IMPROVED DEVELOPMENT OF THE DOMAIN ONTOLOGY
FOR DIFFERENT USER PROFILES

Application Domain: Conformity Checking in Construction

Anastasiya Yurchyshyna, Alain Zarli
CSTB, 290 Route des Lucioles, BP 209, 06904 Sophia Antipolis, France

Catherine Faron-Zucker, Nhan Le Thanh
I3S, UNSA-CNRS, 930 route des Colles, BP 145, 06903, Sophia Antipolis, France

Keywords: Development of the domain ontology, Conformity-checking modelling, User profiles, Semantic search, Validation by usage.

Abstract: This paper presents a formal method for the development of the domain ontology for different user profiles in the context of conformity checking in construction, which is developed to enrich our conformity-checking model. We start by describing our research domain: the conformity-checking in construction. Then we discuss some ontology-based approaches for formalising domain knowledge and particularly focus on methods for the ontology development for different user profiles. In order to efficiently adapt our conformity-checking ontology for different user profiles, we suggest a semantic approach for the improved development of the domain ontology that takes into account the domain knowledge. This semantic approach is based on three main ideas. First, we adapt the knowledge acquisition method developed for our conformity-checking model. Second, we integrate a method of the context modelling of the domain ontology applied by end users by integrating the results of the semantic search. Third, we develop an approach for the adaptation of the domain ontology for different user profiles. Finally, we describe a web-based prototype, the C3R (Conformity Checking in Construction: Reasoning) prototype that integrates our semantic method of the improved development of the domain ontology for different user profiles.

1 INTRODUCTION

This paper presents a formal method for the development of the domain ontology for different user profiles. This work continues and extends our research on the modelling of the conformity checking process in construction, and particularly focuses on the conceptual modelling of domain knowledge and the usage-based validation by end users.

The complexity of the conformity checking problem can be explained by the following factors: (i) the multidisciplinary of the components defining the conformity checking (e.g. modeling of construction regulations, reasoning on conformity), (ii) the interdependence of various actors of the construction domain; (iii) the large amount of the non formalised expert knowledge guiding the process, (iv) the great volumes of construction data to be retrieved and maintained.

The central problem of the conformity checking in construction is to automate the process of checking whether a construction project (e.g. a private house, a public building, a non-building installation) is conform to a set of conformity requirements described by regulation texts. The semantic complexity of this problem requires an expressive formalism for representing the knowledge of the checking process. Recently, multiple approaches for the development of building-oriented ontologies have been developed: e-COGNOS (El-Diraby et al. 2003), ifcOWL (Gehre and Katranuschkov, 2007), buildingSMART (Bell and Bjorkhaug, 2006). Despite the variety of these approaches, these generic ontologies can be hardly used for a specific aim of conformity-checking.
To address these limitations, on our previous research (Yurchyshyna et al., 2008a), we developed a conformity-checking ontology, which integrates not only building-related knowledge, but also the knowledge on conformity regulation texts and the expert knowledge on checking procedures. Developed with the help of domain experts, mostly from CSTB (Centre Scientifique et Technique du Bâtiment), our conformity-checking ontology was a key component of our conformity-checking model.

Complex and multidisciplinary, the construction industry is a field of collaborative work and communication of multiple actors, the so-called “key players of the building-oriented market”, who form the target audience for the development of the construction sector and their needs define the innovation process of the industry. The main key players of the building-oriented market are as follows:

- **architects** generating data related to different aspects of a building;
- **engineers** responsible for generating data related to a specified facility’s system of a building;
- **contractors** dealing with process-related characteristics of a building (scheduling, cost analysis, project management, etc.);
- **consumers** of a building product;
- **building product manufacturers** generating supplementary data related to a building product (e.g. physical and functional characteristics, cost);
- **legal authorities** formulating performance-oriented rules of the development of a building.

Obviously, different actors of the conformity checking in construction have different needs and understanding of the checking process. It also means that they may interpret and use the knowledge from the domain ontology in a different way (e.g. an acoustic engineer may need a very detailed “version” of the domain sub ontology concerning acoustics, in contrast to a final user interested in general non-detailed conformity recommendations).

The formalisation and the integration of the domain tacit knowledge play an important role in the checking modelling. In our previous work (Yurchyshyna et al., 2008b) we describe our approach for formalising expert knowledge, which is also base on the conformity-checking ontology.

From the other hand, it is always interesting to enrich the initial ontology by the knowledge that is acquired from its usage. This research objective motivated our research on the validation of the conformity-checking ontology by integrating the results of the semantic search.

The interdependence of the actors of the application domain (i.e. conformity checking in construction) is, however, an important factor to take into account by the application of a generic domain ontology. It this context, it is a real challenge to enrich the approach for the ontology development by the knowledge of its usage by different user groups and to refine the ontology for different user profiles. This research objective motivates our work on the development of our semantic method of the improved development of the domain ontology for different user profiles, the DOUP method.

The paper is organized as follows. In next section, we discuss some ontology-based approaches for formalising domain knowledge and particularly focus on methods for the ontology development for different user profiles. Section 3 represents our semantic approach for the improved development of the domain ontology for different user profiles (the DOUP-method) and details its three levels. In section 4, we describe the C3R prototype aiming to illustrate the feasibility of our approach. Finally, we describe the ongoing works and the perspectives of our research.

2 TOWARDS THE DEVELOPMENT OF THE DOMAIN ONTOLOGY FOR DIFFERENT USER PROFILES

Our research on the development of the domain ontology for different user profiles is based on three scientific axes. First, we study the main problems of the development of a domain ontology, which is characterised by a large amount of the tacit knowledge. Second, we focus on the factors defining the different usage of the same generic domain ontology by different user groups. Third, we study the methods of the ontological modelling oriented different user profiles.

As a general rule, the development of an industry-oriented domain ontology is characterised by the following factors.

First, a large amount of the knowledge to be formalised is tacit (Polanyi, 1983). For example, the conformity-checking in construction is characterised by: (i) the “know-what” knowledge of the construction industry commonly known by architects; and (ii) the “know-how” knowledge of the checking process shared by conformity experts.
It is thus indispensable to explicit and interpret such tacit domain knowledge formalising it.

Second, the development of the domain ontology is driven by its (future) application. For that reason, the development of a domain ontology is often a part of some more global research task (e.g., conformity-checking modelling, semantic search, etc.). Moreover, the formalisms used for development should be seamlessly interconnected with the formalisms of these research tasks and should be based on interoperable standards. For example, the development of the buildingSMART ontology (Bell and Bjorkhaug, 2006) is coordinated with the elaboration of the SMARTcodes™, the code provisions for code compliance checking (Smartcodes, 2008), as well as regulation-centric knowledge representation formalisms define conceptual architecture of the conformance assistance framework (Kerrigan and Law, 2005).

Third, before formal representation, the domain knowledge should be first interpreted by domain experts. Even if the domain experts are the professionals of the domain, it is obvious that such interpretation remains rather subjective and/or partial. For this reason, it is important to validate the acquired domain ontology by usage.

Fourth, from the different point of view, the expert interpretation sometimes differs from the understanding of end users, who may find the knowledge “not adequate” and “difficult to use”, but fail to express the exact meaning of the concepts used (e.g., a user may find it difficult to distinguish between “main door” and “entrance door”, but does not use these two concepts in the same way).

Fifth; the domain ontology is often defined in a specified context (Hernandez et al., 2007), which should be then validated by usage.

The second axe of our analysis is devoted to the practical usage of the generic domain ontology by different user groups. Such usage may cause problems for the following reasons: (i) the interpretation of the domain knowledge by end users may be different from the interpretation of domain experts; (ii) different groups of end users may have different scope of interest (e.g., an architect and a legal authority need different level of detailing the conformity-related construction ontology); (iii) a large amount of knowledge remains tacit. In other words, it may be difficult for end users to define how they really need to use this knowledge (e.g., in the case of checking the conformity of a public building, a user may interest only in checking its accessibility, but not the acoustic requirements, which are the part of the global conformity-checking).

The problem of the development of a domain ontology for different user profiles represents our third research axe. A general approach for personalising the user's environment and integrating the user profiles into the development of information services is discussed in (Sutterer et al., 2008). The main methods for the automatic creating and application of user profiles are discussed in (Gauch et al., 2007). These methods allow integrating search results tailored to individual users to more complex systems and thus to personalise the application of such systems. In (Sieg et al., 2007), the authors propose a general approach for representing the user context by assigning interest scores to existing concepts in a domain ontology.

We focus on these three research axes aiming the development of the domain ontology for different user profiles to define our semantic approach for the improved development of the domain ontology for conformity-checking in construction, which allows the personalisation of the domain knowledge for different user profiles.

3 SEMANTIC APPROACH FOR THE IMPROVED DEVELOPMENT OF THE DOMAIN ONTOLOGY

Our semantic approach for the improved development of the domain ontology for different user profiles (the DOUP-method) has three levels:

1. Our knowledge representation and acquisition method (the KRA-method) developed for our conformity-checking model.
2. Our method of context modelling of the ontology by integrating the results of semantic search (the CMV-method).
3. Our approach for the adaptation of the domain ontology for different user profiles (the ECMV-method).

3.1 Knowledge Representation and Acquisition Method

We adopt the ontological approach and the semantic web technologies (Berners-Lee, 2001) to develop the knowledge representation and acquisition method (the KRA-method, cf. Figure 1) that allows us to represent complex and multidisciplinary knowledge characterising the conformity-checking process in construction. In this section, we briefly describe the
main ideas of our knowledge representation and acquisition method. A more detailed explanation and corresponding examples could be found in (Yurchyshyna et al., 2008a).

The acquired queries, however, contain only conformity constraints, but have no supplementary information, guiding the checking process: e.g. the scheduling of queries. The forth phase of our method aims thus at the development of semantic annotation of conformity queries. We propose a special RDF annotation of a query, developed according to its tag-based context: possible values for certain tags are concepts/properties of the conformity-checking ontology.

To do it, we combine two main methods of document annotation: annotation by content of the document and annotation by its external sources (Mokhtari and Dieng-Kuntz, 2008). First, we annotate a query by its content. To do this, we define a set of key concepts of this query, which describe what is really checked by this conformity requirement. In other words, we define keyConcept tag in the RDF annotation of a query, which value is a list of concepts from the conformity-checking ontology extracted from the SPARQL representation of this query. We remark also that there is a semantic correspondence between different types of knowledge used for query annotation. For example, a conformity query defining the physical dimensions of a door is annotated by a Door concept from our conformity-checking ontology.

Second, we annotate a conformity query according to external sources. Such annotation allows representing different types of knowledge. First, they are characteristics of the regulation text from which the query was extracted: (i) thematic (e.g. accessibility); (ii) regulation type (e.g. circular); (iii) complex title composed of the title, publication date, references, etc.; (iv) level of application (e.g. national), (v) destination of a building (e.g. private house). Second, they are characteristics of extraction process: (i) article, (ii) paragraph from which a query was extracted, (iii) current number (e.g. 3 query of 1 paragraph of Door article). Third, it is formalised expert knowledge: tacit « common knowledge » on the process of conformity-checking that is commonly applied by domain experts: (i) knowledge on domain and sub domain of the application of a query (e.g. Stairs); (ii) knowledge on checking practice (e.g. if a room is adapted, it is always accessible). Fourth, it is the application context of a query. This group specifies the aspects of query application for certain use cases. For example, the requirements on the maximal height of stairs handrail vary from 96 cm (for adults) to 76 cm (for kids). In this case, it is important to know the destination of a building (e.g. school).

Characteristics and possible values of the first two groups are automatically extracted from the CD
REEF. The knowledge described by the last two
groups is defined partially and/or has to be explicitly
formalised by domain experts.

Figure 2: Conformity-checking model.

The knowledge acquired by the KRA-method is
then used in our conformity-checking model that is
based on the analysis of matchings between the
representations of a construction project and of
conformity queries (Yurchyshyna et al, 2008a, cf.
Figure 2).

3.2 Context Modelling of the Domain
Ontology by Integrating the Results
of Semantic Search

According to the KRA-method, the conformity-
checking ontology is developed with the help of
domain experts and does not depend on the
conformity-checking process. All concepts and
relations of the ontology are defined and validated
by domain experts before the checking process and
can not be changed in the process of checking.
Domain experts also formulate rules of definition of
new concepts, context rules and, in general, they
validate the whole knowledge base of the
conformity-checking process.

In some cases, such definitions can be partial or
inadequate, and it does not represent the real usage-
driven conformity-related knowledge of the
checking process: even the definition of domain
experts is not sufficient to represent the whole
complexity of the checking knowledge. It is thus
important to propose an approach of the acquisition
of another type of the checking knowledge: the
knowledge on the checking practices, which turns
out explicit thanks to a large number of checking
operations by different end non-expert users.

To do this, we developed an approach of the
context-based modelling of the domain ontology,
which is validated by usage, the CMV-method
(Yurchyshyna et al, 2008c). In other words, we
proposed an approach for the evaluation of the
semantic proximity of different concepts/relations of
the conformity-checking ontology, according to the
interpretation of end users.

The CVM-method is based on the semantic
annotations of conformity queries. It aims to analyse
the simultaneous choice of the queries, which are
annotated by the same key concepts, and thus to
define the semantic similarity between these
concepts.

In the KRA-method, the semantic annotations of
conformity queries are developed according to the
tag-based context: possible values for certain tags of
semantic annotations of queries are concepts/properties of the conformity-checking
ontology.

The following example illustrates an annotation
of a query by the door concept of the conformity-
checking ontology.

```xml
<rdf:RDF xmlns:ontocC="domain.owl#"
<Annotation rdf:ID="">
<domainApplication>
<ontoC:Door/>
</domainApplication>
</Annotation>
</rdf:RDF>
```

Our work on modelling the domain ontology for
conformity checking in construction is conducted
under the Semantic Web vision, which guarantees
more advanced capabilities for processing the
knowledge. In particular, we also interest at more
advanced search, the so called semantic search that
gives better results in comparison to traditional
search mechanisms.

The semantic annotation of conformity queries
allows us to propose a user more detailed selection
of queries to be checked. For example, for a user
interested in checking the conformity of a door, we
can propose the semantic search that will give a
semantically richer result: it will interrogate the
domain ontology to define not only the queries
annotated by Door, but also all the corresponding
ones (its subclasses Entrance, EntranceDoor,
FrontDoor, AccessibleEntrance). It also means that
a user can obtain a semantically consistent answer
about the content of the conformity query before
executing it – only by its RDF annotation – and thus
to identify what he really wishes to check.

Technically, such semantic search is based on the
execution of the following SPARQL query against a
base of RDF semantic annotations.
In our example, the search of “door” expression will give the list of queries, which application domain contains “door” and classifies them according to the conformity-checking ontology. In comparison to the traditional search (that results with the only answer “door”), the semantic search will detail the application domain of found queries and classifies them into subclasses: (i) “door”; (ii) “entrance door”, “front door”, “entrance”; (iii) “accessible entrance” (cf. Figure 3).

The advantage of such semantic search is that it is defined according to the general domain knowledge of the construction industry, formalised in the conformity-checking ontology which is independent of an end user, but helps him to detail the search of corresponding conformity requirements and thus to refine the algorithms of their application during this process.

Another advantage of our approach for the semantic search of conformity queries is that the results of the semantic search followed by the user selection of a query can be then used to validate the initial domain ontology.

To illustrate these ideas by an example, let us take three subclasses of a Door class: FrontDoor, Entrance and EntranceDoor, which are defined equivalent in the conformity-checking ontology. They are also used as key concepts to annotate conformity queries (e.g. these three concepts annotate the query “an entrance door of any building should be accessible to disabled persons”). According to our model, for the checking of the conformity of an entrance door of a building, a construction project should be checked to the queries annotated by all these three concepts. A full list of these queries will be thus proposed to an end user.

In some cases, this list turns out redundant when an end user has no interest in some specific queries (e.g. the one concerning the luminosity of an entrance door of a school). It is, therefore, important to evaluate the cohesion between the queries chosen and rejected by an end user and the corresponding key concepts annotating these queries. For example, we can notice that queries annotated by Entrance and EntranceDoor are chosen more frequently than the ones annotated by Door. Intuitively, Entrance and EntranceDoor are semantically closer than Entrance and Door (cf. Figure 4).

To propose a formal definition of the validation of the conformity-checking ontology by usage, we first define our approach on the evaluation of the concepts of the conformity-checking ontology. It is based on three main criteria (Karoui et al, 2007) adapted for the conformity-checking problematic: (i) credibility degree: we suppose that all concepts and properties of the conformity-checking ontology are defined by construction experts, their definitions are pertinent and correct with the credibility degree equal to 1; (ii) cohesion degree: we suppose that our conformity-checking ontology is homogeneous: there are subclasses of a class which are declared equivalent by domain experts (e.g. door, entrance, entranceDoor); (iii) eligibility degree: concepts and relations are defined by experts and added to the conformity-checking ontology, if they are necessary for the formalization of conformity queries.

Our approach of the context-based validation of the conformity-checking ontology by usage is developed according to the same criteria, in order to keep the semantic consistency of the conformity-checking ontology: (i) credibility degree: no concepts or relations can be defined by end non-expert users; (ii) cohesion degree: the distance between the equivalent concepts is then recalculated according to the frequency of their simultaneous choice by end non-expert users (e.g. Entrance and EntranceDoor are chosen more often); (iii) eligibility degree: if some classes of semantically close concepts are defined, it can be interesting to identify the concept characterising the whole class, e.g. EntranceDoor for the class containing Entrance, AccessibleEntrance, FrontDoor, etc. By identifying the representative concept of the class, we can refine
the semantic annotation of the corresponding queries (for example, annotating them only by this concept) and, consequently, the algorithms of expert reasoning (for example, we do not need to schedule queries which are annotated by the concepts of the same class).

To model the semantic distances in the conformity-checking ontology, we base on the calculating of the semantic similarity in content-based retrieval systems (El Sayed et al. 2007) and by adapting the approach of the “intelligent evaluation” (Karoui et al., 2007) of ontological concepts. Currently, we work on the detailed development of the conceptual approach for the evaluation of the concepts of the conformity-checking ontology.

3.3 Adaptation of the Domain Ontology for Different User Profiles

Our method for the development and usage-based validation of the domain ontology remains still generic and not adapted to the variety of different actors of the construction domain. For this reason, it is a real challenge to propose an approach for adapting it for different user profiles; e.g. architect, electric engineer, legal authority, etc.

In order to adapt the acquired domain ontology for different user profiles, we propose to enrich our CMV-method by personalising it for different user profiles: the ECMV-method.

Our ECMV-method contains two main steps. First, we identify the groups of users and the corresponding user profiles. For each user profile, we create a copy of the initial generic domain ontology: e.g. the conformity-checking ontology for (i) architects; (ii) electric engineers; (iii) conformity-checking experts; and (iv) end non-professional users.

Second, we define the scope of interest for each user profile. To do this, we apply the CMV-method for each group of users and modify their copy of the domain ontology according to its usage by the corresponding end users. As result, we generalise or detail the domain ontology according to the scope of interest of user profiles. For example, it is important only for an architect to distinguish between different types of entrances (cf. Figure 5).

It is important to underline that the ECMV-method guarantees the coherence and semantic consistency between the generic domain ontology its facets developed for different user profiles. This coherence is based on the following aspects: (i) credibility degree: the credibility of the facets of the initial domain ontology is 0; the users can only refine the distances between the concepts of the initial domain ontology, but they can not create new concepts; (ii) cohesion degree: the distance between the synonym concepts is recalculated according to the frequency of their simultaneous choice by users of the same user profile (we do not aim at establishing correspondences between different user profiles); (iii) eligibility degree: if users of the same user profile define semiantically close concepts, these concepts are grouped by a representative concept which is the closest super class of these semantically close concepts (e.g. Entrance, EntranceDoor, FrontDoor classes of the initial domain ontology are grouped by Entrance class of the ontology facet for electric engineers).

4 IMPLEMENTATION: ENRICHMENT OF THE C3R PROTOTYPE FOR DIFFERENT USER PROFILES

In our previous work on the development of the conformity-checking model (Yurchyshyna et al., 2008a), we have developed the C3R (Conformity Checking in Construction with the help of Reasoning) system (cf. Figure 6) that implements the algorithms of reasoning by expert rules according to organised conformity queries. For the checking operation, C3R relies on the semantic search engine CORESE (Conceptual Resource Search Engine) (Corby, Faron-Zucker, 2007), which implements RDF, RDFS and SPARQL languages and answers SPARQL queries asked against an RDF/OWL Lite knowledge base (Sowa, 1984); and SEWESE (Seweze, 2008), the JSP/Servlet/Corese environment for building Semantic Web applications.

The main components of the C3R prototypes are: (i) the knowledge acquisition module (query
formaliser, ontology editor; construction project extractor); (ii) the reasoning module (checking reasoner enabled by the CORESE engine; query scheduler, conformity report generator); (iii) the module on capitalisation of context knowledge (query base generator; annotation editor; expert reasoning explorer; formaliser of usage-based knowledge).

For the C3R prototype, we have defined a conformity-checking ontology that currently comprises around 2200 concepts and 1600 properties. The conformity-checking ontology is written in the OWL-Lite language, which is rather expressive and, at the same time, decidable. We also define about 50 definition rules describing new concepts and properties with the help of the ones from the conformity-checking ontology.

To develop a base of conformity queries for the validation of our approach, we chose 9 regulation texts on the accessibility of public buildings (French regulation base). These regulation documents represent different classes of regulation texts (e.g. norm, circular) and describe the accessibility constraints of different entities: doors, routes, signalisation, etc. With the help of CSTB experts, we have identified about 350 simple text conformity queries that resume these 9 regulation texts, which are partially interpreted (around 65%). Other 35% of identified queries are classified as non-interpretable and are not formalised. For practical validation of our approach, we currently formalised and tested about 100 conformity queries as SPARQL queries.

To adapt the C3R prototype for different user profiles, we define 3 user profiles: architects, engineers, and owners/non-professional end users. For each user profile, we create a facet of the conformity-checking ontology and define. The calculation of the semantic similarity between the concepts of the conformity-checking ontology according to the DOUP-method is not implemented yet. It is an objective of our future work on the incremental implementation of the C3R prototype.

5 CONCLUSIONS

We presented a formal method for the development of the domain ontology for different user profiles in the context of conformity checking in construction. Our semantic approach for the improved development of the domain ontology for different user profiles (the DOUP-method) comprises three components: (i) the KRA-method aiming to represent the knowledge for the conformity-checking modelling; (ii) the CMV-method aiming context modelling of the ontology by integrating the results of semantic search of a query to check; and (iii) the ECMV-method, which adapts the CMV-method of the development of a domain ontology for different user profiles. We also described the conceptual architecture of the C3R prototype and presented the current work on its implementation, to illustrate feasibility of our approach.

One possible limitation of our work is that we do not establish the semantic correspondences between different facets of the domain ontology. This very interesting research problem is not taken into account by our semantic approach for the improved development of the domain ontology for different user profiles, and can be seen as a probable axe for future research.

Our future works focus on the further incremental development of the conformity-checking ontology and the C3R prototype, as well as their evaluation by domain experts and end users. In particular, we will detail the DOUP-method to adapt the C3R prototype for different user profiles, as well as to create different facets of the conformity checking ontology.

REFERENCES


cwese


