Krist V. Gernaey, Jakob K. Huusom and Rafiqul Gani (Eds.), 12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering.
31 May – 4 June 2015, Copenhagen, Denmark

On the Process of Building a Process Systems Engineering Ontology Using a Semi-Automatic Construction Approach

Canan Dombayci[a], Javier Farreres[b], Horacio Rodríguez[b], Edrisi Muñoz[c], Elisabet Capón-García[d], Antonio Espuña[a], Moisès Graells[e]

[a]*Chemical Engineering Department, ETSEIB, Universitat Politècnica de Catalunya, 647 Diagonal Avenue, 08028 Barcelona, Spain*
[b]*Software Department, CEIB, Universitat Politècnica de Catalunya, 187 Comte D'Urgell, 08036 Barcelona, Spain*
[c] *Centro de Investigación en Matemáticas A.C., Jalisco S/N, Mineral y Valenciana, 36240 Guanajuato, México*
[d] *Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 1, 8093 Zürich, Switzerland*
[e] *Chemical Engineering Department, EUETIB, Universitat Politècnica de Catalunya, Comte d'Urgell 187, 08028 Barcelona, Spain*
moises.graells@upc.edu

Abstract

This work presents a novel systematic approach for the construction of domain ontologies. The suggested approach uses a semi-automatic construction methodology. For this study, parent-child concept pairs are taken from a previous work. Novel contributions include building and completing branches, introducing new relations, and resolving inconsistencies and contradictions. For the process systems engineering (PSE) domain the ISA88 Standard is chosen as a promising starting point for automatic text processing. Finally, this work concludes with a discussion of the ISA88 Standard based on the conclusions that can be obtained from the application of this semi-automatic construction methodology.

Keywords: Domain Ontology, Automatic Ontology Construction, Ontology Construction Methodology, ISA88, Standards.

1. Introduction

There has been a growing interest in the field of PSE for using ontologies as knowledge models (Muñoz, 2011), intelligent data analysis of databases (Roda and Musulin, 2014), intelligent software applications (Morbach et al., 2010) and many other applications, each one using different ontology models. The use of a general domain PSE ontology based on already existing and accepted standards would not only facilitate the development of these and other applications, but also their integration and coordination. Departing from the parent-child concept pairs resulting from a previous work (Farreres et al., 2014), this work describes the systematic procedures used to refine the relationships to be modelled in the core of the ontology. Particularly, an ontology can be defined as "an explicit specification of a conceptualization" (Gruber, 1993). In the field of Computer Engineering, large ontologies for common knowledge have been developed together with tools based on internet databases, reference books and many documents from the Internet

(Vivaldi and Rodríguez, 2002). This is expected to lead to semi-automatic procedures to help experts to develop domain ontologies, although a great effort is still required to filter the noise obtained from extensive indiscriminate searches.

In parallel, expert teams have devoted a commendable effort to produce standards. Using language analysis tools, and in order to combine efforts for creating domain ontologies, the texts defining standards seem very well suited to be taken as a source to synthesize manual and automatic approaches in the most efficient way. Thus, the aim of this work is to investigate the semi-automatic development of domain ontologies in general, and to validate the hypothesis that a document defining a standard is an efficient starting point from which the ontology may be later extended and enriched.

The ISA88 Standard for batch control is a document that is assumed suitable for creating a domain ontology because it is a technical document clearly defining models and terminology, including the explanation of processes, data structures and language guidelines. The assumption is that intelligent selection of texts will reduce noise and allow fast and straight identification of concepts and relations.

In a previous work (Farreres et al., 2014), the ISA88 document was processed using a number of tools, and a set of concept pairs were extracted and manually revised. Each pair was a proposed relation, either between a parent concept and its children, or between a concept and its parts. This study is a continuation of such preliminary work for building a domain ontology. This building process is explained in a systematic way describing the suggested methodology and the quality of the resulting ontology is assessed.

2. Methodology

Common strategies for identifying concepts for a domain ontology manage concepts from the most specific ones to the most general ones (also reverse) or from the most important concepts to the most specific and most general ones (López, 1999 and Corcho et al. 2003). Conversely, this work identifies the concepts from a technical standard that is considered the basis for the domain. As a result, the proposed construction methodology is significantly different from the other strategies, and the ontology is built from these identified concepts by using a semi-automatic construction methodology. The following sections present a brief description of the steps of the methodology.

2.1. Phase-1: Building branches by composing parent-child concept pairs

A core ontology is built using parent-child concept pairs from Farreres et al. (2014). In the case of ISA88 Standard 266 concepts are added with 188 ‘is a’ relations. The ontology is enriched with the meronymy (‘partOf’) relations and 81 doubtful cases are left apart.

2.2. Phase-2: Doubtful cases, introducing ‘partOf‘ relationship and information from the figures in the standard

First, doubtful parent-child concept pairs (e.g. concepts with ‘and’ and ‘or’) are identified and solved manually. Next, graphical information is extracted by a human expert and the newly detected relations and concepts are introduced to further enrich the ontology. The ISA88 Standard contains figures that cannot be processed via automatic pattern matching owing to the fact that figures cannot be processed by the computers. This is revealed as a

difficulty for the automation of the process and it is solved manually processing the standard by a human expert.

2.3. Phase-3: Adding commonsense and ontology pruning

Until Phase-3 knowledge introduced in the ontology was all from the text and graphics of the supporting Standard. However, this may include neither commonsense nor the knowledge shared within a domain that must not be conveyed in a text because of its obviousness. After Phase-1 and Phase-2, conceptual gaps occur as consequence of the lack of explicit commonsense knowledge in the text.

Some simple and straightforward cases of commonsense knowledge are then addressed using a set of pre-established rules for introducing implicit information that can be understood without any explanation. A particular case is identified and solved following the rules 1 and 2. Thus, the UnitRecipe concept, composed of two nouns (1), would be moved as a child of Recipe; and RegulatoryControl, composed of an adjective and a noun (2), would be moved as a child of Control. 141 concepts were refined in this way and additional 'is a' relations are added to the ontology.

$$\text{Noun}_A + \text{Noun}_B \xrightarrow{\text{'is a'}} \text{Noun}_B \quad (1)$$

$$\text{Adjective}_A + \text{Noun}_C \xrightarrow{\text{'is a'}} \text{Noun}_C \quad (2)$$

Another particular case led to the creation of additional 'partOf' relations. Container, Combination, and similar concepts are often extracted as parent concepts from 'is a' relations. However, these concepts indicate the aggregation of further concepts. In this case the actual relation that correctly represents the model is not an 'is a' relation, but a 'partOf' relation. In the case of the ISA88 based PSE ontology, 20 concepts were refined concluding with adding 34 new 'partOf' relations and 22 new concepts.

Another case of commonsense rule will lead to the identification (Zhu et al., 2009) and elimination of redundant relations already presented by other relations.

3. Example of methodology (GeneralRecipe segment)

Figure 1 shows the 'is-a' segment originating in GeneralRecipe after Phase-1 and Phase-2. It is next used as an example to discuss the methodology steps detailed as follows:

i). Container, Collection and Combination concepts: Figure 1 shows GeneralRecipe as a subconcept of Container. As explained in Section 2.3, Container is better modelled by a 'partOf' relation. Figure 2 shows the new added 'partOf' relations.

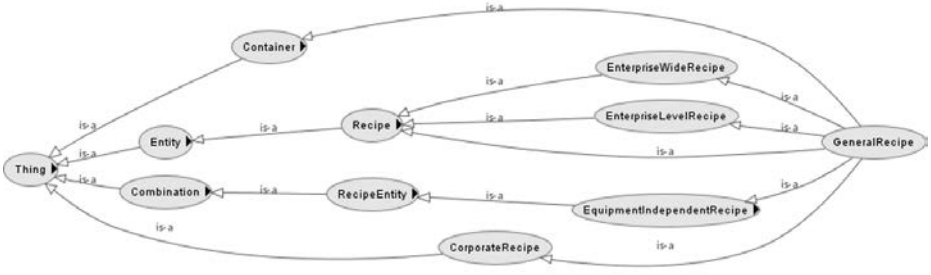

Figure 1. GeneralRecipe segment after Phase-1 and Phase-2

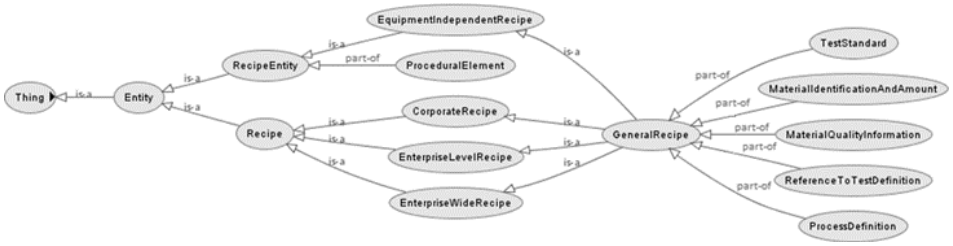

Figure 2. GeneralRecipe segment after Phase-3 (Only 'is a' and 'partOf' relations related with GeneralRecipe segment are represented, in case of clarity)

ii). Introducing commonsense knowledge: In addition, concepts given by compound nouns need reviewing (Section 2.3). For instance, it is apparent that the EnterpriseWideRecipe concept should be moved as a child concept of Recipe as shown in Figure 3a as before. Figure 3b shows the result.

iii). Removing redundant relations: Since there are two ways to reach Recipe from GeneralRecipe, the relation between GeneralRecipe and Recipe can be removed. This redundant relation is shown in Figure 3b.

Finally, Figure 2 shows the final diagram after implementing the proposed methodology.

4. Quality measure results

The direct pattern matching of the raw text after Phase-1 and Phase-2 results in a flat ontology (Figure 4a), lacking commonsense. Although there are no standard metrics for the quality of ontologies, some topological information have been used for measuring. One topological aspect taken into account is the width and depth of the ontology. Figure 4b shows a partial snapshot of the ontology after Phase-2 and Phase-3. It is clear that the first ontology is flat while the second one is deeper and tree-like, which has been considered as a sign of quality and improvement.

Additionally, Figure 5 shows the differences in the number of concepts per level after each Phase. Phase 1 generates an ontology that has most of its concepts at depth 2. After Phase-2 the number of concepts increases but there are few concepts below depth 2. After adding commonsense, the distribution is more reasonable, with a spread number of concepts between level 1 and 4, and a significant queue arriving to depth 6.

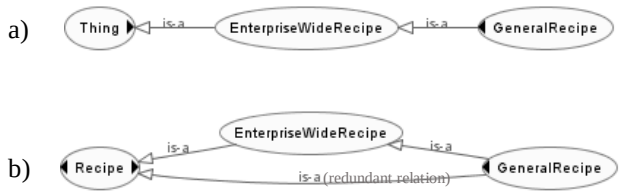

Figure 3. Redundant relations. a) Before introducing commonsense knowledge b) After introducing commonsense knowledge with redundant relation

a)

b)

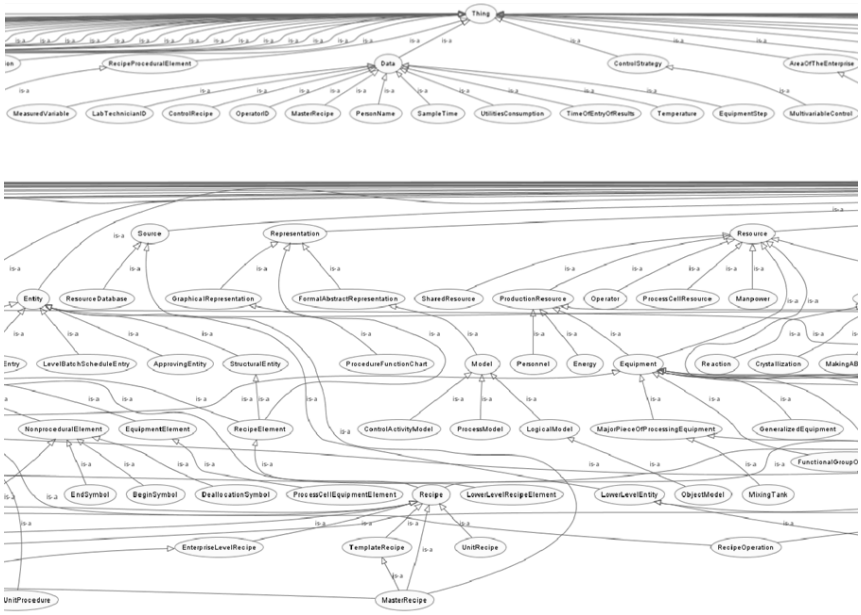

Figure 4. Depth of the ontology. a) After Phase-2. b) After Phase-3

Finally, the improvement obtained by using a text defining a standard (regarding the reduction of noise) is an assumption that can be hardly validated in a quantitative and systematic way. However, the use of another text could provide an illustrative reference. Towards this end, the same methodology was used to process a play of Shakespeare (1611) and nothing was produced: no pattern for the 'is a' or the 'part of' rules is found in the text.

5. Conclusions and future research

This paper proposes a semi-automatic methodology for building ontologies, presents most of the concepts of batch control in PSE by using the ISA88 Standard, and introduces commonsense knowledge to the ontology. The use of technical standards for building domain ontologies has been assumed. Another significant outcome raised in this study is the human expert introduction to the automatic ontology construction. An additional interesting result is that other relations ('input' and 'output') have been detected in addition to the 'is a' and 'partOf' relations for the enrichment process of ontology.

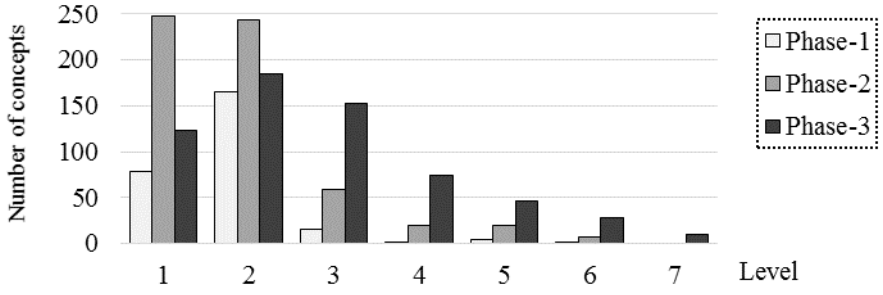

Figure 5. Number of concepts per level

Moreover, the ISA88 Standard consistency could be assessed from the point of view and formalisms of ontology development, and conclusions and guidelines for producing technical standards have been obtained. This includes ambiguous figures, undefined concepts, extended use of synonymy, and use of adjectives in the document which have been identified as problems, not only for the implication of automatic processes, but also for human readers. This contributes an analysis and suggestions for developing technical standards. On the other hand, the ISA88 Standard should be considered as a first case study for a research line, and the learning outcomes from this work are expected to be of practical interest to further ontology developments from other documents (e.g. ISA95).

Further work in this research line includes the enrichment of the ontology by extending the present semi-automatic development of domain ontologies for PSE using other technical documents and automatically searching the Internet to get implicit knowledge. The outcome of this systematic procedure could afterwards be compared to already existing ontologies of the same domain developed manually (Muñoz et al., 2011). Using some similarity metrics in order to validate their completeness. Hence, relevant feedback could be retrieved in order to further improve the systematic approach presented. Furthermore, the findings of this study provide information for developing the methodology as well as improving quality of resulting ontology from ISA88 Standard.

Acknowledgements

Financial support received from the Spanish "Ministerio de Economía y Competitividad", the “Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR” and the European Regional Development Fund, funding the research Projects EHMAN (DPI2009-09386), SIGERA (DPI2012-37154-C02-01), SKATER (TIN2012-38584-C06-01) and the Research Group CEPEiMA (2014SGR1092), is fully appreciated.

References

O. Corcho, M. F. López, A.G. Perez, 2003, Methodologies, tools and languages for building ontologies. Where is their meeting point?, Data & Knowledge Engineering, 46, 41–64.

J. Farreres, M. Graells, H. Rodríguez, A. Espuña, 2014, Towards Automatic Construction of Domain Ontologies : Application to ISA88. Proceedings of the 24th European Symposium on Computer Aided Process Engineering, Budapest, Hungary.

T.R. Gruber, 1993, A Translation Approach to Portable Ontology Specifications, Knowledge Acquisition, 5, 2, 199–220.

ISA, 2000, Enterprise-Control System Integration Part 1: Models and Terminology, ANSI/ISA–95.00.01–2000, Technical report, ISA Standard.

ISA, 2006, Batch control, Part 1 (2006), Part 2 (2001), Part 3 (2003), Part 4 (2006), (ISA-88.01-1995 (R2006)), Technical report, ISA Standard.

M. F. López, 1999, Overview of Methodologies For Building Ontologies, Laboratorio de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid.

J. Morbach, W. Marquardt, A. Wiesner, A. Yang, 2010, Onto CAPE - A Re-Usable Ontology for Chemical Process Engineering, Springer.

E. Muñoz, 2011, Knowledge management technology for integrated decision support systems in process industries, Ph.D. Thesis, Univesitat Politècnica de Catalunya, Spain.

F. Roda, E. Musulin, 2014, An ontology-based framework to support intelligent data analysis of sensor measurements, Expert Systems with Applications, 41, 7914–7926.

W. Shakespeare, 1611, Macbeth <http://shakespeare.mit.edu>, Accessed on 20/09/2014.

J. Vivaldi, H. Rodríguez, 2002, Medical Term Extraction using the EWN ontology, Terminology and Knowledge Engineering, 137-142.

J. Zhu, J. Wang, B. Li, 2009, A formal method for integrating distributed ontologies and reducing the redundant relations, Kybernetes, 38, 1870-1879.