ARTICLE IN PRESS
AUTCON-02221; No of Pages 18
Automation in Construction xxx (2017) xxx–xxx
Contents lists available at ScienceDirect
Automation in Construction
journal homepage: www.elsevier.com/locate/autcon

Enhancing the ifcOWL ontology with an alternative representation for geometric data

Pieter Pauwels[a,*], Thomas Krijnen[b], Walter Terkaj[c], Jakob Beetz[b]

[a] Ghent University, Ghent, Belgium
[b] Eindhoven University of Technology (TU/e), Eindhoven, Netherlands
[c] Institute of Industrial Technologies and Automation (ITIA), Milano, Italy

ARTICLE INFO

Article history:
Received 2 August 2016
Received in revised form 9 January 2017
Accepted 1 March 2017
Available online xxxx

Keywords:
BIM
IFC
OWL
Linked data
Geometry
Data exchange

ABSTRACT

Over the past few years, several suggestions have been made of how to convert an EXPRESS schema into an OWL ontology. The conversion from EXPRESS to OWL is of particular use to the architectural design and construction industry, because one of the key data models in this domain, namely the Industry Foundation Classes (IFC), is represented using the EXPRESS information modelling language. These conversion efforts have by now resulted in a recommended ifcOWL ontology that stays semantically close to the EXPRESS schema. Two major improvements could be made in addition to this ifcOWL basis. First, the ontology could be split into diverse modules, making it easier to use subsets of the entire ontology. Second, geometric aggregated data (e.g. lists of coordinates) could be serialised into alternative, less complex semantic structures. The purpose of both improvements is to make ifcOWL data smaller in size and complexity. In this article, we focus entirely on the second topic, namely the optimization of geometric data in the semantic representation. We outline and discuss the diverse available options in optimizing the data representations used. We quantify the impact of these measures on the ifcOWL ontology and instance model size. We conclude with an explicit recommendation and give an indication of how this recommendation might be implemented in combination with the already available ifcOWL ontology.

1. Introduction

1.1. Interoperable data exchange for Building Information Models

Building information modelling (BIM) is one of the most notable efforts in years regarding information management in construction industry [1]. BIM environments allow to semantically describe any kind of information about the building in a common information environment. The Industry Foundation Classes (IFC) standard [2], developed and maintained by the buildingSMART organisation [3], aims at providing a central *"conceptual data schema and an exchange file format for BIM data"* [4]. Using the IFC data model and instance serialisation formats, BIM data can be exchanged between heterogeneous software applications covering a wide range of use cases including 4D planning, 5D cost calculation and structural analysis. In many common business scenarios, such files consist of partial domain models from different stakeholders that are exchanged frequently in iterative design and planning processes. For building projects in the later planning stages with high amounts of detail (reinforcement, ductwork etc.), for large buildings or buildings with complex geometries, instance files commonly consist of hundreds of thousands of objects and the resulting models result in large files that are resource-intensive to process.

The IFC data model is represented as a schema in the expressive EXPRESS data specification language defined in ISO 10303-11:2004 that *"consists of language elements which allow an unambiguous data definition and specification of constraints on the data defined and by which aspects of product data can be specified"* [5]. Currently, the most commonly used IFC schemas are IFC2×3 (IFC2×3_TC1.exp) and IFC4 (IFC4_ADD1.exp). IFC2×3 is important because it has been used for more than ten years in industry. Hence, numerous sample IFC2×3 implementations in widely adopted software tools are available and in use in practice. A good public real-world data set with IFC2×3 files is available in [6] (Dataset Schependomlaan). IFC4 is important because it is the last version of IFC and thus supersedes IFC2×3. As IFC4 is not yet widely implemented and/or certified in commercial software tools yet, no public IFC4 real-world sample files are available, besides the ones provided as part of the IfcDoc tool [7]. The IfcDoc sample files however include many of the less commonly used

* Corresponding author.
E-mail addresses: pipauwel.pauwels@ugent.be (P. Pauwels), t.f.krijnen@tue.nl (T. Krijnen), walter.terkaj@itia.cnr.it (W. Terkaj), j.beetz@tue.nl (J. Beetz).

http://dx.doi.org/10.1016/j.autcon.2017.03.001
0926-5805/

Please cite this article as: P. Pauwels et al., Enhancing the ifcOWL ontology with an alternative representation for geometric data, Automation in Construction (2017), http://dx.doi.org/10.1016/j.autcon.2017.03.001

IFC data types, including lists of lists, binary representations, tesselated geometry, and so forth. In this article, we will use these sample IFC4 files from IfcDoc as well as the IFC2×3 data set in [6], leading to a set of sample files that is representative for IFC2×3 as well as IFC4, and that is representative for real-world IFC files as well as semantically less common data. All test files are made available at Pauwels et al. [8].

1.2. Background: the ifcOWL ontology

In 2013, the latest version of the IFC schema (IFC4) was published into an ISO standard (ISO 16739) not only as an EXPRESS schema, but also as an XSD schema. The main objective was to make the data model more easily available for flexible usage in XML-based environments. Due to the verbosity of the XML document format and the performance limitations of the commonly available XML processing tools and programming libraries however, this representation has been deemed inappropriate for most practical use cases and was never fully embraced by the industry. In order to benefit from the built-in capabilities to modularise and distribute models across file boundaries and network structures, to harness reasoning and standardized query capabilities and to easily integrate further vocabularies and data sets with BIM models, Semantic Web and Linked Data technologies have come into the focus of numerous research efforts.

It is not the purpose of this article to list all the possibilities that are made available from the mere usage of linked data or semantic web technologies. Yet, Pauwels et al. [9] gives a broad overview of these possibilities, while referring to numerous example implementations worldwide. This article lists examples in three categories: (1) interoperability, (2) linking data across domains, and (3) logical inference and proofs. The usage of semantic web technologies appears to enable use cases mostly in the latter two categories. When a lot of links are made across domains (with IFC being just one of those domains), externally managed product manufacturer data is more tightly connected to building models; improved building performance analysis can be targeted; regulation compliance-checking is achievable; a link with geographical and infrastructure data is possible; and so forth. These use cases make considerable use of the ease of linking data with Linked Data and Semantic Web technologies, as well as the out-of-the-box query functionality. In the case of logical inference and proofs, regulation compliance-checking is a use case often mentioned, in which the logical model of the building (e.g. an ifcOWL representation) can naturally be combined with a rule set that represents the logical model of the building regulation. Checking the compliance with building regulations is then just a matter of starting a logic-based reasoning engine and querying for the compliance-checking result.

Using OWL as a schema modelling language, an ontology for the IFC data model was proposed by numerous authors [10–16]: ifcOWL. These different efforts are currently converging into an agreed common standard. The proposed conversion effort hereby specifically aims to keep the resulting OWL ontology as close as possible to the original EXPRESS schema of IFC. This conversion approach is documented in full detail in [16], including an extensive literature review and comparison of approaches. The proposed ifcOWL ontology is now picked up within buildingSMART International, where it might eventually become a part of the ISO 16739 standard, similar to the way in which the XSD schema became part of this standard. As a result, the IFC data model would be available in EXPRESS, XSD, and OWL (see Fig. 1), allowing the representation and usage of building data in STEP Physical File Format (SPFF), eXtensible Markup Language (XML), and the Resource Description Framework (RDF).

The latest edition of the ifcOWL ontology can be found in the buildingSMART International web pages [17]. The ifcOWL directly imports the EXPRESS ontology [18] and indirectly imports the LIST ontology [19] that was modelled after [20]. The key features of the conversion pattern are summarized in Table 1, while referring to the prefixes defined for the IFC4_ADD1 ontology in Fragment 1. Specific further details, in particular regarding the conversion of aggregation data types, are provided in the following sections.

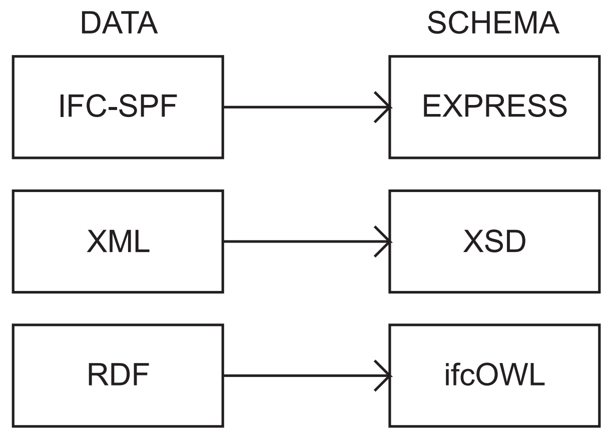

Fig. 1. The IFC data model is available in EXPRESS, XSD, and OWL, allowing to capture and use building data using three different technologies.

1.3. The modular structure of IFC

The current ifcOWL ontology [17] is a direct mapping of the IFC EXPRESS schema. The IFC4_ADD1 (IFC2×3_TC1) EXPRESS schema contains 768 (653) ENTITY data types, 206 (164) enumeration data types, 60 (46) select data types, 131 (117) defined data types, 46 (38) FUNCTION declarations, and 2 (2) RULE declarations. This results in an ifcOWL ontology for IFC4_ADD1 (IFC2×3_TC1) with 1313 (1093) classes, 1580 (1422) object properties, 5 (5) data properties, 13,867 (11,790) logical axioms, and 1158 (1018) individuals. The ontology is thus considerably big and complex to load and use. Furthermore, the ontology takes full advantage of OWL2 DL expressivity ($\mathcal{SHIQ}(\mathcal{D})$), which can lead to a high number of assertions when handed to OWL reasoning engines. By consequence, when the ontology is referenced by an instance file in RDF, all 1313 classes, 1580 object properties, 13,867 logical axioms and so forth are loaded. Moreover, it might be necessary to produce all available OWL2 DL assertions as well, leading to further overhead and delay in any software application.

Therefore, it would make sense to split the IFC ontology in separate modules, or separate smaller ontologies, so that end users and applications only need to load those ifcOWL modules that are actually going to be used. However, many of the entities and types in the IFC schema are tightly interconnected between the diverse sub-schemas. Hence, in order to make a useful modularisation, a full investigation of the schema needs to be made, and the relation between the different modules would need to be reconsidered to a significant level and detail. Such an investigation is out of scope for this research and paper. Nevertheless, we wish to mention that the modularisation of ifcOWL is an important next step in restructuring the ontology so that it can be more efficiently used in a web context.

1.4. The representation of geometry in IFC

A large part of a typical IFC-SPF file is devoted to the representation of geometric structures, usually employing aggregation data types (e.g. ordered lists of points in Cartesian points, ordered lists of Cartesian points in polylines, and so forth). Therefore, if one aims at reducing the size and complexity of an RDF graph based on ifcOWL, two important options are available: (1) removing the geometry