

The ICOM 3.0 Intelligent Conceptual Modelling tool and methodology

Editor(s): Axel Polleres, DERI Galway, Ireland

Solicited review(s): Andreas Harth, AIFB, University of Karlsruhe, Germany; Stijn Heymans, SemanticBits LLC, Herndon, VA, U.S.A.; anonymous reviewer

Pablo R. Fillottrani^a, Enrico Franconi^b
Sergio Tessaris^b

^a *Universidad Nacional del Sur, and Comisión de Investigaciones Científicas de la provincia de Buenos Aires, Argentina*

prf@cs.uns.edu.ar

^b *Free University of Bozen-Bolzano, Italy*

{franconi,tessarisi}@inf.unibz.it

Abstract. ICOM (version 3.0) is an advanced conceptual modelling tool, which allows the user to design multiple extended ontologies. Each project can be organised into several ontologies, with the possibility to include inter- and intra-ontology constraints. Complete logical reasoning is employed by the tool to verify the specification, infer implicit facts, devise stricter constraints, and manifest any inconsistency. ICOM is fully integrated with a very powerful description logic reasoning server which acts as a background inference engine. The intention behind ICOM is to provide a simple conceptual modelling tool that demonstrates the use of, and stimulates interest in, the novel and powerful *knowledge representation* based technologies for *database* and *ontology* design.

Keywords: DL, OWL, ER, UML, ORM

1. Introduction

ICOM is an advanced conceptual modelling tool, which allows the user to design multiple ER or UML class diagrams with inter- and intra-model constraints. Complete logical reasoning is employed by the tool to verify the specification, infer implicit facts, devise stricter constraints, and manifest any inconsistency.

For the ontology creation and maintenance tasks, ICOM interface supports ontology engineers in engineering ontologies that meets clear and measurable quality criteria. Indeed, recently we observe the development of large numbers of ontologies which have, however, usually been developed in an ad hoc manner by domain experts, often with only a limited understanding of the semantics of ontology languages. The result is that many ontologies are of low quality - they make poor use of the languages in which they are written and do not accurately capture the author's rich knowledge of the domain. This problem becomes even more acute as ontologies are maintained and extended over time, often by multiple authors. Poor quality ontologies usually require localised "tuning" in order to achieve the desired results within applications. This leads to further degradation in their overall quality, increases the brittleness of the applications that use them, and makes interoperability and reuse difficult or impossible. To overcome these problems tools are needed which support the design and the development of the basic infrastructure for building, merging, and maintaining ontologies.

The leverage of automated reasoning to support the domain modelling is enabled by a precise semantic definition of all the elements of the class diagrams. The diagrams and inter-model constraints are internally translated into a class-based logic formalism. The same underlying logic enables the use of a view definition language to specify additional constraints, not captured at the diagram level. The conceptual modelling language supported by ICOM can express:

- the standard Extended Entity-Relationship data model or the standard UML class diagrams (we are working on providing *Object-Role modelling* ORM [5] as well), enriched with disjoint and

covering constraints and definitions attached to classes and relations by means of view expressions over other classes and relationships in the ontology;

- inter-ontology mappings, as inclusion and equivalence statements between view expressions involving classes and relationships possibly belonging to different ontologies.

The tool allows for the creation, the editing, the managing, and the storing of several interconnected ontologies, with a user friendly graphical interface. The main purpose of the ICOM project is not to provide to the ontology community a robust tool potentially replacing the many other tools available; we do not claim that ICOM is currently more usable than any of the existing conceptual modelling tools for ontology design (such as, for example, [8,1]). ICOM is meant to be a proof of concept, willing to showcase two main points:

1. the effectiveness of using a class diagram graphical syntax for expressing ontologies, even with complex languages;
2. the emphasis to the use of complex automated reasoning tasks to deduce implied facts, as opposed to mere subsumption (classification) and consistency.

The two above points are novel and in our opinion very important in the context of the existing ontology design tools and methodologies (see next Section). Indeed ICOM proves (point 1) the feasibility and the ease of use of a class diagram graphical syntax for expressing ontologies, even with complex ontology languages, by relying on the notion of views (which roughly correspond to OCL constructs) in order to capture the (typically very few) cases where a larger expressivity than graphical class diagrams is needed. ICOM is based on a *deduction-complete* notion of reasoning support relative to the class diagram graphical syntax (point 2). Users will see the original ontology *graphically completed* with all the deductions making sense given the provided ontology, and expressed in the graphical class diagram language itself. This includes checking class and relationship consistency, discovering implied class and relationship interrelations (e.g., subsumption) or cardinality constraints, and in general discovering any implied but originally implicit class diagram graphical construct. Customarily, ontology design tools just provide a support limited to class subsumption and consistency.

ICOM provides a general framework to support the typical tasks involved in such activities:

- Authoring of concept descriptions: in this task the author wants to add a new concept description to the ontology or modify a concept description that was already contained in the ontology. This may happen either in the design phase of the ontology or during the maintenance phase. After producing a candidate description of the concept, the author needs to understand the implicit consequences of his modelling and the interaction of this description with the other descriptions in the ontology.
- Structuring of the ontology: in this task the ontology designer wants to improve the structure of an ontology by inserting intermediate concepts into the ontology diagram. He needs support to decide where to add such concepts and how to describe them.

ICOM is a fairly mature project, its first release has been published in 2000 (see [7,4]). The version 3.0 of the ICOM tool is loosely based on the ICOM tool previously released in 2000 as an Entity-Relationship editor (which had around 3,000 registered installations, mostly in academic environments and for teaching purposes in industry), and a demo of a preliminary version was presented few years ago [3]. The foundations of the user-computer interaction have been radically changed according to the experience of the first ICOM and the research in this last decade. The system has been completely re-implemented, using different graphic libraries. The graphical interface has been completely rewritten to improve the usability and intuitiveness of the tool. Interoperability with other tools is a crucial aspect; so, import and export modules have been developed for XMI 2.x and Description Logics based ontology languages via DIG.

The ICOM tool is written in standard Java 5.0, and it is distributed on Linux, Mac, and Windows machines. ICOM communicates via the DIG 1.1 protocol with a description logic server, such as, for example, RACER. ICOM provides an interface for importing and exporting ontologies in UML-XMI class diagrams format.

ICOM is available as a free download at <http://www.inf.unibz.it/~franconi/icom/>

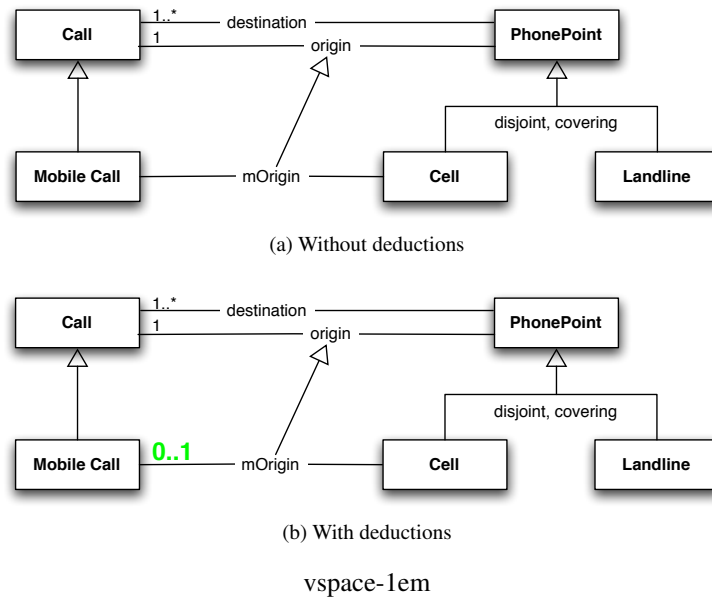


Fig. 1. First ontology design scenario.

2. Related Works

There are several ontology building tools available, but Protégé¹ is the most popular one. Protégé is a free, open source Java-based platform from Stanford University. At its core, Protégé implements a rich set of knowledge-modelling structures and actions that support the creation, visualisation, and manipulation of ontologies in various representation formats. Protégé can be customised to provide domain-friendly support for creating knowledge models and entering data, and also it can be extended by a plugin architecture for building specific tools and applications. Protégé ontologies can be exported into a variety of formats including RDF(S), OWL, and XML Schema. Protégé popularity is based on the robustness of the platform, the broad functionality provided by the plugin architecture, and the ample support from the community of users. It allows two main ways of modelling ontologies, the Protégé-Frames editor, based on Open Knowledge Base Connectivity protocol (OKBC), and the Protégé-OWL editor. The Protégé OWL editor includes a visual tool, and provides means for showing the user inferences obtained from an external OWL reasoner, being in this sense similar to our tool. But the concept and methodology that ICOM promotes, combining graphical design with reasoning support, is

missing in the Protégé-OWL editor. Editing the ontology is done mainly from the tree-based hierarchical representation of concepts, properties and nominals, each in a different panel. Inferences are shown in an extra tab panel, and thus don't have a graphical representation. Also, this inference panel is available only for individual concepts, not for properties. There is no possibility of integrating different ontologies like in ICOM's project/model mechanism. Finally, reasoning usage is limited to finding super- and sub-classes and detecting inconsistencies. On the other hand, the Protégé-OWL editor endorses a more expressive language, since it includes nominals and transitive and symmetric properties.

The NeOn toolkit² is an outcome from an European Commission Sixth Framework Programme project, with the goal of providing methodological and tool support for developing and managing a new generation of semantic applications. The toolkit is an open source, Eclipse-based multi-platform ontology engineering environment, with comprehensive support for the whole ontology engineering life-cycle. It provides an extensive set of plug-ins covering a variety of ontology engineering activities, but these plugins are quite isolated. For example, there are several Human-Ontology interaction plugins with different techniques for visualising ontologies and there is a Reasoner plu-

¹<http://protege.stanford.edu>

²<http://neon-toolkit.org>

gin to provide access to external reasoner. But inferences are shown in a text based fashion and cannot be visualised in the interaction plugins.

The TopBraid Composer³ is a commercial modelling environment for developing ontologies in several formats, implemented as an Eclipse plugin. It has comprehensive support for developing, managing and testing configurations of knowledge models, incorporating an extensible framework for developing semantic client/server or browser-based applications that can integrate disparate applications and data sources. Like in ICOM, inferences are shown graphically in the composer, but these are limited to IsAs hierarchies. No reasoning on associations or roles is done.

Kaon2⁴ is an infrastructure for managing OWL-DL, SWRL and F-Logic ontologies, jointly developed by the Research Center for Information Technologies (FZI), the University of Karlsruhe and the University of Manchester. Its main feature is its own inference engine, with little support in editing ontologies. On the other hand, Ontotrack⁵ is a ontology authoring tool for a subset of OWL Lite that combines a graphical layout with mouse enabled editing features, but with no reasoning support.

Other related tools are: DOME⁶ DERI Ontology Management Environment; SWOOP⁷ from the University of Maryland; IODT⁸ IBM Integrated Ontology Development Toolkit; OilEd⁹ from the University of Manchester; Knoodl¹⁰ from Revelytix, Inc.; Ontolingua¹¹; and Chimera¹² web based services from Stanford University. They all represent ontologies as a text tree-based hierarchy of concepts, with limited or no graphical visualisation. Reasoning usage is very restrictive, mostly only for concepts, and thus inferences have little to do in helping the user designing an ontology. Most of them are no longer under active development.

³http://www.topquadrant.com/products/TB_Composer.html

⁴<http://kaon2.semanticweb.org/>

⁵<http://www.informatik.uni-ulm.de/ki/ontotrack/>

⁶<http://dome.sourceforge.net/>

⁷<http://www.mindswap.org/2004/SWOOP/>

⁸<http://www.alphaworks.ibm.com/tech/semanticstk>

⁹<http://oiled.man.ac.uk/>

¹⁰<http://www.knoodl.com>

¹¹<http://www.ksl.stanford.edu/software/ontolingua/>

¹²<http://www.ksl.stanford.edu/software/chimaera/>

3. Example User Scenarios

In this section we will show few basic common use cases, emphasising in particular the added value of the functionalities of ICOM. Deductions automatically performed by the ontology editor may help the ontology engineer in validating the design. In fact, if the derived constraints make little sense to the engineer they may help in suggesting changes, or they may show serious but subtle conceptual mistakes. ICOM adopts as a neutral language to represent ontologies UML-like class (or ER) diagrams. The advantage of this choice is that examples may be clear even to people who do not have familiarity with more classical ontology languages, due to its graphical nature and to the fact that UML class diagrams are quite widespread within computer technologies. Note that we use the notation of cardinality constraints as *participation constraints* as it is customary in ER diagrams, but as opposed to the UML standard that uses them as *look-through constraints*.

3.1. Ontology Design

Complete reasoning over expressive ontology languages supports the ontology engineer in creating and maintaining ontologies tasks. The following examples show the kind of insights that a reasoning enabled system could provide during the ontology editing.

Let us consider the ontology represented by the diagram in Figure 1a. The ontology states that mobile calls are a kind of calls (IsA link between entities); that phone points are partitioned between cell points and landline points (i.e., any phone point is either a cell or a landline point, but not both: they form a covering and disjoint IsA hierarchy). Each call has at least one destination phone point (mandatory participation of `cell` to `destination`), while it has exactly one origin phone point. Mobile calls are related to cells through the `mOrigin` relationship. Finally, the binary relationship `mOrigin` is included in the binary relationship `origin`.

Which are the consequences of the above ontology? The system could automatically complete the diagram in the way depicted in Figure 1b. The added constraint states that in the above scenario it is necessarily true that each mobile call may have an origin from at most one cell point. In order to understand why this is true, consider the following. The `mOrigin` binary relationship is included in the `origin` binary relationship, i.e., any pair in `mOrigin` is also among the pairs

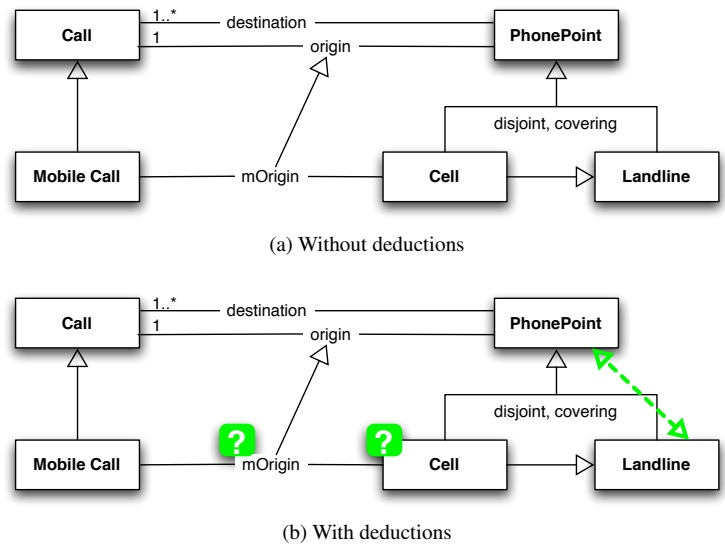


Fig. 2. Second ontology design scenario.

in origin. Since each call participates exactly once as first argument to the `origin` relationship, if I take a generic sub class of calls, such as the class of mobile calls, and a sub relationship of the `origin` relationship, such as `mOrigin`, then we can conclude that necessarily each mobile call participates at most once as first argument to the `mOrigin` relationship. Nothing can be concluded about the minimum participation, since the `mOrigin` relationship may not contain all calls in the `origin` relationship.

Let us now consider the same diagram as Figure 1 with an additional IsA link stating that each cell point is among the landline points (see Figure 2a). The system is now able to conclude and present the user new facts about the elements in the diagram (see Figure 2b). Firstly, the `cell` entity is inconsistent, i.e., does not have any instance, since the disjointness constraint in the IsA link states that there is no element in common between cell and landline phone points. The empty set denoted by the `cell` entity is the only set which can be at the same time disjoint and a subset of another set. Since the phone point entity is formed by the union of the cell and landline entities, and the cell entity is inconsistent, the landline entity becomes equivalent to the phone point entity. Moreover, since there is no cell point, there is no pair in the `mOrigin` relationship as well (i.e., it is inconsistent): the diagram states that any second argument of the `mOrigin` relationship should be of the cell type. The `Mobile Call` entity is not inconsistent, since it may be populated by calls which

have no `mOrigin` at all (this is possible, since there is no mandatory participation constraint).

Let us now add a cardinality constraint, stating that now each mobile call should participate at least once to the `mOrigin` relationship (i.e., a mandatory participation constraint). The change results in the diagram of Figure 3a. Now the system deduces that the mobile call entity is inconsistent as well (see Figure 3b).

3.2. Ontology Integration and Views

In this section we introduce a scenario of usage of the tool in the context of ontology integration by making use of the view facility of ICOM.

Figure 4a shows two ontologies in the phase to be integrated by the ontology engineer. The top ontology describes concepts where information about Italian ISO certified companies is held; in particular, the information about their contact person is specified. The facts described by the diagram state that a company should have at least one employee, and that it should be involved in at least one sector. Among the employees there is the contact person of the company, which should be unique. Moreover, the Italian companies are exactly defined as those companies which are in a country called Italy, while the ISO certified Italian companies are exactly those Italian companies having an ISO certification (specified as a boolean property called `isoCert`). Please note the particular use of the 'slash' "/" operator in front of the completely intensionally defined classes—in the ontol-

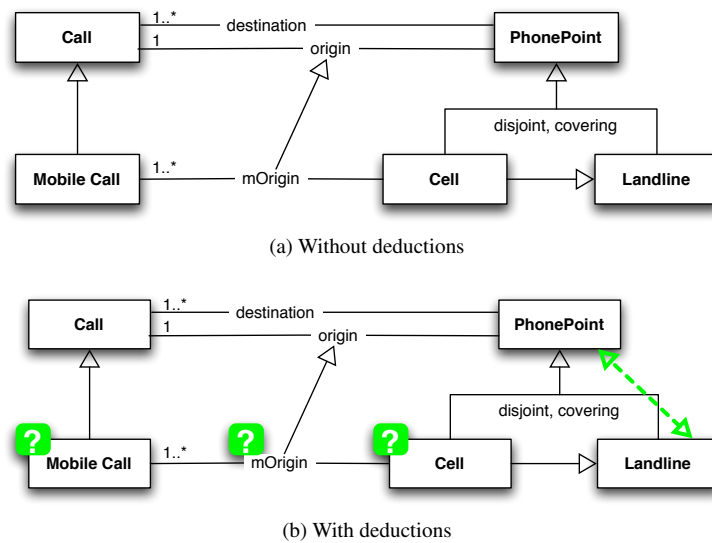


Fig. 3. Third ontology design scenario.

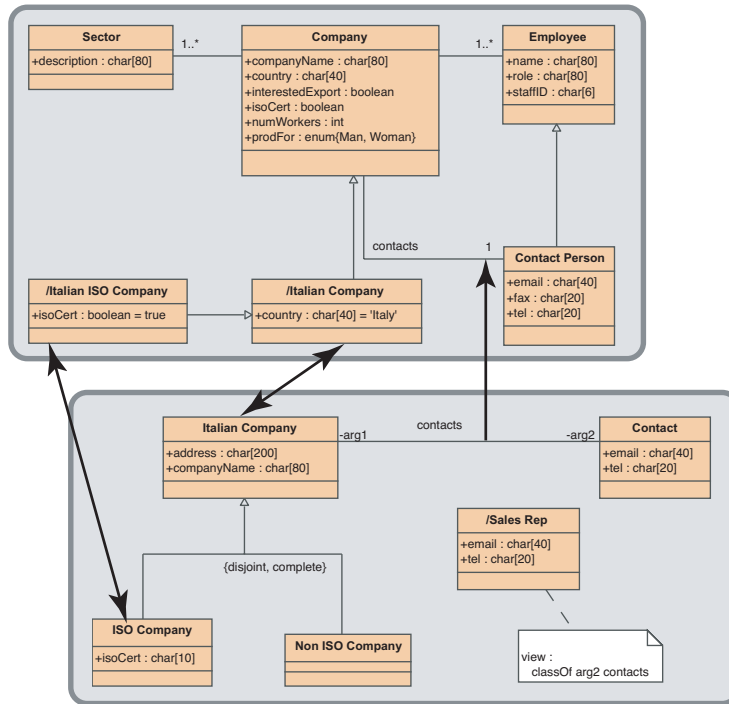
ogy the classes *Italian Company* and *Italian ISO Company* are completely defined by means of the properties specified in the diagram. This is the simplest case of a view defined in the ontology.

The lower ontology of Figure 4a describes a slightly different perspective about the same domain. Still, there are Italian companies and their contact persons (but now without any cardinality constraint, and without mentioning that contact persons of companies should be employees), plus the specification that the companies are either ISO certified or not—here the ISO companies are identified by the code of their ISO certification institution. In addition, the ontology includes the view class *Sales Rep*, which is completely defined by means of its attributes together with the view expression stating that sales representative is the range of the *contacts* association. Note that the view definition can be written in any *reasonable* linear ontology language, such as an OCL constraint, or an OWL axiom, or a SQL check constraint, or first order logic sentence. The view definition mechanism is the hook that allows to use the full power of the ontology language—if the user wants. Most of the ontologies will not need to use this hook, and they will be more directly understandable by the engineers. In the case when subtle integration constraints have to be written, views will come at hand, by providing an expressive language to the engineers in a way which is perfectly integrated with the diagrammatical paradigm proposed here.

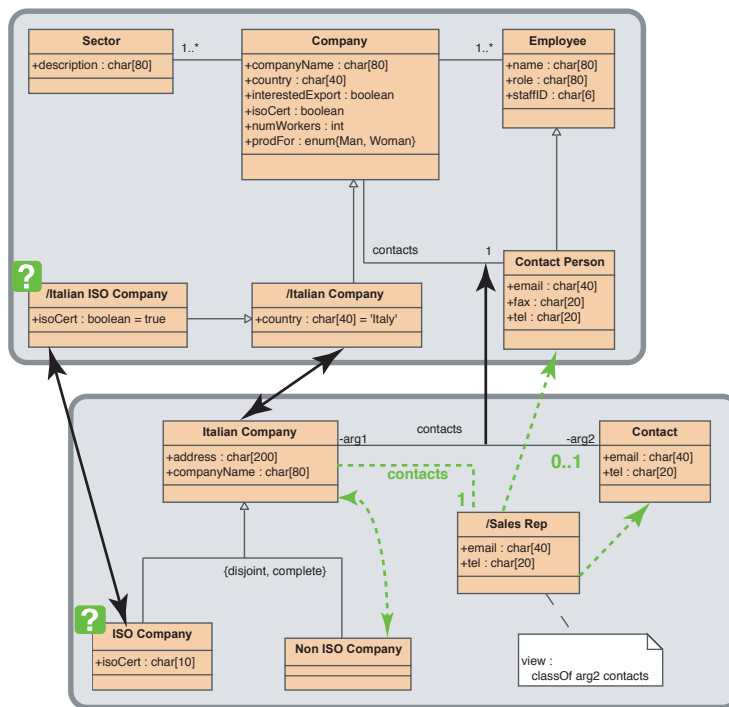
Figure 4a includes also the mappings between the two ontologies. You can see that the Italian ISO companies in the top ontology are declared to be the same as the (Italian) ISO companies in the lower ontology, and that the Italian companies in both ontologies are declared to be equivalent as well. Moreover, the *contacts* association in the lower ontology is declared to specialise the homologous association in the top ontology. Inter-ontology mappings are declared by simply drawing directed links between pairs of classes (or pairs of associations) belonging to different ontologies; these can state either equivalence, or containment, or disjointness.

Now, the whole picture seems very reasonable to any ontology engineer; however there are interesting, unexpected, and clarifying consequences that our design tool will automatically draw—still in a diagrammatic fashion. These are shown in Figure 4b.

The first consequence relates to the equivalence stated between the two Italian ISO certified company class definitions in the two ontologies. The two classes have a type incompatibility in the attribute *isoCert*: one is declared to be a boolean value, while the other is declared to be a string of ten characters. Indeed, the system deduces that if such an integration has to be taken seriously, then the two classes have to be empty in any possible context, since an object in one context which, say, represents an Italian ISO Company by having an attribute *isoCert* with the value *true*, can not be at the same time an instance of a class whose *isoCert* is declared to be of an incompatible type

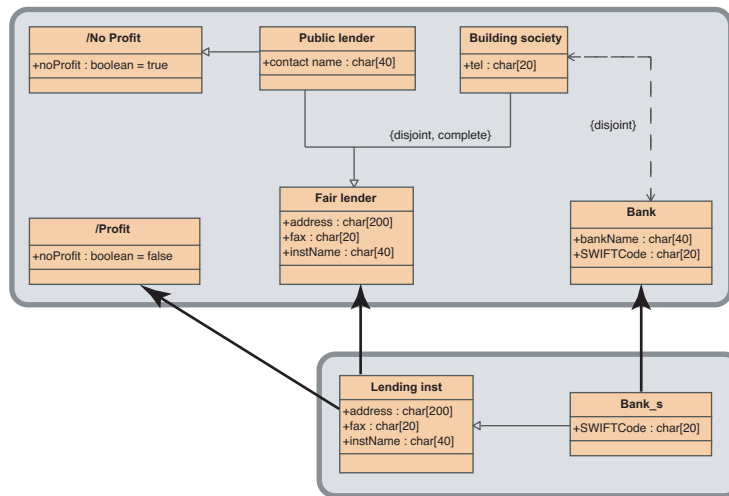


(a) Without deductions

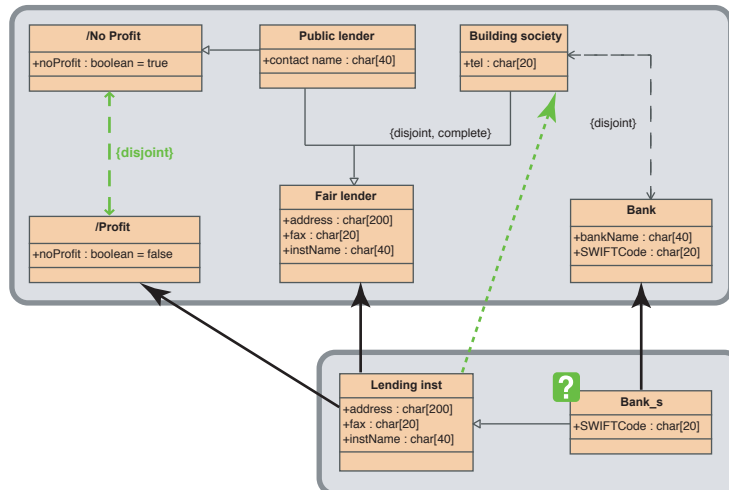


(b) With deductions

Fig. 4. The first integration scenario.



(a) Without deductions



(b) With deductions

Fig. 5. The second integration scenario.

(i.e., string). Therefore, such an instance can not exist, and, as a matter of fact, no instances of the two classes can not exist at all. This first deduction by the tool (indicated by the question marks on the corner of the two classes) is actually a hint to the designer to actually take care of this *data reconciliation* problem, by, for example, providing local conversion functions between the two attribute types. Please note that the tool also correctly derives the fact that any object which is instance of the Italian Company class (in any of the two contexts/ontologies) should also be an instance of a non ISO company. In fact, since no Italian ISO companies can exist in the current version of the scenario, any Italian company will necessarily be a non ISO certified

company. This is made explicit by the dashed equivalence link added by the tool between the Italian Company class and the Non ISO Company class in the lower ontology.

If we go on with the analysis of the deductions made by the tool, we see that a stricter cardinality constraint has been deduced: now any Italian company can have at most one contact person (in the sense of the lower ontology)—this is the $[0..1]$ cardinality constraint found at the right end of the `contacts` association. The system has deduced this stricter constraint by observing that Italian companies are companies, which have exactly one contact person (in the sense of the top ontology); moreover, each Italian com-

pany should have contact persons (in the lower ontology sense) among the contact persons in the top ontology sense. Therefore, no Italian company can have more than one contact person (in the lower ontology sense). The lower bound is not derived since the specialisation of the `contacts` association may not necessarily consider all the Italian companies. So, this is an example of a deduction which is not just an IS-A link or an inconsistent class, which are the only kind of deductions that the most advanced ontology design tools (like, e.g., OIEd, or Protege) are capable of.

Another deduction which can not be done by any other ontology design tool is the one which makes explicit a `contacts` association in the lower ontology between Italian companies and sales representative, plus the now stricter cardinality constraint stating that each Italian company has exactly one sales representative. Please note how powerful this deduction mechanism is: an isolated class is automatically fully put in context, by considering all the possible constraints which may relate it with the other terms of the integrated ontologies. As a matter of fact it can be proved that the design tools derives *all* the correct constraints (and only those).

Finally, we note that the tool derives also that sales representative are both contact persons in the top ontology sense and contact persons in the lower ontology sense.

All these deductions may help the ontology engineer in validating the design—if the derived constraints make sense to the engineer; they may help in suggesting changes; or they may show serious but subtle conceptual mistakes. The next case scenario shown in Figure 5a is an example of the latter case.

In this new integration scenario, the top ontology describes fair lenders which are partitioned into public lenders and building societies. Public lenders are no profit companies, and in addition it is stated that banks are not building societies.

In the lower less detailed ontology, we have the generic class of lending institutions which specialises into the bank class. We also assume that actually the lower ontology, in spite of the fact that it uses more generic terms, describes a world which is actually a portion of the world described more accurately by the top ontology.

A very natural integration between the two ontologies is pursued by the ontology engineer: she/he states that banks of the lower ontology are among the banks of the top ontology, and that lending institutions of the

lower ontology are fair lenders and profit companies as defined in the top ontology.

The consequences of this integration attempt are immediately drawn by the tool as depicted in Figure 5b. As a first (more or less obvious) deduction we can observe that the profit and the no profit classes are derived to be disjoint, as expected. However, it turns out—from the big question mark at the corner of the `Bank_s` class—that no banks can exist according to the lower ontology! This is somehow unexpected, since we thought we were playing a rather simple game in this case. Why is this? A quick glance at the attribute types shows that they are perfectly compatible this time. The reason is the following. First of all, we can derive that lending institutions are building society (as pointed out by the tool); in fact, lending institutions are fair lenders, which can be either public lenders or building societies. On the other hand we have that lending institution are profit companies, which are provably disjoint from public lenders. therefore, any lending institution has necessarily to be a building society. At this point, we are closer to the answer to our original question about the inconsistency of the `Bank_s` class. Those lower ontology banks are at the same time a kind of top ontology banks and building societies (by transitivity). However the two latter classes are disjoint, hence no common instance can exist—i.e., no bank can be a lending institution according to this integration system.

Of course, there should be something wrong the way the two ontologies have been integrated, and this calls for a revision of the mappings. The engineer should either omit the mapping stating that lending institutions are necessarily profit companies, or the mapping stating that lending institutions are necessarily fair lenders. In both cases, the outcome will be acceptable by the engineer, and she/he should choose the one that best fits further analysis of the domain that she/he may have done after this unexpected discovery.

4. The Ontology Editor

The Ontology Editor works on *projects*, which may contain one or more UML class diagrams. The diagrams are referred as *models*. Multiple projects can be opened at the same time, but objects cannot be moved across them. Only one project is visible at a time and the editing of each project is independent. The user can switch between different projects using the tabs at the top of the project area.

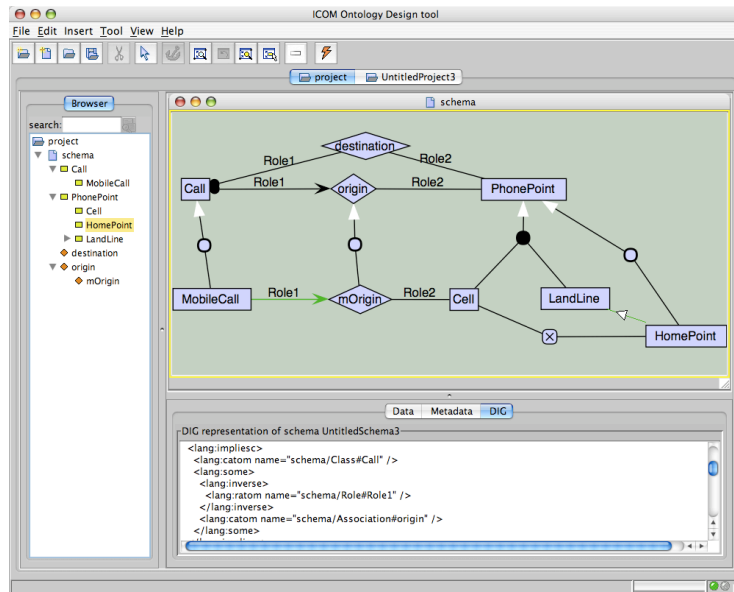


Fig. 6. The ontology design scenario in ICOM.

Figure 6 shows the main window of the Ontology Editor editing a single model. Classes are represented by boxes and n-ary associations by diamonds. Associations may have so-called association classes specifying their attributes. IsA relationships are represented as arrows with a disc in the middle (e.g. see *MobileCall* and *Call*).

The tool does not implement special visual techniques for handling very large ontologies. The tasks that it supports, i.e. authoring of concept description and structuring the ontology, are not aimed at working simultaneously with thousands of concepts. However, a set functionalities that are very useful in managing such ontologies are available. First, the interface is zoomable, that is, the level of detail and size of the icons that represent the model can be smoothly changed by pressing the right mouse button and dragging left to zoom out or right to zoom in. Also, the window can be panned by pressing the middle button and dragging. This allows the user to focus the attention in a specific region of the ontology. There are also two dedicated buttons for zooming: one will show the complete graph, and the other will zoom in to show the selected elements. Selection works by left-clicking on icons or by left click and drag; also, there is a button for expanding the selection to all connected nodes, which is very useful in combination with the zoom-to-selection button. Finally, custom automatic layout algorithms for ontologies are under development. These

combine known layout algorithms for drawing large graphs, with special conventions used in ontologies, like IsAs hierarchies are drawn top-down and associations are drawn in the middle of related concepts. New metrics to measure the "quality" of ontology graphs were developed with this purpose.

4.1. Editing Models

Most of the model editing is done in the project panel, where each model in the project is displayed in a separate model panel. In addition, two dialogues are used to elicit additional information about model objects. The attribute domain dialogue allows the domain of attributes to be set. The definition dialogue enables the characterisation of a class or association by means of a view written in the language described in the next Section.

Objects can be created by selecting the appropriate button in the toolbar, or an entry under the Diagram menu, or a contextual menu in the project area. Most object-creating operations require further inputs to complete the operation. Usually, the user is requested to select an existing object in the diagram (e.g. during the creation of an IsA relationship). In this case, the system will highlight only objects in the diagram suitable for the specific operation.

All the objects of the diagram have a name. Upon their creation the system allocates a new fresh name,

which can be edited by the user. To improve the identification of the nodes, when icons become smaller because of the zoom level, all the nodes show their name on a tool-tip when the mouse is hovering over them. Names are scoped by the model they belong; e.g. classes with the same name in different models are considered different.

Metadata fields can be associated to every kind of objects. These fields are ignored in the reasoning process.

The creation of a new class adds a new box in the diagram with a new default name. Every class can optionally have attributes. Attributes are added and edited by means of a specific attribute dialogue. Similar to classes, attributes of the same name in different models are considered different. Attributes of the same name within the same model represent the same attribute. For each attribute, a domain should be indicated. There the set of possible domains is not predefined, and the user is allowed to enter an arbitrary name. Unlike the classes and associations, domains have a global context. Therefore, domains of the same name in different models are considered to be the same.

Associations are created by default with no roles. N-ary associations can be specified by adding new roles to existing ones. The creation of a new association introduces a corresponding association class; which can be edited as a normal class (e.g. it can have attributes).

Adding new roles to an existing association requires the user to select the association and a class which restricts the domain of the argument of the association corresponding to the role. Similar to class and associations names, role names have a model scope.

For example, assume there are two models $M1$ and $M2$, each one with a binary association `lives` having the roles `subject` and `object`. Note that, being association scoped over models, from the global perspective there are two associations $M1:lives$ and $M2:lives$. Now, the modelling of the domain requires that $M2:lives$ is more specific (i.e. a subset) of $M1:lives$. Since also role names are scoped over each model, overall there are four different roles. Therefore, the more specific association ($M2:lives$) inherits the roles of the general one, ending up being of arity four (namely the roles $M2:object$, $M2:subject$ and $M1:object$, $M1:subject$).

Roles denote the connection of a class to an association and it is also used to express the cardinality constraints of a class in an association. A role may have two constraints: *totality*, or the minimum cardinality, and *uniqueness* representing the maximum cardinality.

In the current version of the system, the numbers expressing cardinality are restricted to be 0 and 1. A minimum cardinality of 1 indicates that all instances of a class must participate in the association at least once (i.e. mandatory constraint). A maximum cardinality of 1 indicates that all instances of a class can only participate once in the association (i.e. functional constraint).

Within a project, equivalence and subset role mappings can be defined between roles in the same or different models. These allow a better characterisation of the relationship between associations across different models. In the former example, $M2:lives$ can be set as a binary association by saying that $M1:object$ contains $M2:object$, and that $M1:subject$ contains $M2:subject$.

The system enables the user to specify inheritance relationships among classes and associations. The relationships can be arbitrary (e.g. cycles are allowed) provided that classes can only inherit from classes, and associations from associations. Formally, the inheritance is expressed by the inclusion (subclass) constraint.

On the diagram, inheritance is specified by means of IsA links (in the diagram indicated by arrows with a circle in the middle) connecting nodes. IsA links can be specified one-to-one, or many-to-one. The latter groups together more than one (association) class and restrict all of them to be a subclass of the link target.

The possibility of grouping more than one descendant, not only provides way of visually organising the layout of the model; but enable the user to specify additional constraints among the (association) classes. In particular, the *covering* and *disjointness* constraints. The first one expresses the fact that the (association) class is equivalent to the union of the specified descendant; while the second constraints the grouped (association) classes to be mutually disjoint.

Note that disjointness among classes is not assumed by default; so, in absence of a specific constraint, (association) classes may overlap.

4.2. Inter-Model Axioms

Additional constraints among classes and associations can be expressed by means of intra- as well as inter-model axioms. The Ontology Editor provides four types of axioms: Node Definition, Equivalence, Subsumption and Disjointness. As discussed in Section 1 these constraints provide a powerful modelling tool in the context of data integration and ontology mapping.

```

⟨class⟩ ::= ⟨class-name⟩
          | not ⟨class⟩
          | and ⟨class⟩+
          | or ⟨class⟩+
          | project ⟨role-name⟩ ⟨association⟩
          | atleastr ⟨number⟩ ⟨role-name⟩ ⟨association⟩
          | atmostr ⟨number⟩ ⟨role-name⟩ ⟨association⟩

⟨association⟩ ::= ⟨association-name⟩
                | not ⟨association⟩
                | and ⟨association⟩+
                | or ⟨association⟩+
                | select ⟨role-name⟩ ⟨class⟩

```

Fig. 7. View definition language

Each class and association can be fully defined by means of a view expression. The view expression language is more expressive than the diagrammatic definition language, so enables the expert user to add constraints that cannot be expressed by the UML diagram alone.

The adopted view language (see Figure 7) based on the DLR description logic. A definition has a global context, meaning it can express inter-model relationships as well as intra-model relationships. The view language includes two syntactic sorts: one for classes and one for associations. Full boolean operators are allowed, plus a *selection* operator (selecting tuples in an association with a specific class type in some named role argument) and a unary *projection* operator (projecting an association over a named role argument). A generalised projection operator with cardinality restrictions is available as well.

Since a definition can refer to objects in different models, a name-prefix is used in definitions to distinguish objects with the same name but from different models. The name-prefix used is the model's name followed by a colon symbol. For example, `class1` in `Model1` and `class1` in `Model2` would be referred to as `Model1:class1` and `Model2:class1` respectively.

Any two (associations) classes in any model can be related by semantic relationships stating their equivalence, subsumption, or disjointness. Creating one of these relationships requires the user to specify source and target node. The system prevents the creation of a relationship between non-homogeneous nodes by restricting the scope of the second node to be selected.

4.3. Exporting and Importing Projects

ICOM projects can be saved and retrieved in an own XML format, preserving the meaning of all elements including view definitions. It is also possible to import UML class diagrams saved in the XMI format. The tool only recognises the subset of XMI determined by classes, associations, attributes, roles and primitive datatypes defined within a UML model. Functional and mandatory constraints on roles are the only type of imported constraints. Aggregation relationships in the UML model are ignored. We are currently working on exporting projects in XMI files, but this translation would be necessarily carried out with some loss of meaning because, for example, not all view definitions can be expressed in XMI even with attaching OCL expressions to the model elements.

5. Automated Reasoning

Although the Ontology Editor can be used as a standalone modelling tool, exploiting its full capabilities requires the coupling of the system with a Description Logic reasoner. Without such an automated reasoning tool the Ontology Editor would be unable to perform *deduction-complete* automated reasoning over the models. As we noted, this includes checking class and relationship consistency, discovering implied class and relationship inter-relations (e.g., subsumption) or cardinality constraints, and in general discovering any implied but originally implicit class diagram graphical construct.

Instead of implementing its own dedicated reasoner, the Ontology Tool can exploit any DIG enabled DL reasoner (see [2]). Being DIG a standardised communication protocol, the user can choose the most suitable DL reasoner (e.g. the one used by other in-house project), or upgrade to the latest version of the preferred reasoner without being forced to upgrade to a different version of the Ontology Editor.

We adopted the DIG protocol rather than the OWL API [6] because it is language independent rather than based on a Java model. Currently we are in the process of evaluating the migration to OWLLink [9], which is the successor of DIG protocol for the latest OWL 2 standard.

To enable the automated deduction mechanism, the user must specify the URL of the server where the DIG reasoner is available. Note that it is not required that the DIG server should be installed in the same com-

puter as the Ontology Editor. Once the URL to the DIG server has been provided, the system checks the connection to the reasoner to verify its availability. Upon success, the menu item and toolbar button enabling the automated deduction will be enabled.

The so called *verification process* can be computationally expensive, so it is activated only on user's request. This process includes the following operations. The selected project is encoded into an appropriate Description Logics knowledge base and shipped to the DIG reasoner. Each class, association in the project is checked for satisfiability (i.e. non-emptiness). For each class, association in the project, its equivalent peers, and super-classes are determined. For each class-role-association triple, the system calculates the stricter minimum and maximum cardinality constraints. To perform these operations, the system formulates a sequence of queries to be sent to the DIG reasoner. Accordingly to the received answers the Ontology Editor infers properties of the models in the project. To perform these operations, the system formulates a sequence of queries to be sent to the DIG reasoner, which is linear in the number of project elements. Accordingly to the received answers the Ontology Editor infers properties of the models in the project. The algorithm for this inference is quadratic in the number of concepts and roles, and linear in the number of axioms and IsA links. Thus, the tool can reasonable manage projects with several hundreds of elements, calling a current state-of-the-art reasoner.

After the verification process, the system provides the user with a visual account of the deductions by modifying the appearance of the model diagrams in the project. All unsatisfiable objects will appear in red in the model diagrams. An object is unsatisfiable when necessarily describes an empty set of tuples of objects. Additional non explicit deductions will appear in green, to be distinguished from the user specified elements of the diagrams. Semantically equivalent objects are connected with newly inserted equivalent axiom links. Objects discovered to hold an inclusion relationships between them are connected with subsumption axiom links. Cardinality constraints of roles for that are more restrictive than those originally specified. Although the deductions are displayed on the actual diagrams, it is up to the user to decide whether they should be permanently added to the models or discarded. The rational behind this behaviour is that the automated reasoning process may detect unwanted deductions caused by a wrong modelling of the domain. In this case the user should correct the project before

any subsequent editing. Another reason is that, in spite of the fact that only the non-trivial deductions are presented, the user is satisfied by the fact that they are implicit without the need of having them explicitly asserted.

The user can discard the deductions and the entire project will be returned to its original state (and any information about unsatisfiability will be discarded). Editing one of the models in the project will also discard the deductions before the editing is carried out. Alternatively, the equivalence, subsumption association, and role cardinality deductions can be added permanently to the project by committing them.

6. Conclusions and Future Works

In this paper we presented ICOM, an advanced conceptual modelling tool grounded on more than ten years of research on the use of automated reasoning to support the development and integration of ontologies. ICOM employs a diagrammatic based language to represent most of the constructs used in ontology design; although it enables the use of non graphical ontology languages, experience with users demonstrates that the design of the diagrammatic language is sufficiently expressive to describe rich domains. Moreover, deductions are expressed within the same diagrammatic language; providing a uniform view over design and analysis of models.

By means of use cases we demonstrated the importance of exploiting basic reasoning tasks (as subsumption) in order to provide richer information on ontologies. This is a crucial step towards guaranteeing the quality of the ontologies designed using a tool like ICOM.

The research and development of ICOM continues on two main tracks: from one side we are improving the modelling workflow by considering alternative modelling languages and reasoning services; while on the other hand we are enhancing the user experience by improving the graphical user interface and the interoperability.

We are currently considering the adoption of modelling features from ORM [5] conceptual modelling methodology and representation. Its adoption would have the advantage of leveraging the vast research which has been carried on supporting the user in the modelling tasks; including the integration of natural language generation. The use of ORM modelling style would require also a redesign of the reasoning tasks in

order to align the inferences to the new graphical representation.

On the interface we are improving the automatic layout algorithms and working on the support of undo actions. We also plan to include a role browser tab to show the role hierarchy in the same style of the class browser. Moreover, we are improving the interoperability with other tools by tackling the import and export compatibility with XMI and OWL.

References

- [1] Sean Bechhofer, Ian Horrocks, Carole A. Goble, and Robert Stevens. OilEd: a reason-able ontology editor for the semantic web. In Franz Baader, Gerhard Brewka, and Thomas Eiter, editors, *Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence*, volume 2174 of *Lecture Notes in Computer Science*, pages 396–408. Springer, 2001.
- [2] Sean Bechhofer, Ralf Möller, and Peter Crowther. The DIG description logic interface. In Diego Calvanese, Giuseppe De Giacomo, and Enrico Franconi, editors, *Proceedings of the 2003 International Workshop on Description Logics (DL2003)*, volume 81 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2003.
- [3] Pablo R. Fillottrani, Enrico Franconi, and Sergio Tessaris. The new ICOM ontology editor. In Bijan Parsia, Ulrike Sattler, and David Toman, editors, *Proceedings of the 2006 International Workshop on Description Logics (DL2006)*, volume 189 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2006.
- [4] Enrico Franconi and Gary Ng. The i.com tool for intelligent conceptual modeling. In Mokrane Bouzeghoub, Matthias Klusch, Werner Nutt, and Ulrike Sattler, editors, *Proceedings of the 7th International Workshop on Knowledge Representation meets Databases (KRDB 2000)*, volume 29 of *CEUR Workshop Proceedings*, pages 45–53. CEUR-WS.org, 2000.
- [5] T.A. Halpin, A.J. Morgan, and T. Morgan. *Information modeling and relational databases*. Morgan Kaufmann series in data management systems. Elsevier/Morgan Kaufman Publishers, 2008.
- [6] Matthew Horridge, Sean Bechhofer, and Olaf Noppens. Igniting the OWL 1.1 touch paper: The OWL API. In Christine Golbreich, Aditya Kalyanpur, and Bijan Parsia, editors, *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions*, volume 258 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2007.
- [7] Matthias Jarke, Christoph Quix, Diego Calvanese, Maurizio Lenzerini, Enrico Franconi, Spyros Ligoudistianos, Panos Vassiliadis, and Yannis Vassiliou. Concept based design of data warehouses: The DWQ demonstrators. In Weidong Chen, Jeffrey F. Naughton, and Philip A. Bernstein, editors, *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, page 591. ACM, 2000.
- [8] Holger Knublauch, Ray W. Ferguson, Natalya Fridman Noy, and Mark A. Musen. The protégé OWL plugin: An open development environment for semantic web applications. In Sheila A. McIlraith, Dimitris Plexousakis, and Frank van Harmelen, editors, *Proceedings of ISWC 2004: Third International Semantic Web Conference*, volume 3298 of *Lecture Notes in Computer Science*, pages 229–243. Springer, 2004.
- [9] Thorsten Liebig, Marko Luther, Olaf Noppens, Mariano Rodriguez, Diego Calvanese, Michael Wessel, Matthew Horridge, Sean Bechhofer, Dmitry Tsarkov, and Evren Sirin. OWLlink: DIG for OWL 2. In Catherine Dolbear, Alan Ruttenberg, and Ulrike Sattler, editors, *Proceedings of the OWLED 2008 Workshop on OWL: Experiences and Directions*, volume 432 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.