there would be a model in which some object is connected, via some association chain, to two different master objects. This would contradict the granularity of O . Hence, by Theorem 4, all associations of the chains attached to A ; i.e., all associations of O are functional in the direction to A . \square

The Definition 16 of generalized granular ontologies includes one more requirement: each non-master object must be connected to some master object. In this way, the ontology can guarantee that granules cover the entire database. In first order ontologies, any association R is “mandatory” in the sense that any object in $dom(R)$ is connected to some object of $rng(R)$. This is why there are no elegant ways to introduce the latter requirement in the Generalized Granularity Theorem 5.

As an application of the Generalized Granularity Theorem, one can obtain, in particular, the following generalization of Theorem 1 (Granularity Theorem from Barzdins et al. (2014)):

Theorem 6. (Granularity Theorem 2) *Consider any restricted UML ontology O that is consistent in the sense that it allows for a model, in which all classes and associations are non-empty. Then, O is granular iff it is a tree ontology.*

Restricted UML ontologies and tree ontologies are defined in the above Definitions 2, 1 respectively.

Proof. Let us prove that if O is a consistent restricted UML ontology in which no more than one association connects two classes, and which is granular, then it is a tree ontology. The converse is trivial.

So, consider O as defined by some UML class diagram D . Obtain the natural first order translation O' of O .

According to Proposition 4, O' is a locally constrained ontology.

According to Proposition 2, the class diagram of O represents a proper association diagram of O' .

According to Proposition 6, since O is a granular ontology in the sense of Section 1 with A being its master class, then O' is a granular ontology in the sense of the above Definition 16 with A in D being its master node.

According to Proposition 5, since O allows for a model, in which all classes and associations are non-empty, then O also allows for a model in which all associations are non-singular.

Hence, we can apply Theorem 5 to conclude that the diagram D represents a tree with A as the root node, in which no more than one association connects two classes, and all associations of D are functional in the direction to A .

And finally, since in O no more than one association connects two classes, for each object, there is only one chain of associations in the diagram D connecting it to some master object. Hence, all associations of O must be mandatory in the direction to A .

This completes the proof that O is a tree ontology. \square

12 Finite Model Setting

The finite model setting considered in this Section, was explored for UML class diagrams in Balaban and Maraee (2013).

One might object that allowing for infinite models (i.e., infinite databases, see Section 4) makes the results of the paper less interesting.

Fortunately, the model transformations used in the above proofs are based on Permutation Lemma; hence, they do not add new objects, and finite models are transformed into finite ones. This allows obtaining of the (somewhat less elegant) *finite model versions* of the above lemmas and theorems.

To formulate these versions, we must replace:

a) phrases like “the axioms of the ontology O imply a closed formula F ” with “a closed formula F is true in all finite models of the ontology O ” (the former is stronger than the latter as a premise, but weaker – as a consequence);

b) phrases like “the axioms of the ontology O do not imply a closed formula F ” with “a closed formula F is false in some finite model of the ontology O ” (the former is weaker than the latter as a premise, but stronger – as a consequence);

Consider, for example, Theorem 2 about connecting chains in locally constrained ontologies. Its proof is based on Lemma 3 that explicitly declares that it does not add new objects; thus, it transforms finite models into finite ones.

By using the same Lemma 3, the following finite model re-formulation of Theorem 2 can be proved.

Theorem 7. (First Locality Phenomenon) Assume O is a locally constrained ontology. If, in all finite models of O , Q is a proper association chain, and there is a finite model of O in which all associations of Q are non-empty, then there is a finite model of O in which the chain Q connects some objects.

Similarly, Lemma 4 remains valid for finite models. This allows to prove the finite model re-formulation of Theorem 3:

Theorem 8. (Second Locality Phenomenon) Assume O is a locally constrained ontology. If, in all finite models of O , Q is a proper association chain, and there is a finite model of O in which all associations of Q are non-singular, then there is a finite model of O in which Q is a non-singular chain (and thus, connects some objects as well).

Similarly, Lemma 6 remains valid for finite models. This allows to prove the finite model re-formulation Theorem 4.

Theorem 9. (Third Locality Phenomenon) Assume O is a locally constrained ontology, and in all finite models of O , Q is a proper and functional association chain. Assume there is a finite model of O in which all associations of Q are non-singular. Then, in this model, all associations of Q are functional.

Lemma 7 and the Generalized Granularity Theorem (Theorem 5) allow for finite model formulations as well.

Theorem 10. (Generalized Granularity Theorem). Assume O is a locally constrained ontology, and in all finite models of O :

- a) D is a proper association diagram of O ; and
- b) O is a granular ontology with A in D being its master node.

Assume, additionally, that there is a finite model M of O in which all associations are non-singular.

Then the association diagram D represents a tree with A as the root node. And, in all finite models of O , all associations are functional in the direction to A .

13 Conclusions

The present paper considers ontologies formalized as sets of axioms in first order languages (first order ontologies).

The main results of the paper are Theorems 2, 3, 4, 5 establishing certain *locality phenomena* that appear when all the constraints used in the ontology are *local* in the sense that (roughly) each constraint involves at most one association (more precisely – see Definition 13 of locally constrained ontologies.).

Establishing and proving of Theorem 5 (Generalized Granularity Theorem) and Theorem 6 (Granularity Theorem 2), thus, solving an open problem formulated in Barzdins et al. (2014), was the main goal of the paper. Theorems 2, 3, 4 were used in the proof of Theorem 5.

Computational complexity issues of processing queries to databases described by locally constrained schemas will be considered in a separate paper.

The principal proof method of the paper (Permutation Lemma) does not apply to ontologies containing association inclusions (super- and sub-associations) or associations connected to super-classes. However, it seems, that if association inclusions are the only kind of non-local axioms of the ontology, then the locality phenomena established above (the Granularity Theorem included) should appear in such ontologies as well. Could the method of the Permutation Lemma be refined to prove this? Or, is a completely new method needed?

Acknowledgements

This work has been supported by the European Regional Development Fund (ERDF) project Nr. 2014/0020/2DP/2.1.1.1.0/14/ APIA/VIAA/072 at the Institute of Mathematics and Computer Science, University of Latvia. The author is grateful to Karlis Cerans, the head of the project, for helpful comments and suggestions.

References

- Areces, C.E. (2000). *Logic Engineering. The Case of Description and Hybrid Logics*, PhD Thesis, University of Amsterdam.
- Baader, F. (1991). Augmenting Concept Languages by Transitive Closure of Roles: An Alternative to Terminological Cycles, In: *Proc. 12th int. joint conf. on Artif. intell. (IJCAI-91, Sydney, Australia)*, 1, 446-451.
- Balaban, M., Maraee, A. (2013). Finite satisfiability of UML class diagrams with constrained class hierarchy, *ACM Trans. Softw. Eng. Methodol.*, 22(3), 24.
- Barzdins, J., Rencis, E., Sostaks, A. (2014). Data ontologies and ad hoc queries: a case study, In: *Databases and Information Systems: Proc. 11th Int. Baltic Conf.*, (Baltic DBIS 2014, Tallinn, Estonia), Tallinn Univ. Technol. Press, 55-66.
- Berardi, D., Calvanese, D., De Giacomo, G. (2005). Reasoning on UML class diagrams, *Artif. Intell.*, 168(1-2), 70-118.
- Ding, Z., Peng, Y. (2004). A probabilistic extension to ontology language OWL, In: *Proc. 37th Hawaii Int. Conf. Syst. Sci. (HICSS-37, Big Island, Hawaii, 2004)*.
- Lenzerini, M., Nobili, P. (1990). On the satisfiability of dependency constraints in entity-relationship schemata, *Inf. Syst.*, 15(4), 453-461.
- Kaneiwa, K., Satoh, K. (2010). On the complexities of consistency checking for restricted UML class diagrams, *Theor. Comput. Sci.*, 411(2), 301-323.
- Klein, M. (2004). *Change Management for Distributed Ontologies*, SIKS Dissertation Series, 2004-11, Vrije Universiteit, Amsterdam.
- Mendelson, E. (2009). *Introduction to Mathematical Logic*, 5th ed., Chapman and Hall.
- Ozdamar, L., Ulusoy, G. (1994). A local constraint based analysis approach to project scheduling under general resource constraints, *Eur. J. Oper. Res.*, 79(2), 287-298.
- Tobies, S. (2000). The Complexity of Reasoning with Cardinality Restrictions and Nominals in Expressive Description Logics, *J. Artif. Intell. Res. (JAIR)*, 12, 199-217.
- UML (2016). The Unified Modeling Language, *Object Management Group*, <http://www.omg.org>

Received August 29, 2016 , accepted September 19, 2016