

INCREMENT: A Mixed MDE-IR Approach for Regulatory Requirements Modeling and Analysis*

Nicolas Sannier and Benoit Baudry

Inria Rennes Bretagne Atlantique
Campus de Beaulieu
35042 Rennes Cedex, France
{nicolas.sannier,benoit.baudry}@inria.fr

Abstract. [Context and motivation] Regulatory requirements for Nuclear instrumentation and control (I&C) systems are first class requirements. They are written by national safety entities and are completed through a large documentation set of national recommendation guides and national/international standards. [Question/Problem] I&C systems important to safety must comply to all of these requirements. The global knowledge of this domain is scattered through these different documents and not formalized. Its organization and traceability relationships within this domain is mainly implicit. As a consequence, such long lasting nuclear I&C projects set important challenges in terms of tacit expertise capitalization and domain analysis. [Principal ideas/results] To tackle this domain formalization issue, we propose a dual Model-driven Engineering (MDE) and Information Retrieval (IR) approach to address the nuclear regulatory requirements domain definition, and assisted traceability based on the acquired requirements model. [Contributions] In this paper, we present the Connexion metamodel that provides a canvas for the definition and capitalization of the nuclear regulatory requirements domain. We also present an hybrid MDE/IR-based approach, named INCREMENT, for acquiring, modeling and analyzing these regulatory requirements. This approach is supported by a tool that is developed in the context of the CONNEXION project, which gathers French major nuclear I&C industrial actors.

Keywords: Nuclear Instrumentation and Control Systems, Regulatory Requirements, Standards, Metamodeling, Traceability, Information Retrieval.

1 Introduction

In addition to their systems requirements, systems with high level of security, privacy, or safety must also conform to regulatory requirements. For example, in the avionics domain, most regulators impose the application and compliance to

* This work is partially supported by the French BGLE Project CONNEXION.

the RTCA DO-178B/C. All healthcare related products in the USA must comply with the Health Insurance Portability and Accountability Act (HIPAA). Costs of noncompliance and incentives toward conformance are significant [1,2] and many initiatives, such as OPENCOSS [3], have emerged to tackle the regulatory requirements compliance issue from the safety certification perspective.

In the nuclear domain, regulatory requirements are completed using a large set of national recommendation guides and national/international standards. Putting these requirements in an international context showed important gaps between requirements and practices in different countries [4]. Since January 2011, the French nuclear industry and academic partners have joined forces in the CONNEXION project¹ to develop the major innovations in the design and implementation of the future nuclear power plants' Instrumentation and Control (I&C) systems. One aspect of the project consists in the formalization and the understanding, from a high level global perspective, of regulatory requirements the nuclear industry partners has to face in their licensing projects.

In this paper, we aim to address the following research questions. (1) How to formalize and organize the domain knowledge in a way that is relevant from an industrial experts perspective? (2) Once formalized, How to browse and manipulate this knowledge? (3) As the domain is large, not formalized, and hard to handle, can we analyze this domain and retrieve traceability links between regulatory requirements?

To tackle these questions, we propose a mix of Model-driven Engineering (MDE) and Information Retrieval (IR) to respectively address domain formalization and requirements traceability. The paper contributions are organized around the INCREMENT approach (Instrumentation aNd Control Regulatory Requirement Modeling Environment) that respectively addresses the challenges previously introduced. In particular, they consist in: **(1) the domain formalization** by proposing a metamodel that allows a high level capitalization of a requirements corpus and its organization. This metamodel was built through intensive interactions with our industrial partners. **(2) A tool-support basis** to gather partial knowledge from the textual documents, and manipulate such models. This tool basis is evaluated with both empirical and industrial feedback. **(3) The proposal of an original hybrid approach, mixing both meta-modeling and information retrieval** to support better domain analysis and that has been empirically evaluated.

The remainder of the paper is organized as follows. Section 2 presents the I&C Regulatory requirements global picture as well as an illustrative example. In section 3, we present the metamodel that supports the INCREMENT approach. Section 4 presents the environment we built on top of the metamodel while section 5 discusses the hybridization MDE-IR in our approach and the empirical evaluation of its benefits. In section 6, we expose threats to validity of our work. Section 7 reviews related work while section 8 concludes the paper.

¹ <http://www.cluster-connexion.fr>

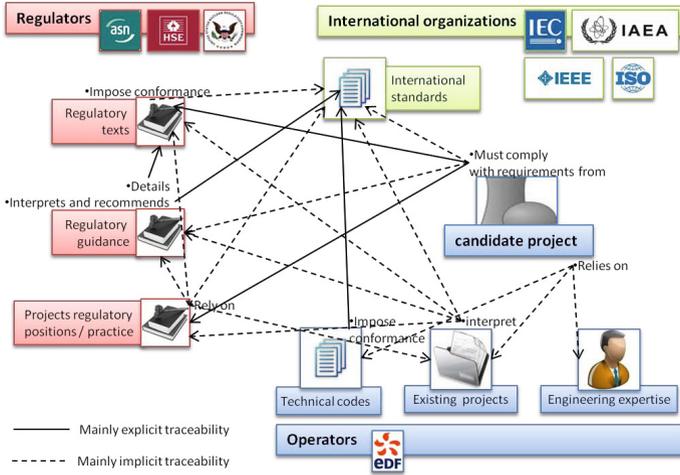


Fig. 1. Global picture of the Nuclear Regulatory Requirements Domain

Requirements for Software in Category 1E Programmed Systems
Reliability
 Reliability is addressed within qualitative perspectives
Ea 2.1 Software design and documentation shall allow performing verification and validation methods in order to demonstrate ... An acceptable practice, related to methods and techniques of verification, is described in chapters 6 (verification) and 7 (software/component integration) of the IEC 60880 publication (1986) ...
 Similarly, simulation is an acceptable technique for the validation of the executable program, especially for time performances. This technique can be combined with prescriptions of chapter 8 of IEC 60880 publication (1986).

Fig. 2. V&V in French regulatory text RFS II.4.1.a

2 Analyzing Nuclear Regulatory Requirements in the Large: An Example

Figure 1 proposes a global picture of the nuclear regulatory landscape in terms of concepts and traceability concerns. To illustrate the heterogeneity of the domain (different stakeholders, different perspectives, different levels of details), we propose ask a simple question and browse the corpus to find out elements related to it. More details are proposed in our previous work [5], and we recall it for the sake of clarity regarding the paper contribution.

Considering specific analysis such as finding V&V regulatory requirements in safety systems for different countries, one should initially think that these requirements are close enough to be compared. We propose an example of what nuclear operators have to face from the regulatory text perspective, and refine it to the normative level in two different contexts: France and USA.

At the Regulatory Level. In France, in the RFS (basic safety rule) II.4.1.a (2000), the requirements or principles are written in French. About the concern Verification and Validation, Figure 2 proposes a translation. In the USA, we shall consider the 10CFR50 and in particular following excerpt in Figure 3.

<p>Par55a(a)(1): Codes and Standards (a) Quality standards, ASME Codes and IEEE standards, and alternatives. (1) Structures, systems, and components must be designed, fabricated, erected, constructed, tested, and inspected to quality standards commensurate with the importance of the safety function to be performed. ... (h) Protection and safety systems. (2) Protection systems. For nuclear power plants ... must meet the requirements stated in either IEEE Std. 279 ... or in IEEE Std. 603-1991 ... (3) Safety systems. Applications ... must meet the requirements for safety systems in IEEE Std. 603-1991 and the correction sheet dated January 30, 1995.</p> <p>Appendix A to Part 50—General Design Criteria for Nuclear Power Plants I. Overall Requirements Criterion 1- Quality standards and records. Structures, systems, and components important to safety shall be designed, fabricated, erected, and tested to quality standards commensurate with the importance of the safety functions to be performed. ...</p>

Fig. 3. US Regulation: 10CFR50 and Appendix A

At this level, we can agree that there are mainly common points regarding verification and validation even if it is not mentioned in the US regulation (apart from the word "tested". In France, independent V&V is already explicit. Fitness to specification (validation) is present. Both of them mention quality assurance programs. The notion of compliance with standards is expressed everywhere with more or less importance. Software safety life cycle is approached using different terms or enumeration of activities in the US, fitness to specification, V&V methods in France). We also observe the emergence of different level of application of standards as acceptable approaches (FR, USA), best in process and applicability (USA) and mandatory items (USA).

At the Regulatory Guidance Level. There is no document at this level in France. Nevertheless, the RFS explicitly mention that use of Chapter 6, 7 and 8 of the IEC60880 (1986) are acceptable practices for software V&V of category 1E systems. The French safety authority has endorsed the RCC (Rules for Design and Construction) series issued by EDF (considered as a technical operator code in Figure 1). In particular, RCC-E (for electrical devices) requires conformance with several international standards such as IEC60880, IEC62138, etc. depending on the safety function category performed by the software. In the US, it is described partially into the regulatory guide 1.168 that will later lead us to the analysis of the IEEE standard 1012.

At the Normative Level. The next step finally leaves us with two documents from the IEC and IEEE community. If both IEC60880 and IEEE1012 deal with software validation and verification, the chosen perspective of description is rather different.

IEC 60880 (chapter 8) deals with: 1. independence of the verification; 2. verification plan; 3. design verification; 4. implementation verification (with both general purpose and application-oriented languages and respective test reports); 5. configuration of pre-developed software. IEEE 1012 deals with: 1. software V&V processes: management, acquisition, supply, development, operation, maintenance; 2. software V&V reporting, administration and documentation; 3. detailing a software V&V plan outline.

If we want to sum the two standards, IEC 60880 expresses objectives to reach whereas IEEE 1012 details activities to perform to reach these objectives.

More generally, there is a gap between the IEC corpus, which is specifically written by the IEC subcommittee SC45-A and that issues nuclear specific to nuclear industry and IEEE standards which are not always nuclear specific.

This example describes two different regulatory practices with their own particularities and the issue is to formalize this domain if we want to be able to compare them [6]. We address in next section our first research question, concerning the domain formalization issue.

3 Formalizing the Nuclear Regulatory Requirements Domain

3.1 Toward a Domain Specific Modeling Approach in the Industry

One major issue when working with industrial partners is their level of adoption of MDE or, at least, modeling concerns [7]. In the CONNEXION project, we face research and development engineers with very heterogeneous background, from senior project leaders with mainly low expertise on modeling to junior and senior engineers with or without knowledge in MDE neither in requirements engineering (or from the Systems Engineering point of view).

The key question here is to propose a modeling approach that meet our partners' intuition of how the domain should be represented and that is close to their current practice. As a consequence, we decided to go through a classic MDE-based approach and the creation of a domain specific metamodel.

3.2 The Connexion Metamodel

In figure 4, we propose an excerpt of the Connexion metamodel we built with our industrial partners in the CONNEXION project. This metamodel² structures the different kind of elements one may find while looking at the nuclear regulations.

1. **Modeling Regulations Atomic Elements.** Though our industrial partners mainly focus on requirements, they also want to put these requirements in context and keep the document structural information. The metamodel does not only focus on requirements and its different specializations but also on side elements such as definitions, recommendations, descriptive texts, etc. This whole set of typed elements (*TypedElement*), contained into a *TypedElementCorpus* are acquired through the documentation or may be tacit knowledge (*NonWrittenElement*) acquired from past or existing projects.
2. **Modeling the Regulatory Hierarchy.** We defined the different types of documents and their structure that we handle in the project. It goes from the different regulatory documents to standards as well as documents from the licensee (engineering documents, technical codes, etc.). These documents are modeled as a *Corpus* of refinable *Documents* and possess a composite structure of refinable *DocumentFragments*.

² A more detailed version as well as specific perspectives are available at <http://wp.me/P1tUd5-6I>

3. **Modeling Clustering of Elements.** Built as-is, the proposed requirements referential includes an important amount of information that must be organized. To analyze this set of elements, Different kinds of wrappers are required. These *TypedElementWrappers* define: (1) structural similarities that clusters *TypedElements* regarding their nature (regulatory, normative, engineering), but also (2) thematic similarities within a *Topic* collection, (3) large general elements that are specific to a *Project* or a more general *GenericProject*.
4. **Modeling the Bridge between Requirements, Architecture and Qualification.** As high level ambiguous requirements [8,9], it is very difficult to cope with the traditional set <Requirement, Architecture, Qualification>. The architecture part is addressed separately in the CONNECTION project with a specific metamodel. However, we link the architecture elements through the satisfaction of design rules, that are industry-based clauses. It is the same process for qualification and certification concerns while defining a justification. As a consequence, we have defined high level *DesignRules* that allow an indirect validation of the related requirements as well as *Justifications* to address the safety evidence process. These concerns are close to the actual OMG proposal around the SAEM/SACM standard metamodels [10].
5. **Modeling Elements Interactions.** Carlhamre et al. [11] defined a set of interdependencies in order to address requirements prioritization and planning. However, regulatory requirements are more abstract, more complex and cannot be seen in terms of marketable or temporal priority.

To tackle the traceability concern we highlighted in figure 1, we propose a set of traceability links where comparison links are made to define equivalence between or conflicting elements. On the other hand, the interaction links are made to describe relationships within the requirements domain. We have defined two families of *Interactions*: *ElementsComparisons* that describe equivalence and conflict links, as well as *ElementInteractions* that describes inter-requirements relationships such as basic *References* or more evaluated *Generalization*, *Contribution* relationships. These relationships are complementary to those defined by Zhang et al[12] or Maxwell et al.[13].

For the nuclear industry, which owns a very precise vocabulary, determining the correct metamodel often depends on the terminology of terms and concepts. The metamodel fitness is built among a long iterative process. To the best of our knowledge, the process of building a domain specific modeling language in the industry has not much been assessed in terms of activity length or number of iterations. We mainly spent two years of interviews and meetings to built and provide examples for this metamodel with various minor (new attributes, renaming attributes) and major changes (brand new concepts, major concept shifts). For the major changes, we had three different versions of the metamodel.

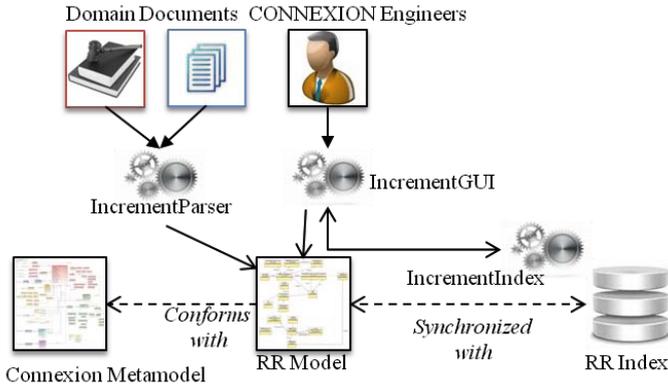


Fig. 5. The INCREMENT Contribution

If we analyze the different factors for our major changes, the following aspects had some impact on the metamodel life cycle:

- Concepts arise or change from the natural, mutual, and iterative process of domain understanding and elicitation.
- Being proactive and proposing concepts or providing examples may help to make domain experts describe tacit knowledge [14]. Clustering requirements within topics is a natural practice as the number of requirements grows up but was not determined explicitly.
- Academic and industrial definitions or visions upon particular concepts can co-exist but may lead to the final choice or definition modification. We observed a significant granularity gap between our visions of topic/theme definitions. We defined topics or themes as a set $\langle \text{topic name, topic signature, tracks} \rangle$, which is close to the Gotel proposition [15] instead of an industrial hierarchical and composite structure.
- Basic state of the art concepts and structures may not ensure their adoption as experts are very demanding concerning their own domain lexicon. They may question academic approaches though concepts are similar but not expressed in the "correct" way.
- Experts are also involved in different projects where they acquire new perspectives or ideas related to the CONNEXION project.

4 Breathing Life into a Regulatory Requirements Model

Providing a domain metamodel is one first thing. Making this metamodel manageable for nuclear engineers and ensure its adoption, with a concrete representation, is one another, and much more difficult, point. For the nuclear industry, most of our partners do read and interpret UML diagrams. However, it's hard

for them understand (meta)modeling. In particular, our partners wrongly associate metamodels, models that conform to them and tooling that manipulate and exploit models.

This metamodel allows to formalize the I&C regulatory requirements domain. However, CONNEXION engineers require additional features in order to provide ways to populate and analyze models that are the concrete domain representations and knowledge capitalization. To this end, we have proposed different software solutions that leverage the metamodel and, in addition to the Connexion metamodel, form the global INCREMENT approach (Fig. 5). In particular, we propose:

- *IncrementParser* as a configurable parser, that allow us to extract information from the regulation and model them as Connexion model elements.
- *IncrementGUI* is a graphical environment for the model browsing and analysis and is presented in Fig. 6.
- *IncrementIndex* is a model-based indexing and searching engine. It leverage the metamodel information to propose an indexing step based on the model elements. We address this concern in the next section.

A Parser for Systematic Requirements Model Acquisition. Despite the variety of documents, regulation can be organized with respect to reading rules, which allow the readers to have a systematic and efficient reading of the document. Some rules may be explicit, written in the documents, or implicit and provided by domain experts such as keywords. It is worth noticing that these reading rules are specific to each document and may evolve from one another.

It is possible to automate the extraction of textual information and generate an instance of the metamodel, leveraging the reading rules. To perform the extraction task, we have developed a configurable parser (*IncrementParser*) that uses, for each document, a set of regular expressions that defines the parsing rules to determine the different fragments types while reading the input file. All clauses of standards are then typed and generated as model elements.

This tool was evaluated on the acquisition of 8 international nuclear standards, and validated by sampling. We provide some details in Table 1. 1. software sys-

Table 1. Details from the 8 Acquired International Standards

Standard	1st year of publication	# of pages	Structure	Reqs.	Recoms.	Defs.	Ind. documents
IEC60880-2006	1986	110	15 sections and 10 normative or informative annexes	308	92	43	939
IEC60987-2007	1989	30	13 sections and 3 informative annexes	53	17	18	219
IEC61226-2009	2009	32	7 sections and 1 informative annex	67	12	22	261
IEC61500-2009	1996	14	10 sections	43	10	8	136
IEC61513-2011	2001	98	8 sections and 5 informative annexes	238	48	62	1098
IEC62138-2004	2004	47	6 sections	180	48	36	555
IEC62340-2007	2007	22	9 sections + 1 informative annex	46	4	26	226
IEC62566-2011	2011	52	17 sections + 2 informative annexes	243	33	14	646
totals	405	107 1st level structures	1178 structural elements	264	94	229	4080

tems performing category A functions (IEC 60880), B and C functions (IEC 62138); 2. hardware design requirements (IEC 60987); 3. classification of safety functions (IEC61226); 4. data communication (IEC 61500); 5. general criteria for I&C systems (IEC 61513); 6. common cause failure (IEC 62340); and 7. development of HDL-programmed Integrated Circuits (IEC62566).

These 8 standards cover a large scope from very general concerns (IEC 61513, IEC 60880, IEC 62138) to very precise ones (IEC 62340, IEC 62566). Publication dates vary from 2001 to 2011 but first publication record start in 1986. These 8 standards illustrate the diversity of the documentation in terms of temporal evolution, scope heterogeneity, amount of statements, etc.

An Environment for the Corpus Browsing and Analysis. *IncrementGUI*, illustrated in Fig. 6, proposes a domain description over three dimensions. The left part proposes to navigate the model through the different wrappers: Type-Elements, semantic wrappers, topics, projects, interactions, design rules, and justifications. The center part details the content of these wrappers and focus on the elements types and verbatims. The right part proposes a detailed and navigable view of the selected element through its attributes and its references.

We originally proposed a first prototype to represent such requirements models with a diagram perspective, which exhibits interactions, similarly to the the visualization proposed by Carlhamre et al. [11]. However, this perspective was not adopted as very far from their common vision of requirements and what they

The screenshot displays the IncrementGUI interface, which is organized into three main vertical panels:

- Categories (Left Panel):** A tree view titled "TypedElementCorpus" showing a hierarchy of elements. The "EngineeringElementsCorpus" is expanded, showing sub-categories like "corpus" (with IAEA NS-R-1, IAEA SS-R-2, IEC 60709, IEC 60964, IEC 60965, IEC 61226) and "projects" (with FLA3, HPC). Other categories like "Redundancy", "Diversity & Common Cause Fa...", "Data communication", "Communication inside one zor", "Embedded security features", "Independence", "SA5", "TotalCoverage", and "partialCoverage" are also visible.
- Content (Middle Panel):** A list of "StandardRequirement" elements. The selected item is "Requirement 22: Safety classification All items important to safety shall be identified and shall be classified on the basis of their function and their safety significance." Other requirements listed include "Requirement 27: Support service systems Support service systems that ensure the operability of equipment forming part of a system important to safety shall be classified accordingly.", "The design basis for information systems, including their measurement devices, shall take into account their importance to safety. The intended safety function of each system and its importance in enabling the operators to take proper pertinent actions in anticipated operational occurrences or accident conditions shall be identified in its design basis and shall be used as an input to any I&C categorization method selected.", "The reliability requirements and failure criteria should be identified from the safety categorization process of IEC 61226 and from plant safety requirements, emerged from, among others, regulatory bodies. A failure of an information system means that the information is degraded and not sufficient and precise enough to understand or perform a safety task properly. A single failure within a system is any failure of a component, e.g., a sensor, a processor or a display unit.", "ECS.1 : Safety categorisation : The safety functions to be delivered within the facility, both during normal operation and in the event of a fault or accident, should be categorised based on their significance with regard to safety.", and "ECS.2 : Safety classification of structures, systems".
- Attributes (Right Panel):** A detailed view of the selected requirement. It shows:
 - id:** 3
 - name:** AP3
 - verbatim:** En Grande Bretagne, les co...
 - classification:** -
 - inputAuthor:** AREVA
 - date:** sept.-13
 - rationale:** On veut protéger les systèm...
 - description:** limiter communication ent...
 - category:** null
 - applicationContext:** general design
 - systemTarget:** all
 - status:** EXPERTADDITION
 - informationNote:** Pratique acceptes issu
 - flexibility:** 0
 - maturity:** 0
 - criticality:** 0
 - scope:** (dropdown menu)
- References (Bottom Right Panel):** A section titled "References" showing a list of related elements: interactions, inProjects, justifiedin, parent, inTopics, corpus, totallyCoveredBy, partiallyCoveredBy, fromProject, and Independence. A "Property details" button is also present.

Fig. 6. The Increment environment

would like to manipulate. In particular, they wanted to manipulate the textual documents they are used to read and analyze. The second proposition is based on the IncrementGUI interface (Fig. 6), and is based on a requirements library metaphor with a strong focus put on elements verbatim.

IncrementGUI layout had a "sufficient correctness", said differently, the prototype was close enough of the industrial intuition of what the domain concepts and the supporting tool should be. The prototype is currently under evaluation regarding navigation and the basic CRUD (create, read, update, delete) features.

Though we had a domain formalization through the metamodel, acquired from the documentation and domain experts, we also wanted to perform analyses on this domain. In particular, we wanted to be able to search within our documentation and ease the definition of topics or retrieve requirements. Due to the textual nature of our model elements, these features were not possible at the modeling level. However, information retrieval techniques for requirements traceability offer such analysis capability. In next section, we propose to describe our hybridization of MDE and IR for requirements analysis and traceability concerns.

5 Hybridizing MDE and IR in the Tool

In a previous work [16], we presented the challenges to keep both a model and a index synchronized in order to have a consistent use of information retrieval on such requirements models. Basically, the main concept of indexing engines is based around the Document, and its fields. Fields are textual entries that may describe any property of the document, its verbatim among others, but without any further semantics such as typing, structure, references, etc.

Fig. 7 describes the mapping we operated to perform the hybridization. The Clusters of the Metamodel can be used as different indexes. Model element types are stored as a field as well as `TypedElement` and `DocumentModelElement` attributes. By the way, instead of slicing flat standard documents, we use the `Connexion` model as root for a richer indexation, provided the elements attributes.

One major drawback of Information retrieval approaches, in particular, TF-IDF similarity scoring, is the huge amount of false positives candidate links that are generated [17]. In the literature, this is handled through the use of an arbitrary cut-off value upon the document's score below which, the document is not considered as a valuable candidate link [18].

Fields inherited from the model, in particular type and classification, can be used efficiently to filter the candidate link generation and remove inconsistent document while searching this large amount of documents. In particular, our metamodel defines not only requirements but a substantial set of different concepts that are also indexed as they may provide contextual information upon their neighbor requirements. With large requirement index such as we have, removing inconsistent information from the expert further operation is crucial as it prevents the expert from rejecting inconsistent (by construction) elements.

Table 2 presents the results of our experiments regarding the search space reduction while leveraging the model information against a standard approach

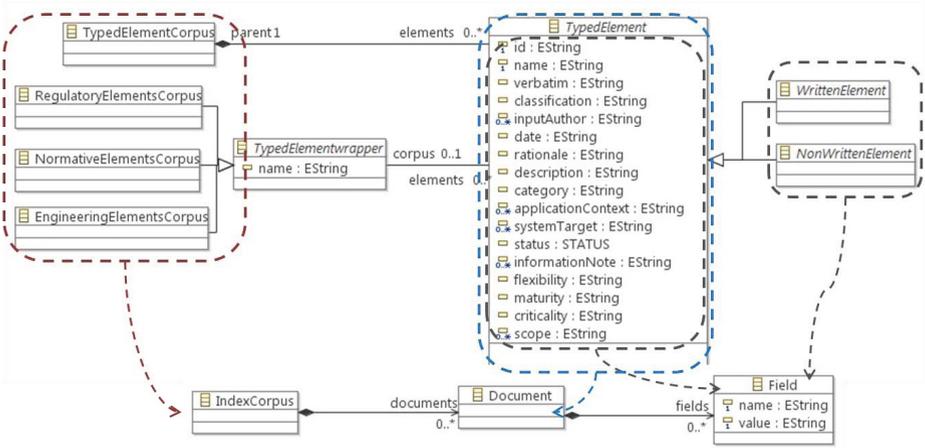


Fig. 7. Hybridizing MDE and IR

with an empirically defined cut-off value (0.2). It proposes a comparison between a « standard » information retrieval TF-IDF similarity scoring based on a simple index (set of flat documents, without any information but the documents text) (left half) and the same scoring based on our model-based index with richer information provided by the model (right half). In particular, the table provides, respectively, the total number of retrieved documents, the number of candidate links (documents whose score are above the cut-off value), the number of retrieved requirements and recommendations, and the proportion on "noise" (documents that are not requirements or recommendations). The right half proposes the number of retrieved documents and the number of requirements and recommendations using our model-based index. Finally we propose an evaluation of our approach concerning the search space reduction against the standard approach.

The cut-off value has a rather low impact on the "noise" removal. Worse, it also cuts possible consistent elements, from a typing point of view. On the other side, the model-based index does not remove these elements and also remove the related noise (which are type inconsistent) and further reduced the research space, compared to the standard approach, at an average 65% more. This tends to show that our model-based pruning heuristic performs much better than an arbitrary cutoff-value in our particular context.

6 Threat to Validity

External Threats. Our study is based on an empirical analysis and informal industrial feedbacks on our work. However, we have not evaluated the consistency and the adoption of the tool at a larger scale. We plan to extend this evaluation

Table 2. Leveraging the Model’s Information to Reduce the Number of Candidate Links

query	Standard Index (with a cutoff value)				Model-based Index		
	links	candidates	Rqts/Rcms	% noise	links	Rqts/Rcms	reduction
config. mngt.	438	221	72	67,42	106	106	52,04%
cmon. cause failure	602	154	17	88,96	115	115	25,32%
specification	668	576	216	62,5	216	216	62,5%
independence	102	64	14	78,13	14	14	78,13%
validation	404	347	96	72,33	96	96	72,33%
verification	555	445	169	62,02	171	171	61,57%
quality assurance	421	259	84	67,57	106	106	59,07%
defence in depth	141	81	8	90,12	14	14	82,72%
integration	280	237	65	72,57	65	65	72,57%
self supervision	125	92	25	72,83	25	25	72,83%
modification	271	214	70	67,29	70	70	67,29%
diversity	114	103	16	84,47	16	16	84,47%

with more formal and quantitative measures in a more advanced dissemination phase.

Internal Threats. We progressively defined the metamodel according to our interviews with experts and analysis of the domain. As our partners have no experience in metamodeling, we have to manage this task and further empirically validate it with them.

Construction Threats. The metamodel is still evolving. Factors of change we identified can be biased with our partners’ MDE practice, though many factors are known from the RE community. The metamodel evolution is more related to adding new concepts or refining and organizing concepts. We have not observed inconsistent concepts or associations. Our partners are highly experiences I&C experts and they do know what concepts have to be defined and then refined.

7 Related Work

On Using Models and Indexes. To the best of our knowledge, there is no existing work that proposes requirements traceability using information retrieval that is based on a domain metamodel and its instances. Recently, Dumitru et al. [19] or Tung et al. [20] mined Softpedia products information to propose enhanced recommending systems. These approaches mined static contents that do not require further synchronization accordingly to a more changing and dynamic model. Moreover, these approach did not address traceability but recommendations and were not concerned with the search space reduction issue but with small and precise information.

On MDE Approaches for Requirements Modeling and Analysis. Apart of the popular goal-oriented approaches such as KAOS [21], i* [22] that are

specification oriented, or URN [23] that has been recently standardized, more domain specific modeling languages have been proposed. Some are based on UML/SysML profiling. Panesar et al., in CRESCO, proposed a UML profil for the certification task and was specific to the IEC 61508 standard [24]. In a similar approach, Zoughbi et al. proposed a UML profile that was specific to the requirements to code traceability within the DO-178B perspective [25]. In a different context, de la Vara and Panesar proposed the metamodel SafetyMet [3] and aimed to fit a more general purpose but was activity-driven. Helming et al. [26,27] had different concern and dealt with requirements models version management with EMFStore in Unicase. All these work did not take into account the diversity of concepts and traceability issues that are contained in regulations. They did not also embrace the global domain in the large, but focus, at most, one specific standard, or very specific requirements analyses.

On Regulatory Requirements Analysis. About regulatory requirements and compliance concerns, extensive studies had been done in healthcare domain and, particularly around HIPAA. In [28], production rules are developed to translate regulatory texts and formalize forms of legal knowledge and ambiguity. In [29], the authors derive rights and obligations from HIPAA and compare different stakeholders' interpretations. In [30], specific legal statements from multiple jurisdictions are refined using a requirement specification language. Statements are then neighbored and similar ones are organized to identify gaps, conflicts and try to reconcile them. In [13], the authors focus on explicit external cross-reference links and propose a legal cross-reference taxonomy. In [31], the authors use User Requirements Notation (URN), a combination of NFR and i* frameworks and use-case maps, to model both the regulation and a hospital business process. All these works, however, did not consider regulatory requirements in the large but focused, in the small, very specific aspects of a regulations such as privacy or data breaches analyses.

On Requirements Traceability and Information Retrieval. Natural language processing (NLP) and information retrieval approaches have been previously used for Requirements Analysis. At the system's scale, it has been pioneered by Sawyer et al. [32] within the REVERE project and distinguish between requirements types. Kiyavitskaya et al. [33] use GaiusT to extract rights, obligations, on both HIPAA and equivalent Italian regulations. It relies on text decomposition in a parse tree. Cleland et al. [34,35,36] use NLP and probabilistic techniques to trace regulatory requirements from HIPAA in several software applications. Leuser and Ott [37] also wanted to tackle requirements traceability in large specifications in the automotive domain at Daimler, but leverage specifications in controlled natural language and already formalized domain knowledge. Tackling the candidate link generation is a major issue in the IR community. Niu and Mahmoud proposed to rely on clustering algorithm to sort between good and bad quality clusters [17]. Our approach is based on a pre-processing enrichment of the documents, synchronized with the model information and can be seen as a complementary work.

8 Conclusion

In this paper, we addressed the question of formalizing the regulatory requirements for the nuclear domain. In this domain, nuclear I&C engineer face a large amount of regulatory and normative requirements as well as tacit practices. All these requirements express multiple different concerns, scatter and hinder the domain knowledge capitalization.

In the context of the CONNEXION project, we iteratively defined a metamodel that defines the different domain concepts as well as its organization. We proposed an Hybrid MDE/IR approach and a tool to assist engineers in the quest for the domain navigation, manipulation, and analysis. We evaluated the hybridization of MDE and IR in terms of non arbitrary candidate link search space management. For the particular nuclear I&C domain, we have shown an average 65% reduction of this search space, without having to rely on a cut-off value.

As the CONNEXION goes on, our current work is on an improvement of the definition in the metamodel of tacit (non written) requirements and practices. We also plan to address the requirements variability inside such requirements Metamodel. In particular, we want to address the meaning of regulatory requirements variability, find the good variability formalism and evaluate the impact of requirements variability in terms of design rules validity and architecture design.

References

1. Otto, P.N., Antón, A.I., Baumer, D.L.: The choicepoint dilemma: How data brokers should handle the privacy of personal information. *IEEE Security & Privacy* 5(5), 15–23 (2007)
2. Maxwell, J.C., Antón, A.I., Swire, P.: Managing changing compliance requirements by predicting regulatory evolution. In: *RE 2012*, pp. 101–110 (2012)
3. de la Vara, J.L., Panesar-Walawege, R.K.: Safetymet: A metamodel for safety standards. In: Moreira, A., Schätz, B., Gray, J., Vallecillo, A., Clarke, P. (eds.) *MODELS 2013*. LNCS, vol. 8107, pp. 69–86. Springer, Heidelberg (2013)
4. RHWG, W.R.H.W.G.: Harmonisation of reactor safety in wentra countries. Technical report, WENRA (2006)
5. Sannier, N., Baudry, B.: Defining and retrieving themes in nuclear regulations. In: *RELAW 2012*, pp. 33–41. IEEE (2012)
6. Johnson, G.: Comparison of iec and ieee standards for computer-based control systems important to safety. In: *IEEE Nuclear Science Symposium Conference Record*, vol. 4, pp. 2474–2481. IEEE (2001)
7. Hutchinson, J., Rouncefield, M., Whittle, J.: Model-driven engineering practices in industry. In: *ICSE 2011*, pp. 633–642 (2011)
8. Kamsties, E.: Understanding ambiguity in requirements engineering. In: *Engineering and Managing Software Requirements*, pp. 245–266. Springer (2005)
9. Breaux, T.D., Antón, A.I.: Analyzing regulatory rules for privacy and security requirements. *IEEE Trans. Software Eng.* 34(1), 5–20 (2008)
10. OMG: Documents associated with software assurance evidence metamodel (saem) version 1.0 - beta 1, <http://www.omg.org/spec/SAEM/1.0/Beta1/> (2010)

11. Carlshamre, P., Sandahl, K., Lindvall, M., Regnell, B.: och Dag, J.N.: An industrial survey of requirements interdependencies in software product release planning. In: RE 2001, pp. 84–93 (2001)
12. Zhang, W., Mei, H., Zhao, H.: A feature-oriented approach to modeling requirements dependencies. In: RE 2005, pp. 273–284 (2005)
13. Maxwell, J.C., Antón, A.I., Swire, P.: A legal cross-references taxonomy for identifying conflicting software requirements. In: RE 2011, pp. 197–206. IEEE (2011)
14. Sawyer, P., Gervasi, V., Nuseibeh, B.: Unknown knowns: Tacit knowledge in requirements engineering. In: RE 2011, p. 329 (2011)
15. Gotel, O., Morris, S.J.: Out of the labyrinth: Leveraging other disciplines for requirements traceability. In: RE 2011, pp. 121–130 (2011)
16. Sannier, N., Baudry, B.: Toward multilevel textual requirements traceability using model-driven engineering and information retrieval. In: MoDRE 2012, pp. 29–38 (2012)
17. Niu, N., Mahmoud, A.: Enhancing candidate link generation for requirements tracing: the cluster hypothesis revisited. In: RE 2012, pp. 81–90. IEEE (2012)
18. Chen, X., Grundy, J.: Improving automated documentation to code traceability by combining retrieval techniques. In: ASE 2011, pp. 223–232. IEEE Computer Society (2011)
19. Dumitru, H., Gibiec, M., Hariri, N., Cleland-Huang, J., Mobasher, B., Castro-Herrera, C., Mirakhorli, M.: On-demand feature recommendations derived from mining public product descriptions. In: Proceedings of the 33rd International Conference on Software Engineering, ICSE 2011, pp. 181–190. ACM, New York (2011)
20. Thung, F., Wang, S., Lo, D., Lawall, J.: Automatic recommendation of api methods from feature requests. In: ASE 2013, pp. 290–300 (2013)
21. van Lamsweerde, A.: Requirements Engineering - From System Goals to UML Models to Software Specifications. Wiley (2009)
22. Yu, E.S.: Towards modelling and reasoning support for early-phase requirements engineering. In: Requirements Engineering, RE 1997, pp. 226–235. IEEE (1997)
23. Amyot, D., Mussbacher, G.: User requirements notation: The first ten years, the next ten years (invited paper). *Journal of Software (JSW)* 6(5), 747–768 (2011)
24. Panesar-Walawege, R.K., Sabetzadeh, M., Briand, L.C.: A model-driven engineering approach to support the verification of compliance to safety standards. In: ISSRE 2011, pp. 30–39 (2011)
25. Zoughbi, G., Briand, L.C., Labiche, Y.: Modeling safety and airworthiness (rtca do-178b) information: conceptual model and uml profile. *SOSYM* 10(3), 337–367 (2011)
26. Li, Y., Narayan, N., Helming, J., Koegel, M.: A domain specific requirements model for scientific computing. In: ICSE 2011, pp. 848–851 (2011)
27. Helming, J., Koegel, M.: Managing iterations with unicas. In: ICSE 2010, pp. 313–314 (2010)
28. Maxwell, J.C., Antón, A.I.: Developing production rule models to aid in acquiring requirements from legal texts. In: RE 2009, pp. 101–110 (2009)
29. Breaux, T.D., Antón, A.I., Doyle, J.: Semantic parameterization: A process for modeling domain descriptions. *TOSEM* 18(2) (2008)
30. Gordon, D.G., Breaux, T.D.: Reconciling multi-jurisdictional legal requirements: A case study in requirements water marking. In: RE 2012, pp. 91–100 (2012)
31. Ghanavati, S., Amyot, D., Peyton, L.: Towards a framework for tracking legal compliance in healthcare. In: Krogstie, J., Opdahl, A.L., Sindre, G. (eds.) CAISE 2007. LNCS, vol. 4495, pp. 218–232. Springer, Heidelberg (2007)

32. Sawyer, P., Rayson, P., Garside, R.: Revere: Support for requirements synthesis from documents. *Information Systems Frontiers* 4(3), 343–353 (2002)
33. Kiyavitskaya, N., Zeni, N., Breaux, T.D., Antón, A.I., Cordy, J.R., Mich, L., Mylopoulos, J.: Automating the extraction of rights and obligations for regulatory compliance. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) *ER 2008*. LNCS, vol. 5231, pp. 154–168. Springer, Heidelberg (2008)
34. Cleland-Huang, J., Czauderna, A., Gibiec, M., Emenecker, J.: A machine learning approach for tracing regulatory codes to product specific requirements. In: *ICSE 2010*, pp. 155–164 (2010)
35. Mirakhorli, M., Shin, Y., Cleland-Huang, J., Çinar, M.: A tactic-centric approach for automating traceability of quality concerns. In: *ICSE 2012*, pp. 639–649 (2012)
36. Cleland-Huang, J., Heimdahl, M., Huffman Hayes, J., Lutz, R., Maeder, P.: Trace queries for safety requirements in high assurance systems. In: Regnell, B., Damian, D. (eds.) *REFSQ 2011*. LNCS, vol. 7195, pp. 179–193. Springer, Heidelberg (2012)
37. Leuser, J., Ott, D.: Tackling semi-automatic trace recovery for large specifications. In: Wieringa, R., Persson, A. (eds.) *REFSQ 2010*. LNCS, vol. 6182, pp. 203–217. Springer, Heidelberg (2010)