



HAL
open science

Requirements Engineering enabled by Natural Language Processing and Artificial Intelligence for Nuclear Safety Demonstration

Emir Roumili, Jean-François Bossu, Vincent Chapurlat, Nicolas Daclin,
Alekssei Iancheruk, Robert Plana, Jérôme Tixier

► To cite this version:

Emir Roumili, Jean-François Bossu, Vincent Chapurlat, Nicolas Daclin, Alekssei Iancheruk, et al.. Requirements Engineering enabled by Natural Language Processing and Artificial Intelligence for Nuclear Safety Demonstration. CSDM 2020 - Complex System Design and Management 2020, Dec 2020, Paris, France. hal-02974291

HAL Id: hal-02974291

<https://imt-mines-ales.hal.science/hal-02974291v1>

Submitted on 12 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Requirements Engineering enabled by Natural Language Processing and Artificial Intelligence for Nuclear Safety Demonstration.

Emir ROUMILI, Jean-François BOSSU, Vincent CHAPURLAT, Nicolas DACLIN, Aleksei IANCHERUK, Robert PLANA, Jérôme TIXIER

Abstract This article describes a new method to elaborate and conduct the safety demonstration phase for nuclear infrastructure in phase with System Engineering principles. More particularly, this method focuses on helping and guiding engineers involved in nuclear safety demonstration into a better assessment considering the increasing complexity and the expectations from the licensing owner who must accommodate for any new nuclear project. This article claims that MBSE coupled to Natural Language Processing (NLP) and Artificial Intelligence (AI) is facilitating the Nuclear safety demonstration process and is allowing to anticipate any issue the project may face during the design, construction, commissioning and exploitation phases.

1 Introduction

It is understood that Nuclear Power Plant (NPP) design and building projects are more and more complex. This is mainly due to an increase in the number and nature of requirements induced by nuclear safety prescriptions. This situation encourages the development of new System Requirements Engineering approaches.

This is also due to the increasing volume and heterogeneity of data from both preliminary documents and lessons learnt from past analyses that are more and more requested in design phase in projects. Therefore, techniques to couple requirements engineering and data treatment are needed to demonstrate safety on complex installations. These are mainly related to a total, global and as exhaustive as possible understanding of safety requirements. In such a context, digital technology is essential for the processing of this mass of data. The need is to become able to perform requirements engineering more rapidly, efficiently, while being consistent and while maintaining a high level of reliability in our analyses of these documents.

The proposed solution is at the crossroads of System Engineering, particularly Model Based System Engineering, Natural Language Processing techniques and

Emir ROUMILI
IMT Alès
7 Rue Jules Renard, 30100 Alès
Assystem Engineering Operation Services
Tour Egée, 9-11 Allée de l'Arche, 92400 Courbevoie
eroumili@assystem.com

tools, both considered for Nuclear Safety domain. This article explains the global method we are proposing, followed by a general explanation of safety and the related industrial issues.

1.1 Tool-based methodology to complex-installation safety requirements

This article illustrates a tool-based methodology to complex-installation lifecycle considering its requirements. Our case study is the consideration of safety demonstration problematics. The proposed method is a part of an ecosystem of solutions aiming to the digitalisation of safety demonstration activities (more on that in section 2). Digitalisation is to turn information into a form that can be read easily by a computer [1]. This allows the processing of this data with the assistance of computers computational power. The purpose of this tool-methodology is to automatically create a project data base with well-formed requirements based on predefined structures, called "boilerplates". Thus, this task involves Natural Language Processing, requirements engineering modelling and management. The other parts of this ecosystem develop mostly the architecture modelling of the facilities. In this way, systems and functions can be linked to their respective requirements.

This ecosystem represents the way to proceed during a safety demonstration (defined in section 2). Generally, these steps are done manually with a written approach. This prevents a global vision of the safety demonstration and is longer than the proposed approach. This article focuses on the first tool: a methodology to extract and well-define requirements. The next sections briefly discuss the nuclear safety demonstration which is the core of our research work. Then, the documents of interest for this demonstration are addressed to highlight the contributions of such a method and the development of such a tool. The proposed approach and its related areas are developed as well as the contribution of artificial intelligence. Last section presents the conclusion and research opportunities.

2 Safety Demonstration

The proposed solution is illustrated through the demonstration of safety, especially the one relating to the nuclear industry. Indeed, the standards that regulate it are amongst the strictest in the world, on installations of high complexity and becoming more and more complex every year. Let us also point out that we see projects flourishing in countries where environment conditions and geopolitical issues are more demanding that also requires additional requirements to be embarked. Finally, business models need to be more precise as governments are all facing funding issues and they are looking to have project delivery model more robust concerning costs and timeline. The nuclear safety demonstration is at the

heart of the nuclear industry. It is the most important element and will always remain the limiting factor for all nuclear activities. Indeed, despite the fact that nuclear energy has an impressive efficiency and represents 77 % of the total electricity production in France in 2014 [2], it remains an energy that worries the public opinion. 40 % of the French population estimated in 2013 that the use of nuclear energy was rather inconvenient. [2]

The IAEA (International Atomic Energy Agency) defines nuclear safety as : *“The achievement of proper operating conditions, prevention of accidents or mitigation of accident consequences, resulting in protection of workers, the public and the environment from undue radiation hazards.”* [3] According to this definition, demonstration of safety is defined as follows: “Assessment of all aspects of a practice that are relevant to protection and safety; for an authorized facility, this includes siting, design and operation of the facility.” [3]

2.1 Problematic of safety demonstration in engineering

The demonstration of safety is a long, iterative process requiring a thorough analysis of regulatory texts (IAEA, ASN, feedback, etc.). This analysis will subsequently lead to a manipulation of these texts with an industrial objective: **“Demonstrate that a particular activity or installation is considered as to be safe in our country’s nuclear safety authority’s reference system and in the objectives we have set ourselves in the safety report.”**

To ensure that all operations are carried out safely, validation of the demonstration of safety is mandatory to obtain permission to license, build, operate, dismantle, etc. In this context, any demonstration of safety is part of an industrial project and is therefore a balance between different constraints of scope, schedule, budget, quality, resources etc. [4] The research, analysis, organization, and links that need to be established between reference documents and the installation or activity being demonstrated can quickly become time-consuming and thus lead to an increase in cost. However, the reduction of time and therefore of these costs, which is necessary in a competitive industrial world, may lead to an incomplete analysis. There is a need to increase the efficiency of these processes. This is what we are trying to answer through the proposal of our tool-based method.

It is to note that everything evoked with regard to nuclear safety requirements, which is at the crossroads of many different fields and requires excellence, leading to many documents to be analysed, is equally applicable to other industrial fields, nuclear or non-nuclear.

2.2 Body of regulatory text

The highest authority concerning nuclear industry is the country's nuclear safety authority. This authority must be independent [5] to ensure that it is not subject to political, operational or other influence. In France, the ASN (Nuclear Safety Authority) issues several types of documents, some are mandatory and some are

recommendations. The ASN has responsibility on pronouncing the final verdict on the demonstration of safety. It is therefore advisable to consider all the recommendations. In the figure below, the various documents issued in the context of the safety and linked to the nuclear safety demonstration.

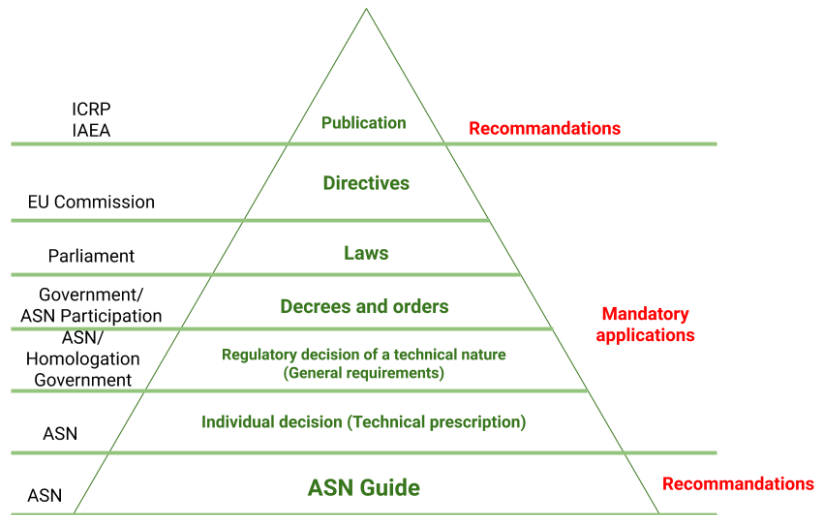


Figure 1 Body of regulatory text and their applications

The International Atomic Energy Agency (IAEA) is an international organization under the aegis of the United Nations seeking to promote the peaceful uses of nuclear energy and to limit the development of its military applications. In this role, the IAEA informs and publishes standards for the stability and safety of nuclear installations. These standards are recommendations, but since the IAEA is an extension of the safety authorities of the nuclear industry founding countries and is often adopted as a standard by nuclear industry emerging countries, their standards have a prominent place. We have chosen to use primarily the documents issued by the IAEA in our application case.

3 States of the art

3.1 State of the art of requirements engineering.

Stakeholders' requirements (i.e. needs specification that could remain often unclear and difficult to understand and formalise) and system requirements engineering (i.e. more classically viewed as a translation of these needs that engage the realisation of the expected solution) is globally hereafter called Requirements Engineering. It is a discipline shared by all engineering professions (software engineering, mechanics, energy, etc.). Systems Engineering [6]

recognizes it as the most upstream decisive phase in any project, i.e., products and/or services satisfying stakeholders¹ requirements.

Among technical, environmental and safety aspects, those requirements take also into account constraints and limits of cost, time, feasibility and quality. Activities included in requirements engineering [7] are:

1. Collecting the requirements from all stakeholders [8] and regulatory prescriptions.
2. Compiling and collating the requirements.
3. Establishment of the requirements.
4. Ensuring the expected qualities of the requirements (e.g. SMART).
5. Tracing, tracking and reporting the progress of requirements.

There are two categories of requirements [9]: functional requirements and non-functional ones. Functional requirements refer to the features and services provided by the system. Non-functional requirements constrain and specify how the system must meet the functional requirements in terms of performance, security, reliability, ergonomics, level of maintainability, portability, re-usability, etc. They may also concern aspects related to the project management such as cost and lead times. This categorization is important for the proper writing of requirements and their interactions with the installation architecture.

In our case, we focus on the notation of requirements which relies on two methods:

- Goal-oriented methods, or prescriptive methods, having as a basic concept the objective aiming to justify the need such as GBRAM [10] or KAOS [11].
- Scenario-oriented methods where one is focused on the sequence of interactions between the considered system and its environment to carry out an activity: Scenario-based Requirements Analysis [12].
- Properties based requirements engineering [13] is based on a theoretical framework allowing engineers to particularly focus on the expected properties (both functional and non-functional) that should characterize the system of interest.
- More global approaches allowing to model and formalize the requirements by using specific modelling languages such as SysML as proposed in [14]. In the same idea, [15] promotes MBSA approach and dedicated modelling languages here considered for the safety purpose.

There are also mixed methods based on the couple (goal, scenario) such as URML (Unified Requirements Modelling Language). These methods are interesting in the

¹ Any person, group of persons or organization that may influence or be affected by the system or its development, directly or indirectly [28]

height they take over the objectives and in the practicality they have in being linked to the scenarios. [16]

Natural language to express requirements remains frequent given its usability by stakeholders, regulatory and historic documents. However, these requirements, often written in free natural language, are unclear, ambiguous and unsuitable to automatic processing. Controlled Natural Languages (CNLs) address the informal aspect of free natural language by providing a formalized subset of natural language [17].

Among techniques relying on CNLs, the Boilerplates, or text patterns, are originally proposed to help engineers and analysts write quality documents and make them easier to read [18]. The boilerplates rely on the hole model concept. A boilerplate is a simple sentence model in which the grammar and part of the vocabulary are given. It is made with fixed elements and configurable attributes (in-between tags $\langle \rangle$) whose nature is known (e.g. the $\langle \text{user} \rangle$ shall be able to $\langle \text{capability} \rangle$). We could summarize:

Requirement Expression = Boilerplate + Placeholder values

We deemed it advisable to use "boilerplate" for our project. The boilerplate concept is adopted by several approaches and projects. We can cite the work "Pattern-based security requirements" [19] which applies it to system safety analysis where security requirements are defined using a set of boilerplates. This work is consistent for our research also positioned in the safety demonstration context. This context also involves studying and analysing regulatory texts, return of experience, etc. to draw up reference frameworks of applicable requirements. We can automatically extract this information from these unstructured texts.

3.2 State of the art of information extraction

Natural Language Processing is used in various application to processes natural language including Information Extraction (IE). We are applying Information Extraction techniques and tools that aim to extract the relevant data from a collection of documents. These are commonly divided into three fields [20] :

- Named-Entity Recognition: identification of specific entities in texts, by associating them with a defined type (names of organizations etc.).
- Relationship extraction: extracting relations existing between two entities in a text (A is done by B).
- Event extraction: filling automatically in an information structure represented as a template, associating different information elements with a given event. (Purchase event: date, amount etc.).

This differs from information retrieval, which is concerned with searching for a set of documents relevant to a given query. Symbolic models are the first work in

the field of information extraction. They use a set of rules manually defined by experts or a form of learning to extract information. [21]. In this context, rules are most often composed of words and other attributes derived from linguistic processing. These systems are usually composed of a consistent set of rules, with a possible overlap between some of them. Therefore, a set of constraints is necessary to trigger them. [20]. Concerning numerical models, their objective are to learn how to automatically associate classes to a set of elements [22]. These methods differ in the degree of supervision they require depending on the availability of training data. Within the framework of a module for knowledge extraction from texts, the recognition of named entities including disambiguation and the extraction of relationships are the indispensable branches of the process. They determine which information (entities, relationships) is used for further processing.

4 Contributions

4.1 Overview

The purpose of this method is to facilitate the safety demonstration of nuclear installations. We start from the premise that the demonstration of safety is based on three aspects (cf. figure 2):

- A requirements repository that is coherent, feasible and as complete as possible in terms of:
 1. Normative texts (IAEA, ASN, etc.), which therefore require a study and analysis of this text in order to derive applicable and generic reference systems of requirements (top-down approach).
 2. Feedback from experiences (lessons learnt), equivalent but uplifting work (bottom-up approach). By starting from experience, avoid making past mistakes but also to be able to transpose the generic requirements to particular cases of nuclear installations without missing the perceived problems and thus the financial and time losses that have resulted in the past.
- Modelling system of interest requirements considering and refining these more or less generic requirements, thus promoting a new requirements engineering approach in the MBSE (Model Based System Engineering) context.
- Demonstrate all requirements, particularly hereafter considered the safety requirements. This demonstration aims to prove and assume with a sufficient level of confidence that the modelled alternatives solutions of the system of interest respect these requirements. For this purpose, four strategies of verification and validation are first proposed:

1. To use expertise of these solutions as they are modelled. This induces the responsibility of recognized experts of both safety and nuclear domains and of modelling.
2. Guided modelling, i.e. modelling that reuses for instance pattern models [23], lessons learnt or past and approved solutions models that have been considered 'successful' and that can be used in confidence. This do not allow to furnish safety demonstrations but to guide engineers and obtain models that described already proved solutions.
3. Simulation i.e. the execution of the models of the system of interest. Models can be first considered separately in order to check their relevance for the expected demonstration. Second, they are federated and simulated together [24]. This allows us to obtain simulation that put in light the global behaviour of the system of interest, being more realistic. This federation can be called Digital Mock-Up and this could then evolve little by little towards Digital Twin of the system of interest.
4. Formal approaches allowing to formally establish proofs of evidence, e.g. by calculations or inference techniques that are recognised and by definition undisputable.

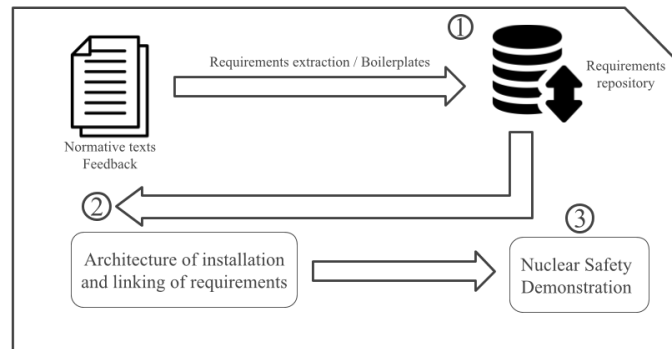


Figure 2 MBSE Guided methodology on nuclear safety demonstration.

This article focuses on the study and extraction of requirements repositories according to the top-down approach. However, the bottom-up approach is envisaged for further work or in an industrial context.

4.2 Contributions of artificial intelligence

It is considered appropriate to work initially on the IAEA's referential. Indeed, as explained in the previous section, IAEA documents are recognised useful in countries with a history of nuclear energy. Moreover, they often constitute the first regulations of emerging countries and a significant basis for their regulations once

they have a sufficient maturity. Thus, by optimizing our tool, initially for these documents, we will be able to make it interesting from the international safety point of view.

As mentioned in section 3.2, our objective is to make NLP algorithms learn how to automatically associate classes to a set of elements. Supervised methods will be used as they are considered more accurate and reliable in comparison to the results produced by unsupervised techniques of machine learning. This is mainly because the input data in the supervised algorithm is well known and labelled. However, these methods can be considered as expensive in the constitution of the dataset and is often requiring the assistance of experts in the field. The choice of AI is motivated by the difficulty underlying the identification of so-called "value" elements. It is already difficult for a human being to make a distinction; it would be even more difficult to ask him to issue rules in order to constitute an expert system. It makes more sense to let algorithms "find" the general patterns behind the choice of particular requirements. Also, data is often already available permitting to not monopolise the company's expert time. The difficulty is compounded by the desire to optimise and save time in order to deploy this type of solution in an environment constrained by costs. Subsequent ontological reasoning may allow the classification of requirements and knowledge management but will not be effective in this identification and extraction work.

4.2.1 Dataset constitution

The Dataset is set up based on the IAEA documents, mainly about risk characterisation in the context of the choice of nuclear sites for new installations.

IAEA Documents	Type
NSG3.2	Geology/Hydrogeological
NSG3.6	Geology/Hydrogeological
SSG9	Geology/Seismic/Bathymetry
SSG35	Seismic
Serie85	Seismic
Serie89	Seismic
SSG21	Volcano
SSG18	Oceanography/Bathymetry/Hydrogeological/Meteorology
NS-R-3 Rev1	Hydrogeological/Meteorology
GSR-Part-7	Meteorology

Table 1 Documents for requirements dataset constitution.

A total of 1141 requirements were extracted from these documents. Since the latest NLP models are pre-trained on large corpora (Wikipedia, DBpedia etc.) to make them learn syntax and semantics, this number is reasonable to get good results. In order to be able to train the model, the requirements were written as they are in the text. Indeed, if these requirements were to be reformulated the algorithm would not learn to recognize the requirements of our documents.

4.2.2 Selection of the model and verification

To train the model from our dataset, we use the BERT (Bidirectional Encoder Representations from Transformers) [25] model. This choice is done for the following reasons:

- The use of "transformers layers" popularized by the publication "Attention is all you need" [26], these layers in NLP allow :
 - Parallel processing of input data, which is time/resource consuming in NLP when using RNN (Recurrent Neural Network) layers because of their temporal component.
 - Multi-head attention allows us to better learn the interdependencies in our input data and thus have better results.
- Designed to pretrain deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context in all layers. This is mainly made possible using Masked LM (MLM) on the words during training.
- BERT obtained new state-of-the-art results on eleven natural language processing tasks. [25]
- Possibility to "fine-tune" the model for our needs. "Finetuning" consists in using a pre-trained model on tasks specific to our needs. We can quote ULM-FIT as one of the first models in NLP that introduced it [27] following the computer vision models allowing transfer learning after training on big sets of image data ("ImageNet", etc.).

A manual work was carried out in order to constitute our dataset of requirements. We are modelling the requirements extraction as a binary classification task: part of the input belongs to the "requirements" class and another part to the "non-requirements" class. A document parser is necessary to extract requirements candidates from raw text before passing them to the model. When the model is trained, the dataset will be divided into 3 parts (cf. figure 3):

- A set for training the model to recognize the requirements.
- A set for validation used during training to adjust model hyperparameters and thus avoid overfitting of the model. Optimized set of hyperparameters will allow us to perform well on new data.
- A test set. This set constitutes requirements that will never be seen by the model and it is on this dataset that the model will be checked.

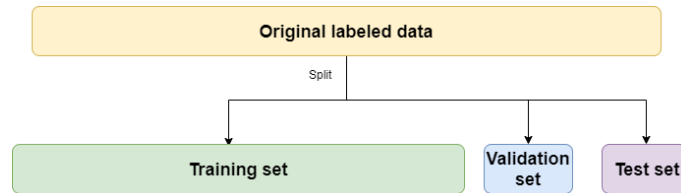


Figure 3 Dataset split

The results will be analysed using a confusion matrix (cf. figure 5) typically used in classification models.

Concerning the training, we have used the model pre-trained by Google teams (requiring huge computing capacities) and "fine-tuned" it on our requirements classification task to extract them afterwards. This consists of a recalculation of the superficial layers of the neural network. The deep layers having been pre-trained in an unsupervised manner on about 11,038 unpublished books (BookCorpus) and 2,500 Million words from text passages of English Wikipedia. After training the classification algorithm, we present here the results of F1 score on our test dataset (thus never seen by our algorithm). This measure is calculated from the precision and the recall. Precision is the number of correctly identified positive results divided by the number of all positive results, including those not identified correctly. Recall is the number of correctly identified positive results divided by the number of all samples that should have been identified as positive. The product in the numerator directly affects the score if there are extremes.

$$F = 2 \cdot \frac{(\text{précision} \cdot \text{rappel})}{(\text{précision} + \text{rappel})}$$

Figure 5 fl score measurement

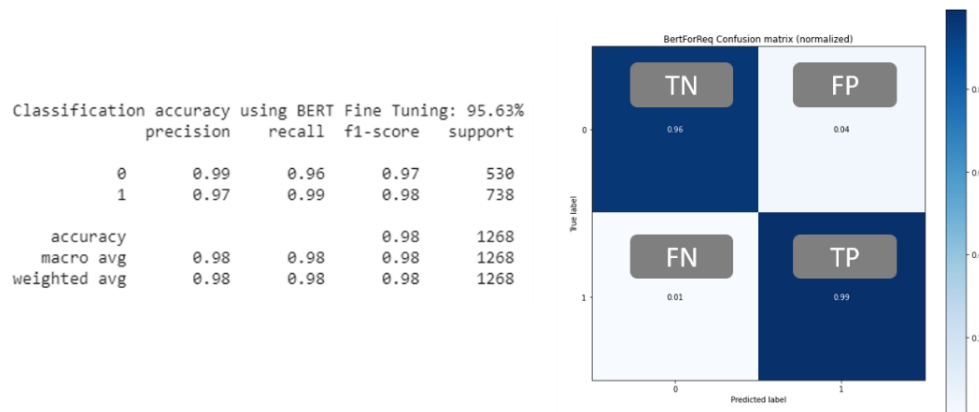


Figure 4 Confusion Matrix and f1 score for Requirements Classification on BERT

An example of extraction is shown in figure 6. The document is a test page that presents specifications for the evaluation of geotechnical aspects in the phase of selection of sites suitable for the construction of nuclear reactors. On this page, points 2.1 and 2.2 are requirements. Point 1.9 is a description of the contents of section 2.

The algorithm gives its results for each of these text blocks, we see that point 1.9 obtains a recognition score of about 0.5. Points 2.1 and 2.2 are selected with a reliability of more than 99%. It is then up to us to set our threshold value.

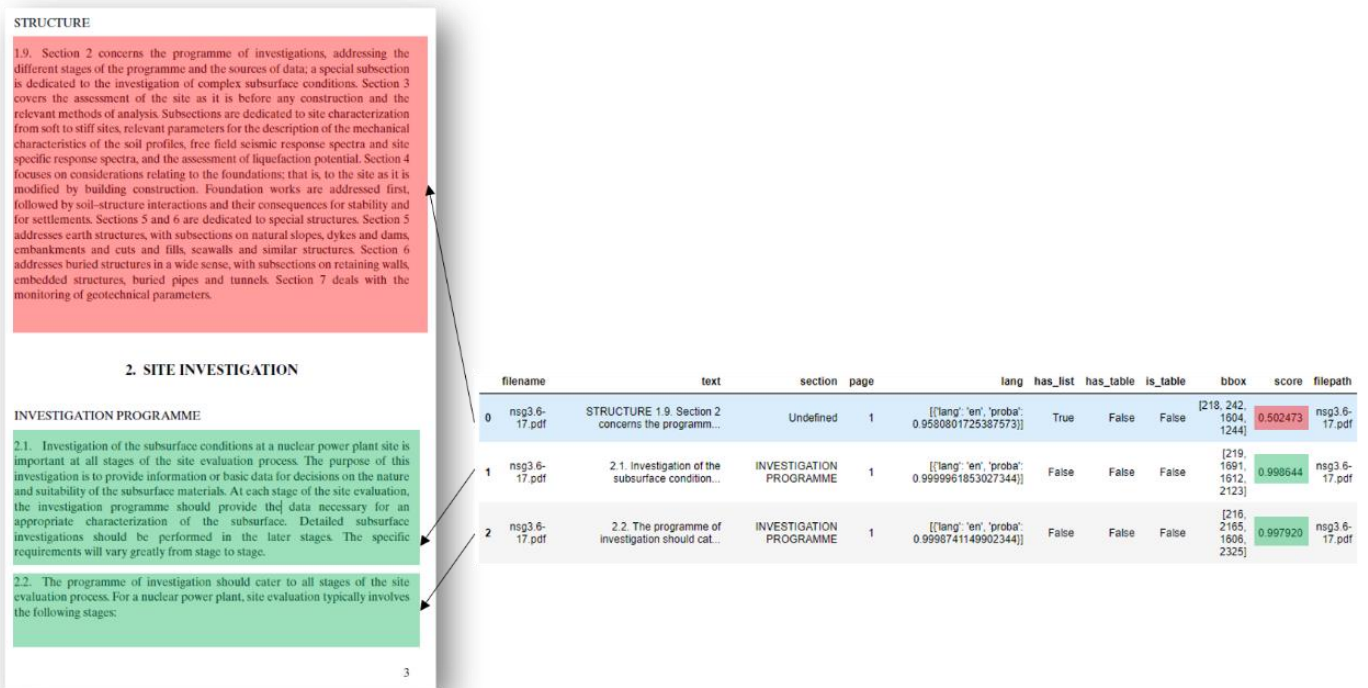


Figure 6 Example of extraction on a page with requirements and descriptive text

5 Conclusion and outlook

There is a risk that the licence for important projects construction and commissioning might be delayed or might never be obtained due to lack of traceability or of reproducibility. To reduce the risk, the use of digital techniques is essential in view of the number of extremely costly non-conformities in almost all complex projects. This research proposes the combination of system engineering and artificial intelligence as well as its application through the demonstration of nuclear safety which is a highly complex discipline and yet addressed in a document-oriented way. This "requirements extraction aid" remains associated to the need of the expected global method. This intent to support and formalise with rigour and robustness specification, analysis and proof of safety requirements. We claim that this requirements extraction technique proposed in this article, as automated as possible, will allow a saving of time and better quality in a project context where it is often what is lacking and may lead to studies that are either longer or of lower quality. As stated in the article, this type of approach can be applied to other areas of engineering since requirements engineering is cross-cutting.

The next steps will address the test of the model on real projects that Assystem is conducting in nuclear industry. Also, once the requirements are extracted, a process of requirements reformulation and modelling (in phase with MBSE practices and principles) via boilerplates and operational scenario modelling will be set up to have quality requirements for projects. This work will be integrated into the architecture modelling of an installation and the choice of the type of demonstration used (see 4.1).

6 References

- [1] Collins, The Collins English Dictionary 13th edition, HarperCollins, 2018.
- [2] Statista, “L'énergie nucléaire en France - Faits et chiffres,” 2017. [Online]. Available: <https://fr.statista.com/themes/2752/1-energie-nucleaire-en-france/>. [Accessed 08 05 2020].
- [3] IAEA, Safety Glossary STI/PUB/1290, International Atomic Energy Agency, 2007.
- [4] PMI, Project Management Body of Knowledge (PMBOK GUIDE) 5th Edition, Project Management Institute, 2013.
- [5] IAEA, Fundamental Safety Principles, International Atomic Energy Agency, 2006.
- [6] Guide to the Systems Engineering Body of Knowledge -- G2SEBoK, BKCASE Editorial Board, 2015.
- [7] M. Chemuturi, Requirements Engineering and Management for Software Development Projects, Springer, 2013.
- [8] I. Karla, S. Gomez , Y. Xin , C. Baron, P. Esteban, C. Yollohtli and A. G. Estrada, “Avez-vous identifié toutes les Parties prenantes ?,” *Congrès international MOSIM 2018, MOdélisation et SIMulation des systèmes, Toulouse, Novembre 2018, France, 2018*”.
- [9] ISO, ISO/IEC 15288 Systems and software engineering — System life cycle processes, ISO, 2008.
- [10] A. I. Anton, “Goal-Based Requirements Analysis,” *ICRE '96: Proceedings of the 2nd International Conference on Requirements Engineering (ICRE '96)*, 1996.
- [11] A. van Lamsweerde, “The KAOS Project: Knowledge Acquisition in Automated Specification of Software,” *American Association for Artificial Intelligence*, pp. 59-62, 1991.
- [12] A. Sutcliffe, “Scenario-based requirements analysis,” *Requirements Engineering*, pp. 48-65, 1998.

- [13] P. Micouin, "Toward a property based requirements theory: System requirements structured as a semilattice," *Systems Engineering*, Vols. 11, Issue 3, pp. 235-245, 2008.
- [14] J.-F. Pétin, D. Evrot, G. Morel and P. Lamy, "Combining SysML and formal methods for safety requirements verification," *22nd International Conference on Software & Systems Engineering and their Applications*, 2010.
- [15] C. Haskins, "Foundations for model-based systems engineering and model-based safety assessment," *Systems Engineering*, Vols. 22, Issue 2, pp. 146-155, 2019.
- [16] S. F. a. N. H. a. Berenbach, "A modeling language to support early lifecycle requirements modeling for systems engineering," *Procedia Computer Science*, vol. 8, pp. 201-206, 2012.
- [17] T. Kuhn, "A Survey and Classification of Controlled Natural Languages," *Association for Computational Linguistics ETH Zurich and University of Zurich*, 2014.
- [18] H. E. a. J. K. a. D. Jeremy, *Requirements Engineering*, Springer, 2015.
- [19] S. T. a. F. S. a. Daramola, "Pattern-based security requirements specification using ontologies and boilerplates," *Requirements Patterns (RePa), 2012 IEEE Second International Workshop*, pp. 20-32, 2012.
- [20] L. Jean-Louis, *Approches supervisées et faiblement supervisées pour, UNIVERSITE PARIS 11 - PARIS SUD ECOLE DOCTORALE EDIPS*.
- [21] I. Muslea, "Extraction Patterns for Information Extraction Tasks: A Survey," *The AAAI-99 workshop on machine learning for information extraction. T. 2. Orlando Florida.*, 1999.
- [22] Y. S. C. a. D. Roth, "Exploiting Syntactico-Semantic Structures for Relation Extraction," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies-Volume 1, Association for Computational Linguistics*, pp. 551-560, 2011.
- [23] V. Chapurlat, F. Pfister, M. Huchard, C. Nebut and J.-L. Wippler, "A proposed meta-model for formalizing Systems Engineering knowledge, based on functional architectural patterns, *International Journal Systems Engineering, INCOSE Eds.*, 2012.

- [24] V. Chapurlat and B. Nastov, "Deploying MBSE in SME context: revisiting and equipping Digital Mock-Up," *6th IEEE International Symposium on System Engineering (ISSE)*, 2020.
- [25] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018.
- [26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.
- [27] J. Howard and S. Ruder, "Universal Language Model Fine-tuning for Text Classification," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2018.
- [28] Sharp, Finkelstein and Galal, "Stakeholder Identification in the Requirements Engineering Process," *Proceedings of 10th International Workshop on Database & Expert Systems Applications (DEXA)*, pp. 387-391, 1999.