Twenty Years of Configuration Knowledge Modeling Research. Main Works, What To Do Next?

M. Mohammad Amini, Michel Aldanondo, Élise Vareilles, T. Coudert

Twenty Years Of Configuration Knowledge Modeling Reasearch. Main Works, What To Do Next?

M. Mohammad Amini[1], M. Aldanondo[2], E. Vareilles[3], T. Coudert[1]
[1]INP-ENIT, University of Toulouse, Tarbes, France
[2]IMT Mines Albi, University of Toulouse, Albi, France
[2]ISAE-SUPAERO, University of Toulouse, France
(maryam.mohammadamini@enit.fr, thierry.coudert@enit.fr, elise.vareilles@isae-supaero.fr, michel.aldanondo@mines-albi.fr)

Abstract - **A configuration software (configurator) associates a knowledge base (KB) with a knowledge processing unit (PU). The KB describes all possible combinations of components while the PU overlays this knowledge with the customer requirements. Our work deals with the KB and the approaches, models, or tools for modeling configuration knowledge. Our goal is to present a small quantitative literature survey highlighting two work streams: the first one gathers modeling works dealing with constraint-based approaches while the second deals with ontologies, description logic, or object-oriented modeling approach. We will also consider hybrid approaches. We will present a quantitative analysis of published materials in Web of science over the last twenty years. The keywords occurrence versus time will also be studied in detail to identify tendencies in configuration knowledge modeling..**

Keywords - **Configuration knowledge modeling, constraints satisfaction problem, ontology, UML, OWL, rules**

I. INTRODUCTION

Nowadays, in one hand, mass customization plays an important role in competitive environments. In fact, mass customization helps to efficiently answer customers' demands when they want it, and at a competitive price. On the other hand, the context of industry 4.0 implies that companies have to answer quickly to many customers' demands and to offer solutions which match perfectly numerous and often complex requirements. In this context, configuration tools play an important role that will permit to satisfy the customer requirements.

This article deals with product configuration or customization. Configuration is a kind of design activity where products are defined given predefined sets of components linked by compatibility constraints [1]. Configuration activity is supported by software tools called Product Configuration Software (PCS) [2]. A PCS gathers (i) a knowledge base (KB) which contains what is called a generic model of the product and (ii) a processing unit (PU) that interacts with the user to assist him during the configuration activity. The generic model contains all component families and product attributes with all compatibility constraints that modulate their possible combinations. One generic model represents one family of products with all possible options and alternatives. The processing unit (PU) is basically responsible for constraint propagation. This means that for each requirement inputted by the user, the constraints are propagated and reduce the definition domain of the other variables linked by the constraints. Consequently, after each requirement input, the solution space or product family size is iteratively reduced and leads to a product solution that satisfies the customer. For more details, do not hesitate to consult [1].

Many works have been published about the processing unit of PCS (constraint propagation, answer set programming, inference engine, rule-based processing…), which is not the same with the modeling of the generic model necessary to set up product configurators [3]. Our purpose is to quantify the amount of work on this subject and what could be interesting tendencies. This is achieved with a quantitative literature survey that deals with two workstreams relevant to this subject: the first one gathers modeling works dealing with constraint approaches while the second deals with ontologies, description logic, or object-oriented modeling approaches. Hybrid approaches that associate will be also considered.

As an example, we will explain several articles in the realm of workstreams:

[4] used description logics-based languages such as Ontology Inference Layer (OIL) and DARPAAgent Markup Language to represent configuration knowledge in the context of the semantic web. Unified Modeling Language (UML) which is a useful modeling language, has also been applied to graphically represent configuration knowledge.

[5] introduced a domain ontology called Kumbang to model variability from both feature and architectural perspectives in software product families. Kumbang can also be used to develop modeling languages, manage variability in software product families based on customer requirements. Natural language and a UML were used to describe Kumbang.

[6] used the Unified Modeling Language and the Object Constraint Language (UML/OCL) as standard configuration knowledge representation languages to represent configurable products then integrate them into the Model Driven Architecture (MDA). In the end, UML/OCL has been applied in several industrial projects.

The following two articles deal with Hybrid approaches:

[7] represented an ontology-based approach for modeling product configuration knowledge. It used Web Ontology Language (OWL) to define different classes and their relationship to formalize product configuration then used

Semantic Web Rule Language (SWRL) to define the constraint and after that used JESS. The proposed approach was applied to configuring the ranger drilling machine. Protégé was used in this article which is an OWL-based tool that could be helpful to model and edit ontologies. [8] represented an Ontology-based method for product configuration knowledge using semantic web technologies. It also used OWL, SWRL, and a rule engine called JESS to improve the product configuration system. In this paper, an actual configuration engine based on derived configuration knowledge was developed as well. The approach was applied to a case for the personal computer. [9] and [10] are two good review articles on product configuration that presented various definitions in this area as well as future roadmaps.

The organization of the paper is as follows. We will first present and discuss the main elements of our survey: keywords, way to use keywords, and literature sources. Then we will present and discuss various literature investigations. We will finally conclude on possible ideas for future works.

II. SURVEY MAIN ELEMENTS

This section discusses the selected keywords, various ways to use them, and finally the article's sources.

A. Identifying keywords

We are dealing with knowledge modeling in the context of product configuration. As previously said, we want to consider and differentiate the works relevant to the model structure and variable identification from those that are more centered on relations or dependencies between configuration variables. Furthermore, we want to have keywords with some increasing strength or filtering level. Consequently, we have identified six groups of keywords:

1. Two less restrictive strength groups of keywords:
   - Group of keywords 1, kw1= configuration and knowledge
   - Group of keywords 2, kw2= configuration and knowledge and modeling
2. One group of keywords relevant to model structuring and variable identification:
   - Group of keywords 3, kw3= configuration and knowledge and (ontology or UML or description logic)
3. Two groups of keywords more relevant to relations or dependencies between configuration variables:
   - Group of keywords 4, kw4= configuration and knowledge and constraint
   - Group of keywords 5, kw5= configuration and knowledge and (OWL or Rule)
4. One highly restrictive group of keywords that considers simultaneously constraint with another previous keyword:
   - Group of keywords 6, kw6= configuration and knowledge and constraint and (ontology or OWL or Rule).

B. Using keywords in queries

Once keywords are defined, we have to decide about how to use them. They can be used to select publications concerning three possible querying fields: article title, article keywords, and article abstract.
Another field "article topic" aggregates the three querying previous fields. Of course, abstract is much less restrictive than title or keywords. We will consider mainly the query on field “topic” but in section 4 we will also discuss the results of query on field “title”.

C. About literature sources

At the beginning of the work, we have been considering three kinds of sources: (i) the web of sciences, (ii) Mendeley and/or Scopus , (iii) CEUR workshops [11]. The main reason to consider the web of sciences is because of its seriousness in the quality of the selection of the indexed elements. Furthermore, its website is quite well organized to filter and categorize queries.

The reason to consider Mendeley and/or Scopus was to avoid relying on the single WoS knowledge base. However, very quickly we face some difficulties in using and understanding Mendeley's results. Furthermore, if Mendeley is easy to access, this is not the case for Scopus.

The last source we wanted to consider was CEUR proceedings, because the knowledge-based configuration workshops community, that runs for more than twenty years, publishes rather frequently their proceedings with CEUR. But it seems that some years are missing because the workshop has been hosted by some conferences like ECAI or CP that prevent open publications of the workshop proceedings.

Consequently, and this is a drawback of this study, the results will only rely on the WoS database, more exactly on the Science Citation Index [12]. In section 4, Science Citation Index and on Conference Proceedings Citation Index queries will be compared.

III. FIRST INVESTIGATIONS

This section presents and discusses the results relevant to the six groups of keywords, with queries on the Science Citation Index considering the field “topic” that gathers: title, keywords and abstract.

A. First raw results

The figure 1 shows the number of articles per year (between 2000 and 2020) on the upper part less restrictive keywords (kw1 and kw2) and on the lower part all the others. In fact, the lower part is just a zoom of the upper part for kw 3,4, 5 and 6.

The upper part dealing with keywords kw1 “configuration and knowledge” and kw2 “configuration and knowledge modeling” show a quite smooth and regular increase. The lower part is not so smooth, but the

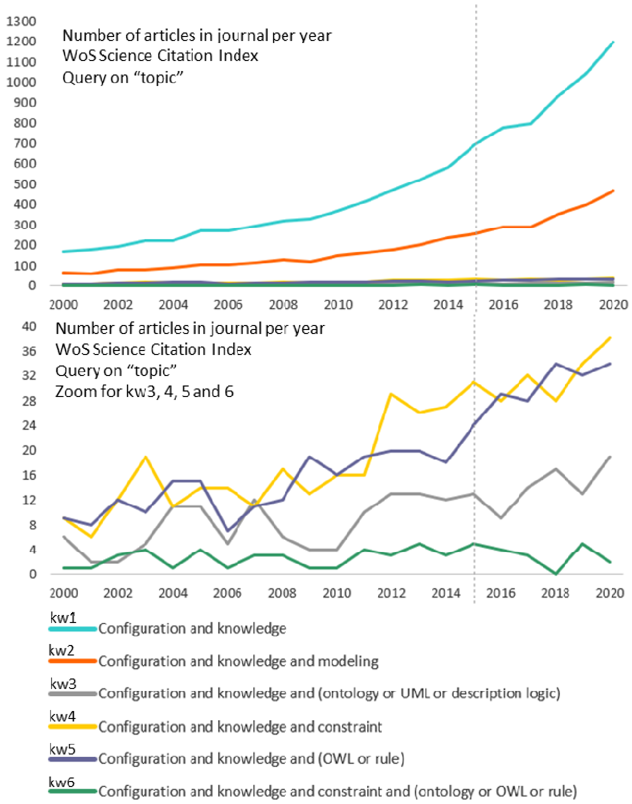

Figure 1. Journal articles / year – WoS-SCI – on topic

tendencies for kw4 "configuration and knowledge and constraint" and kw5 "configuration and knowledge and (OWL or Rule)" shows a regular increase. On the opposite, kw3 "configuration and knowledge and (ontology or UML or description logic)" show a very light increase while kw6 "configuration and knowledge and constraint and (ontology or OWL or Rule)" globally remain constant.

B. Discussion about restrictive keywords

Logically, a very large number of articles can be associated with less restrictive keywords. But a quick look shows that many articles come from scientific domains (biology, medicine…) not related to our topic (engineering, design…). However, the rate of increase sounds quite strong for keywords kw1, 2, 3, and 4. Three possibilities to explain that: (i) more people work on this subject, (ii) people publish more, and (iii) Web of Science indexes more journals. It is probably the association of the three phenomena that explains previous growth. However, we compute some kind of yearly increase for our domain and compare it with simple keywords as "Engineering" and "medicine" as shown in table 1.

We can see that the global increase for less restrictive keywords (kw1 and kw2) is with the same order of magnitude with generic keywords as "engineering" or "medicine". All rates are between 9.9% and 10.3%. For other keywords (kw3, 4, and 5), it is clear that the increases are much lower All rates are between 5.2% and 7.4 %. While keyword 6 is not computable. To compare with a quite hot topic of the moment, the computation has been done for the keyword "machine learning" and shows a yearly or rate of 25%.

It seems to be possible to conclude that:

1. Queries with less restrictive keywords just follow global publication quantitative evolutions.
2. Keywords dealing with relations or dependencies between configuration variables (kw4 and kw5) attract more articles than those dealing with model structuring and variable identification (kw5).
3. Works mixing constraints and model structuring are clearly less frequent.

TABLE 1.
The yearly rate of increase of article quantities

Keywords	Number of publications in year 2000	Average number of publications 2019-2020	Yearly rate of increase
Engineering	9495	64164,5	0,103
Medicine	8263	52238	0,099
kw1 - Configuration and knowledge	169	1119	0,102
kw2 - Configuration and knowledge and modeling	64	429,5	0,103
kw3 - Configuration and knowledge and (ontology or UML or description logic)	6	16	0,052
kw4 - Configuration and knowledge and constraint	9	36	0,074
kw5 - Configuration and knowledge and (OWL or rule)	9	33	0,069
kw6 - Configuration and knowledge and constraint and (ontology or OWL or rule)	1	3,5	no meaning
Machine learning	384	28753,5	0,248

IV DISCUSSION ON SOURCES

We will first compare; journal and conference proceedings survey then deal with investigations on article title or topic.

A. Comparing journal/conference proceedings

In figure 2 we show the result of a query similar to figure 1 except that it considers WoS proceedings. Figure 1 and 2 should be compared.

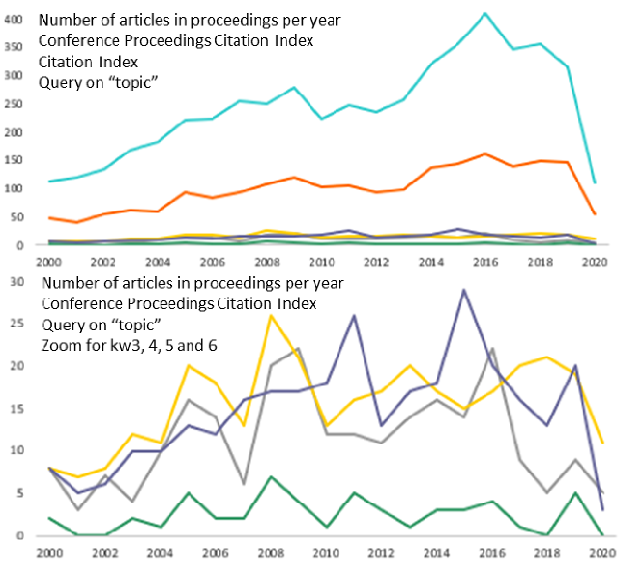

Figure 2. Comparing journal and conference proceedings.

We can see that the quantity of articles is much lower. This can be explained by the fact WoS-SCI indexes much less conference than journals. The quite regular increases of the number of publications which were seen in the result of journals are not the same in the result of proceedings. If an increase can be seen for less restrictive keywords (kw1 and kw2), this is not the case for all other keywords. Figure 2 shows also two phenomena, the economic crisis of 2008 and the Covid crisis of 2019, that seems to induce clear decreases in the years following them. This is especially true for 2020 with a terrible drop of more than 50%. Consequently, this might have pushed authors to consider more journal publishing than conference publishing.

B. Comparing queries on topic and title.

To study that, we have compared queries on article ‘topic” with queries on article “title” on the WoS-SCI journals. Ot shown results show that the quantities of identified articles are roughly divided by an order of magnitude of one hundred. This shows clearly, that making queries on title is too restrictive to derive interesting conclusions.

C. Conclusion on queries parameters

Considering previous results, we only consider queries on journal “articles” or or WoS-SCI made on the “topic” field. However as said in the end of section 3 many papers are not belonging to our subject interest. Next section will therefore add some filters on WoS-SCI queries on “journal” on “topic”.

V FILTERING RESULTS

The idea is to remove out a maximum of out-of-scope papers, for example, papers dealing with medicine, biology, law, business… medicine, arts… For this purpose we have used the categories filters proposed by the WoS website, then in a second step, we refine the results with abstract reading. All the results of this section are shown in figure 3.

A. Filtering with WoS categories

Thirteen WoS categories have been selected to filter results of journal articles queries on field “topic” .
They are:

1. Computer science artificial intelligence
2. Engineering electrical electronic
3. Computer science interdisciplinary application
4. Engineering manufacturing
5. Computer science information systems
6. Engineering multidisciplinary
7. Computer science software engineering
8. Computer science theory methods
9. Operations research management science
10. Engineering mechanical
11. Automation control systems
12. Engineering industrial
13. Engineering civil

It is possible to see in the center part of figure 3, how the raw results of the left part (equivalent to figure 1 – query on journal and topic) are filtered. The quantities of articles are divided by an order of magnitude of 3. Logically, as the quantities are lower, the curve are less smooth and for more restrictive keywords instabilities are present, see kw3 for example. However, a yearly rate of increase can be computed, for less restrictive keywords (kw1 and kw2) around 8% and between 4% and 6% for kw3, kw4, and kw5. About the most restrictive keywords that interest us, kw6, no yearly increase can be computed, but there is no strong difference with or without WoS category filtering.

An interesting point is that the curves dealing with relations or dependencies between configuration variables (kw4 and kw5) are getting much closer to the ones dealing with model structuring and variable identification (kw3). This could lead to the conclusion that in our domain, works and publications are of the same order of magnitude regarding the three groups of keywords while the last one keeps remaining below.

B. Filtering with with abstract reading

Abstract reading has been done only for the paper relevant to the four strongest restrictive keywords kw3, kw4, kw5, and kw6 so in the lower part of figure 4. This means that around 600 abstracts have been quickly looked. The results are shown in the lower right part of figure 3.

The conclusions of the previous section are confirmed. Now with this filter, there is not any significant difference between the three keyword groups (kw3, kw4, and kw5) while the most restrictive keyword (kw6) remains below.

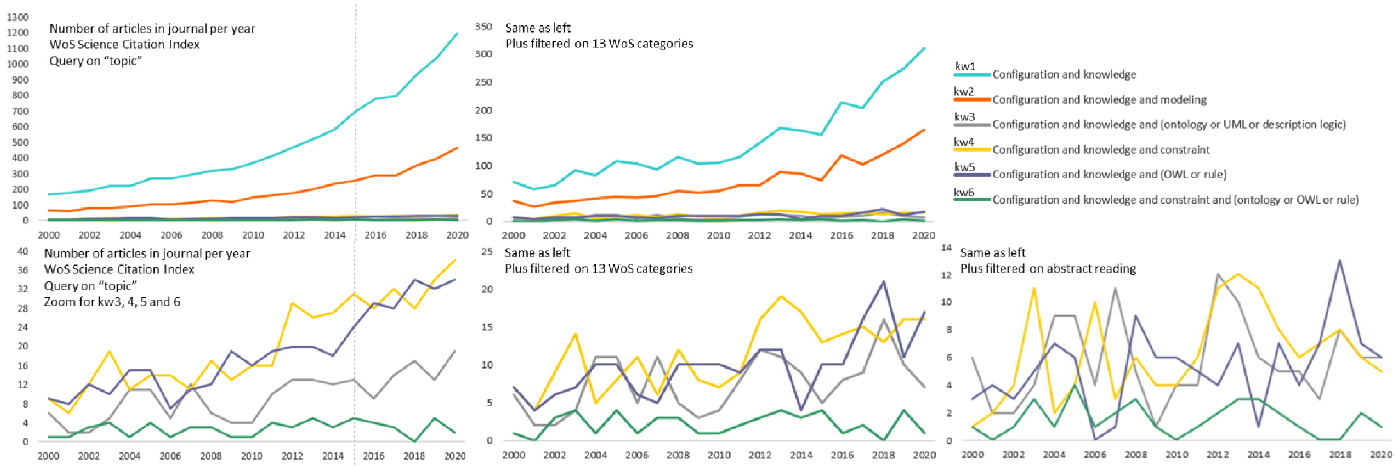

Figure 3. Filtering on WoS categories and abstract relevancy

C. Results discussion

Given the previous element, we have identified the following numbers of articles:

- Kw3: configuration and knowledge and (ontology or UML or description logic): 159 articles,
- Kw4: configuration and knowledge and constraint: 239 articles,
- Kw5: configuration and knowledge and (OWL or Rule): 207 articles,
- Kw6: configuration and knowledge and constraint and (ontology or OWL or Rule): 46 articles.

The initial tendency about the fact that researchers of the configuration domain rather prefer to work and publish on the subject dealing with "relations or dependencies between configuration variables" rather than those dealing with "model structuring and variable identification" is still lightly present but much less than without the filtering. The "constraint" keyword is the most frequent followed by "OWL or "rule".

Among the 46 articles selected with respect to "constraint" and "ontology or OWL or Rule" (kw6):

- 18 are also selected with respect to "ontology or UML or description logic " (kw3),
- 23 are also selected with respect to "OWL or Rule" (kw5),
- 4 are also selected by both (kw3 and kw5).

This shows that researchers naturally tend to work with a single model or formalism. It is indeed simpler, more reassuring, and more productive to work on a single model, tool, or approach. As a consequence, many works address separately the approaches based on constraints, ontologies, or rules, but very few combine two of them. As said in [1] about mixing approaches in what the author call hybrid configuration:" Typically, there is an ontology-like (DL-based) representation for representing components, their compositional relations, and attributes. Additionally, constraints are used for representing n-ary relations between components and for computing and inferring attribute values.".

Associating ontology and constraint approaches to develop configuration software still sounds of interest. Ontology brings the capture of the product concepts hierarchy description and descriptive variable identifications. Constraint diversity allows representing any kind of relation or dependencies that are the key elements of the configuration problem. Since the quite well-cited works of some authors as Dong [7] and [8], Asikainen [5] or Felfernig [4] and [6] were published more than twelve years ago, this domain definitely needs to be investigated again.

VI CONCLUSION

Our goal was to present a quantitative literature survey on the modeling aspect of configuration problems. Keywords around configuration, knowledge modeling, constraints, ontology have been discussed and selected. Web of Science has also been selected and results with different queries on (i) journal or proceeding, (ii) article topic or article title, have been presented.

A first conclusion is that authors seem to prefer to work on (i) constraints or rules than on (ii) ontology or product model structuring. A second one is that it is much harder to be highly specialized in more than one model or tool and consequently to publish on hybrid solutions associating the two previous work streams.

Knowledge modeling and maintenance is a key issue for configurator system deployment. For future studies, considering more efficient ways to model both product structure, product concepts, and configuration dependencies are still a necessity. Working on the association of ontology, object-oriented and constraint-based approaches seems to be very interesting which will generate with no doubt great improvements for configuration system design and deployment.

REFERENCES

[1] A. Felfernig, L. Hotz, C. Baglay, and J. Tihonen. Knowledge-based configuration From Research to Business Cases, Morgan Kaufmann, 2014.

[2] A. Myrodia, K. Kristjansdottir and L. Hvam. Impact of product configuration systems on product profitability and costing accuracy. Computers in Industry vol. 88, pp. 12-18, 2017.

[3] L. Hotz, A. Felfernig, M. Stumptner, A. Ryabokon, C. Bagley, and K. Wolter. "Configuration Knowledge Representation and Reasoning", Chapter 6 of Knowledge-based configuration From Research to Business Cases, Morgan Kaufmann, 2014.

[4] A. Felfernig, G. Friedrich, D. Jannach, M. Stumptner, M. Zanker. Configuration knowledge representations for Semantic Web applications. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, vol. 17 no. 1, 2003.

[5] T. Asikainen, T. Mannisto, T. Soininen, A. Kumbang. A domain ontology for modelling variability in software product families. Advanced Engineering Informatics vol. 21, no. 1, pp 23-40, 2007.

[6] A. Felfernig, Standardized configuration knowledge representations as technological foundation for mass customization. IEEE Transactions on Engineering Management, vol 54, no. 1, 2007.

[7] D. Yang, M. Dong and R. Miao. Development of a product configuration system with an ontology-based approach. Computer-Aided Design vol. 40, no. 8, pp 863-878, 2008

[8] D. Yang, R. Miao, H. Wu and Y. Zhou. Product configuration knowledge modeling using ontology web language. Expert Systems with Applications vol. 36, no. 3, pp 4399-4411, 2009.

[9] L. Zhang. Product configuration: a review of the state-of-the-art and future research. International Journal of Production Research vol. 52, no. 21 pp. 6381-6398, 2014

[10] G. V. Oddsson and K. R. Ladeby. From a literature review of product configuration definitions to a reference framework. Artificial Intelligence for Engineering Design Analysis and Manufacturing vol. 28 no.4 pp.413-428, 2014.

[11] http://ceur-ws.org/

[12] E. Garfield. "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas" Science, vol. 122, no. 3159, 1955, p. 108–111