Coupling Ontology and Probability to manage risks in construction projects

Tran Thi Thuy Phuong[1], Franck Taillandier[1], Cédric Baudrit[2], Denys Breysse[1]

[1] *Univ. Bordeaux, I2M, UMR 5295, F-33400 Talence, France*

[2] *Univ. Bordeaux, INRA-I2M, USC 1368, F-33400, Talence, France*

RÉSUMÉ. *Sous l'effet de causes internes ou externes, les résultats attendus d'un projet de construction (temps, coût et qualité) peuvent être modifiés. Afin de répertorier toutes les sources qui pourraient avoir un impact dans un projet de construction et d'analyser leur propagation de façon systématique, nous avons développé un modèle holistique. Ce modèle est basé sur le formalisme des modèles relationnels probabilistes (MRP) qui couple une structure sémantique du domaine de connaissance (connu sous le nom d'ontologie) avec des modèles graphiques probabilistes (connus sous le nom de réseaux Bayésiens). Nous avons élaboré une représentation générique d'un projet de construction en tenant compte de l'incertitude inhérente à la nature de la connaissance. La notion de risque est capturée à partir de l'écart entre la valeur prédite d'un attribut et sa valeur réelle pour un scénario de construction. Afin de tester le modèle et de mettre en évidence son intérêt, un cas réel simplifié a été instancié s'appuyant sur un projet routier à Hue, Vietnam*

ABSTRACT. *Under internal or external impacts, the expected results (time, cost and quality) of a construction project may be changed. So, the success of construction projects, i.e. reaching conceptives regarding cost, delay and quality, is a major issue. In order to list all sources which could impact a construction project and to formalize uncertainty in a systematic way, we developed a stochastic holistic model. This model is based on the Probability Relational Model (PRM) formalism combining a semantic structure of knowledge domain (known as Ontology) with probabilistic graphical models (known as Bayesian networks). We propose a generic representation of a construction project by taking into account uncertainty inherent to the nature of knowledge. Risks will be considered as being the deviation between the forecast value of some variables and their real value given a construction scenario. In order to test the model and to highlight its interest, a simplified real case will be implemented which is based on the project 2BS in Hue, Vietnam.*

MOTS-CLÉS : *Projet de construction, Gestion des risques, Ontology, réseaux Bayésiens, modèles relationnels probabilistes.*

KEYWORDS: *Construction project, Risk management, Ontology, Bayesian networks, Probabilistic relational model.*

1. Introduction

A construction project can be a process of alteration, conversion, commissioning, renovation, maintenance, refurbishment, demolition, decommissioning or dismantling of a structure [Wor 13]. Each project is a unique (ISO 10006) and temporary process that consists of many activities which are performed and controlled by actors to achieve requirements under the conditions of time, cost, quality. Most construction projects do not reach their expected results (time, cost and quality) due to the internal and external environment variations[LOV 02]. The behavior of the construction project system can thus be apprehended as a random system due to intrinsic uncertainties, we will attempt to develop suitable model for this purpose.

Risk management has an important role in the success of construction project. "The purpose of risk management is to identify potential problems before they occur so that risk-handling activities may be planned and invoked as needed across the life of the product or project to mitigate adverse impacts on achieving conceptives" [KHR 12]. Risk management is divided into : risk identification, risk analysis, risk response planning, control and monitoring the risks [MEH 12]. Many methods can be used in the process of risk identification (literature review, brainstorming, cause and effect diagrams, check list, interviews, questionnaire, ...) as well as for risk analysis (SWOT, Monto Carlo technique, Bayesian Network ...) Each method has advantages and disadvantages. For example : with a questionnaire, in a short period of time, all necessary information can be collected by a large number of people, moreover, results will be collected quickly and easily, but there are some drawbacks : the honesty level of respondents are unknown, different meanings of questions between respondents limited range of questions and answers, low response rate and sampling biases.

In the process of risk identification, the literature review will be privileged because it offers an easy way to access theories, methodologies. Probability framework is commonly used to assess risks [KUO 13] [YIL 14] [JEN 05] [HWA 14] [SER 14] ...Bayesien networks (BNs) "are graphical structures for representing the probabilistic relationships among a large number of variables and doing probabilistic inference with those variables" [NEA 10]. BNs have become popular as a tool for modelling forensic evidence. Many works used BN to analyze risks in construction [LEE 12] [LUU 09] However, BN cannot be used to deal with domains where we might encounter a varying number of concepts (entities) in a variety of configurations and lack the objects (concepts in instantiation). If just BN is applied in a complex project (hundreds of relations), the model quickly lost its expressiveness, many repetitive tasks while modeling, work is so tough and tedious. Another problem is how to systemize semantic knowledge of a complex construction project ? In recent years, many researches have applied ontology for studying risk in construction [TSU 10] [FID 08] [RAH 15] [ZHO 15] [ECE 14] [FUE 07] [LIM 05] ...It established semantic structures of all concepts (entities) which define a construction project. If we just use purely ontology, structures of all concepts will be defined but it does not permit to take uncertainty into account inherent to the nature of knowledge. Although there exists a substantial literature about risk management in civil engineering, none proposes a holistic approach which able to represent the whole system of construction projects by taking uncertainty into account (Combining bayesian networks, ontology).

Section 2 gives an overview about the manipulated mathematical tools. Section 3 proposes a holistic modeling of construction projects based on PRMs. Section 4 shows an implementation of a simplified real case based on road and bridge construction in Hue, Vietnam. Finally, Section 5 draws the conclusion of the work.

2. Presentation of used formalisms

2.1. *Ontology*

Ontology is a structured representation of a knowledge domain. It is composed of a hierarchic organization of concepts and relations that exist between themselves with rules and axioms which constrain them. It provides a reference for the communication as well between machines as between humans and machines by defining the meaning of concepts. [LIM 11]

In our context, we adopt a formal definition of ontology given by a 4-tuple $< \mathcal{C}, \mathcal{R}, \mathcal{A}, \mathcal{I} >$ where $\mathcal{C}, \mathcal{R}, \mathcal{A}, \mathcal{I}$ are disjoint sets containing concepts, relations, attributes and instances. Concept C $\in \mathcal{C}$ (other synonymous terms classes, entities, types, categories) represents a set of terms, things having a common sense in a domain. As the concept Actor (all of people who works in a construction project) . Each concept is described, characterized by :

– A set of concept's relations $\mathcal{R}$ clustered into :

- Taxonomies that define a partial order over the set of concepts characterized by sub-super-concept tree structures and known as specification relation. For example : Contractor is an Actor

- Associative relationships that relate concepts across tree structures, known as nominative relationships describe the names of concepts, locative relationships describe the location of one concept with respect to another. For example : Contractor impacts on Activities in operational phase.

- Partonomies that describe concepts that are parts of other concepts. For example : Worker is apart of Contractor.

– A set of attributes relation $\mathcal{A}$ (properties, functions) such that each attribute relation $C.A$ denotes an attribute of C and $c.A$ denotes the value of attribute $C.A$ for the element c of type C by adopting the objet paradigm. For example : John is an element of type Actor and John.Productivity=slow meaning that John is an actor with a low productivity.

– Instances $\mathcal{I}$ are the elements represented by a concept. For example : In a real project, there are two contractors (Company A, Company B).

2.2. *Bayesian networks (BN)*

A Bayesian Network (BN) [KOS 12] is a graph-based model of a joint multivariate probability distribution that captures properties of conditional independence between variables. Formally, a Bayesian network is a directed acyclic graph (DAG) whose nodes X_i represent random variables, and whose missing arcs encode conditional independences between the variables. This graph is called the structure of the network and the nodes containing probabilistic information are called the parameters of the network . In a Bayesian network, the joint probability distribution (JPD) of the node values can be written as the product of the local probability distribution of each node and its parents. If the set of parent nodes of a node Xi is denoted by Pa(Xi), we have :

$P(X_i, \ldots, X_n) = \prod_{X_i \in (X_1,\ldots,X_n)} P(X_i|Pa(X_i))$

Different methods exist to learn about the structure or the parameters from substantial and/or incomplete data and readers may refer to [Mic 98]. Use of such BNs consists in query expressed as conditional probabilities. The most common task we wish to solve is to estimate the marginal probabilities $P(X_Q|X_E = x_E)$ where X_Q is a set of query variables, and X_E is a set of evidence variables.

2.3. *Probability Relational Model (PRM)*

PRM [GET 07] provides a practical mathematical formalism and an intuitively appealing interface for human experts to model highly-interacting sets of random variables allowing to describe stochastic complex systems. PRM formalism couples a semantic structure of knowledge domain (see section 2.1) [GUI 05] with probabilistic graphical models (see section 2.2) [Jud 91] [KOS 12].

A PRM is a directed acyclic graph whose attribute $C.A$ (by adopting the object paradigm) represent random variables. Formally, PRMs may be formalized by :

– A set of nodes $c.A$ for all $C \in \mathcal{C}$, $c \in \mathcal{I}(C)$ and $A \in \mathcal{A}(C)$ where $\mathcal{I}(C)$ denotes the set of instances of type C and $\mathcal{A}(C)$ denotes the set of attributes relation associated with the concept C.

– A set of parents $Pa(c.A) = (U_1, ..., U_l)$ where U_i may have the form $c.B$ or $\gamma(c.R.B)$ for all $i \in (1, ..., l)$ where R is a slot chain (combination of relations) and γ is an aggregation given a summary of a multiset of random variables $c'.B/(c, c') \in R$ where $R = \{(c, c'), c \in \mathcal{I}(C), c' \in \mathcal{I}(C')\}$.

– The conditional probability distributions $P(c.A|Pa(c.A))$, by assuming that $P(c.A|Pa(c.A)) = P(c'.A|Pa(c'.A))$ for all $(c, c') \in \mathcal{I}(C)^2$. For $u \in \mathcal{V}(U)$, $P(c.A|u) : \mathcal{V}(C.A) \longrightarrow [0, 1]$ defines a probability mass distribution over V(C.A) where V(C.A) denotes the domain of attribute values.

There are several aggregation methods like mode (most frequently occurring value) ; mean value ; median ; maximum ; minimum etc. PRMs define a probability distribution over possible worlds instantiated by a skeleton, denoted $\sigma_{\mathcal{R}}$, specifying the set of concepts $\sigma_{\mathcal{R}}(C)$ for each concept and the relations between the concepts, i.e $\sigma_{\mathcal{R}}(C) =< \mathcal{I}(C), \mathcal{R}(C) >$. The difference between $\sigma_{\mathcal{R}}$ and the ontological model instantiation comes from we do not affect any values to attributes. For a given skeleton $\sigma_{\mathcal{R}}$, the PRM structure induces a ground Bayesian network and we have :

$$P(\mathcal{I}) = \prod_{C\in\mathcal{C}} \prod_{A\in\mathcal{A}(C)} \prod_{c\in\sigma_r(C)} P(c.A|Pa(c.A)) \tag{1}$$

3. Generic model of construction project

Using the PRM, we develop a generic model allowing to formalize any kind of construction project. Based on document of AIA A201-2007 [NIU 15], the model is centered around six main concepts : **external environment (EE), actor, resource, contract, activity and product** (figure 1) . The model formalizes the entities (node) and their semantic relations (arcs).

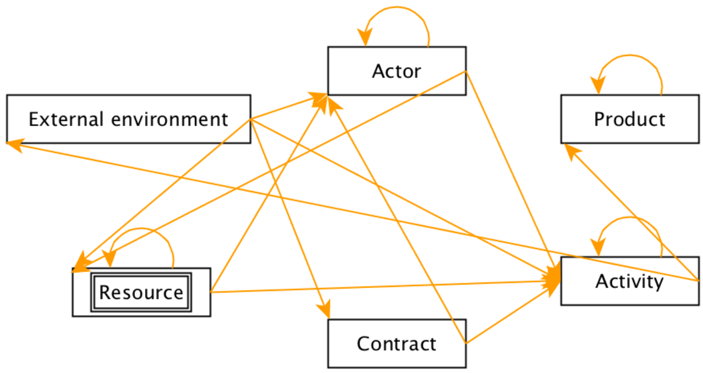


Figure 1 – Relations between construction project entities.

Each concept is characterized by attributes which can impact, more or less directly, the success of construction project. For instance, the Capacity of an Actor may impact the Duration on an Activity, and in fine, the time to achieve the project (one of project objective). Relation between attributes corresponds to conditional probabilistic dependence in our PRM framework. Figure 2 exposes the attributes and the probabilistic dependence between attributes of the six main entities.

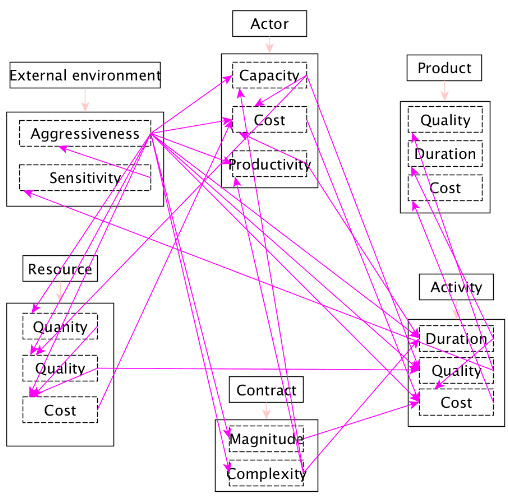
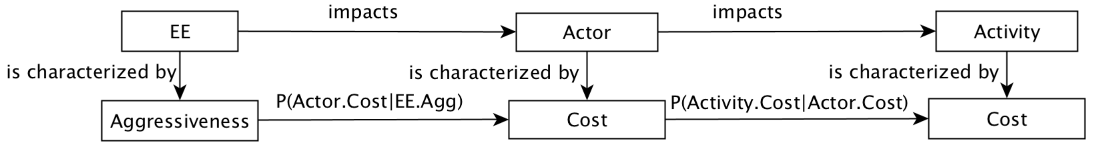


Figure 2 – Dependences between construction project attributes

Obviously, the attribute relation graph has to be consistent with the entity relation graph. A relation between attributes of two different entities requires the being of a relation between these two entities (and reciprocally). Each attribute is assessed according to a common 5-level Likert scale $\mathcal{V}(C.A) = (-2, -1, 0, 1, 2)$ which corresponds respectively to : much more negative, more negative, expected value, more positive, much more positive. An attribute of an entity instance is characterized by a probability distribution defining its belonging to these seven possible states. For instance, Figure 3 exposes two different distributions : P(EE.Aggressivness) and P(Actor.Cost|EE.Aggressiveness) (or P(Activity.Cost|Actor.Cost)).



Input	
Value	P(EE.Agg) (%)
-3	1,7
-2	1,7
-1	17
0	21
1	25
2	28
3	5,6

Output	
Value	P(Activity.Cost|EE.Agg)
-3	?
-2	?
-1	?
0	?
1	?
2	?
3	?

Figure 3 – Input and output probability of attributes

The probabilistic dependence between attributes enables to compute value distribution of an output attribute (e.g. Activity.Cost) considering input values (e.g. EE.Aggressiveness). The calculation is based on BN. In PRM, the CPT are generic and identical for every project. The information about conditional probability distributions is a transposable knowledge. The definition of CPT is thus a major issue of the PRM model. In the current version of the model, the CPT was defined for the six main entities attributes.

4. Instantiation of the project 2BS

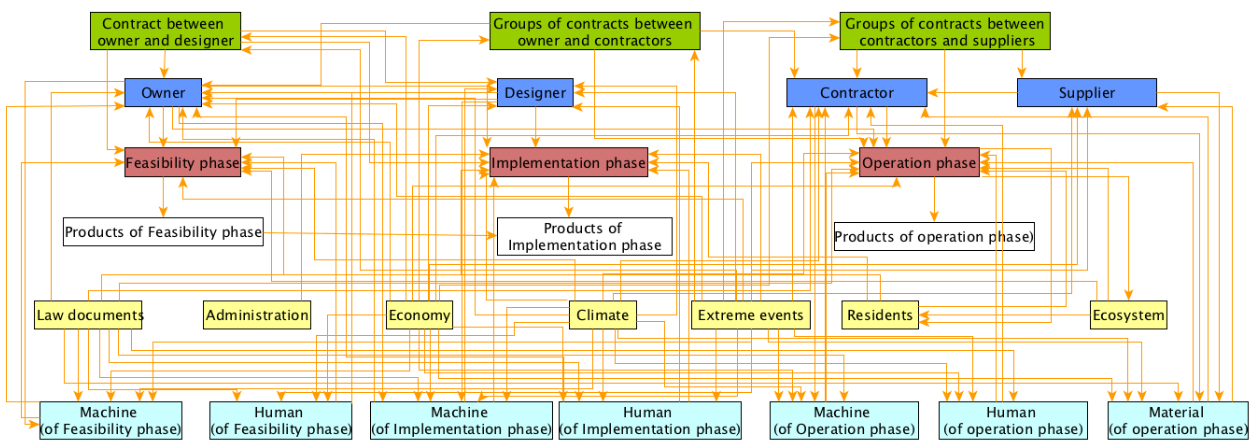


Figure 4 – Relational skeleton (σ_r) of project 2BS

2BS projects concerns a road and bridge project which is in Hue, a province in the central part of Vietnam. It is located near Huong river, building residence (more than 60 houses), infrastructures, and another project. Length of this project is 368.49m (198.29m + 170.2m). Based on plan, cost of project was estimated to more than 6 billion Vietnam dong (VND) (≈ 242 000 euro). Before this project, there was an old road and bridge (Lich Doi) which was built 17 years ago. But because of degradation and increasing demands in transportation, the renovation and upgrading project 2BS was performed during the 2013. According to the plan, project 2BS must be performed in 10 months, but because of many negative impacts (minimum salary of worker increased, inflation of economic, temperature was higher than normal condition, there was a big storm with flood, negative impact of residents around project, some workers were fired and ill, some machines were broken), it last 2 months more and cost of project was increased by more than 0,5 billion VND.

In the project 2BS, the number of objects of the concepts was respectively, 7 for the external environment (Document law, administration, economic, climate, extreme event, ecosystem, third party) , 4 for actors (Contractors, Owner, Designer, Supplier), 7 for resources (machine and human in feasibility phase, machine and human in implementation phase, machine, human and material in operation phase), 3 for contracts (contracts between owner and contractors, contract between owner and designer, contracts between contractors and suppliers), 3 for activities (grouped in feasibility phase, implementation phase, operation phase), 3 for products (corresponding to the same phases). (figure 4)

Figure 5 – Probability distributions of all variables given the scenario in red boxes.

These instances inherit all attributes of concepts which contain them i.e objects of a concept have the same attributes. Therefore, the attributes of Owner, Designer, Contractor, Supplier (objects of concept Actor) are Capacity, Cost, Productivity ; Attributes of Products of Feasibility phase, Implementation phase, Operation phase (objects of concept Product) are Quality, Duration, Cost . . . All relations between objects and their attributes are created according to the generic model described in Section 3. However, an existing relation between two concepts does not imply that all corresponding objects have the same relation. For example : Actor impacts on Activity, but not all objects of Actors impact on all objects of Activity (Owner impacts on activity in Feasibility phase and implementation phase. Contractor impacts on activity in Operation phase, ...). Conversely, relations between attributes of objects are totally like in the generic model (Capacity of Owner impacts on Quality of activities in Feasibility phase, or Capacity of Contractor impacts on Quality of activities in Operation phase,...).

Figure 5 displays the probability distributions over all the variables of project 2BS given a scenario (red boxes). This scenario consists of extreme events, inflation, pay increase, climate conditions, personal accident and breakdown machines. We can see that $P(ProductOperation.Duration = -2|scenario) = 40\%$ (blue circle), i.e. there is a 40% probability that duration of operation phase will be late, which is compatible with the feedback from the real 2BS project.

5. Conclusions and perspective

Risk modeling in construction projects is a complex challenge because of the variety of uncertainty sources and the complexity of interactions. We developed a suitable model to formalize and propagate risks in construction projects. It is based on a specific ontology and a formal modeling pattern, where entities and concepts can be described at various scales, depending on the level of detail which is looked for. A first version of the model has been implemented on a dedicated tool, with the purpose of checking its practical feasibility and consistency. This work is still on-going and the next step will be to check the consistency of the simulations when the project is analyzed at various scales. The model will also be applied to other more important civil engineering projects, and we will finally study how the model can be used in practical project risk management, in order to identify its more risky areas and find efficient risk responses

6. Bibliographie

[ECE 14] ECEM A., DIKMEN I., BIRGONUL M. T., ERCOSKUN K., ALTEN S., « Automation in Construction A knowledge-based risk mapping tool for cost estimation of international construction projects », *Automation in Construction*, vol. 43, p. 144–155, Elsevier B.V., 2014.

[FID 08] FIDAN G., « a Risk and Vulnerability Ontology for Construction Projects », , n° August, 2008.

[FUE 07] FUERTES A., FORCADA N., CASALS M., GANGOLELLS M., « Development of an Ontology for the Document Management Systems for Construction », *Complex Systems Concurrent Engineering*, p. 529–536, 2007.

[GET 07] GETOOR L., FRIEDMAN N., KOLLER D., PFEFFER A., TASKAR B., « Probabilistic Relational Models », 2007.

[GUI 05] GUIZZARDI G., *Ontological Foundations for Structural Conceptual Models*, 2005.

[HWA 14] HWANG B. G., ZHAO X., TOH L. P., « Risk management in small construction projects in Singapore : Status, barriers and impact », *International Journal of Project Management*, vol. 32, n° 1, p. 116–124, Elsevier Ltd and APM IPMA, 2014.

[JEN 05] JENKINSON D., « The Elicitation of Probabilities - A Review of the Statistical Literature », , n° 1988, p. 1–72, 2005.

[Jud 91] JUDEA PEARL, *Probabilistic reasoning in intelligent systems : Networks of plausible inference*, vol. 28, 1991.

[KHR 12] KHRAIWESH M., JORDAN Z., « R Isk M Anagement M Easures in Cmmi », vol. 3, n° 1, p. 149–163, 2012.

[KOS 12] KOSKI T. J. T., NOBLE J. M., « A Review of Bayesian Networks and Structure Learning », *Mathematica Applicanda*, vol. 40, n° 1, p. 53–103, 2012.

[KUO 13] KUO Y.-C., LU S.-T., « Using fuzzy multiple criteria decision making approach to enhance risk assessment for metropolitan construction projects », *International Journal of Project Management*, vol. 31, n° 4, p. 602–614, Elsevier Ltd and IPMA, 2013.

[LEE 12] LEE H., LEE K., PARK M., « RFID-based Real-Time Locating System for Construction Safety Management », *Journal of Computing in Civil Engineering*, vol. 26, n° 3, p. 366–378, 2012.

[LIM 05] LIMA C., FEIS B., « Domain Taxonomy for Construction Concepts : Toward a Formal Ontology for Construction Knowledge », , n° October, p. 394–406, 2005.

[LIM 11] LIM E. H. Y., LIU J. N. K., LEE R. S. T., *Knowledge Seeker – OntologyModelling for Information Search and Management*, 2011.

[LOV 02] LOVE P. E. D., « Using systems dynamics to better understand change and rework in construction project management systems », vol. 20, n° August 2015, p. 425–436, 2002.

[LUU 09] LUU V. T., KIM S.-Y., TUAN N. V., OGUNLANA S. O., « Quantifying schedule risk in construction projects using Bayesian belief networks », *International Journal of Project Management*, vol. 27, n° 1, p. 39–50, 2009.

[MEH 12] MEHDIZADEH R., « Dynamic and multi-perspective risk management of construction projects using tailor-made Risk Breakdown Structures », , n° Ed 209, 2012.

[Mic 98] MICHAEL I. JORDAN, « Learning in Graphical Models (Adaptive Computation and Machine Learning) », 1998.

[NEA 10] NEAPOLITAN R. E., « Learning Bayesian Networks », *International journal of data mining and bioinformatics*, vol. 4, p. 505–19, 2010.

[NIU 15] NIU J., ISSA R. R. A., « Advanced Engineering Informatics Developing taxonomy for the domain ontology of construction contractual semantics : A case study on the AIA A201 document », *Advanced Engineering Informatics*, vol. 29, n° 3, p. 472–482, Elsevier Ltd, 2015.

[RAH 15] RAHMANI M. B., « A taxonomy for causes of changes in construction », *PhD Proposal*, vol. 1, n° September, 2015.

[SER 14] SERPELLA A. F., FERRADA X., HOWARD R., RUBIO L., « Risk Management in Construction Projects : A Knowledge-based Approach », *Procedia - Social and Behavioral Sciences*, vol. 119, p. 653–662, Elsevier B.V., 2014.

[TSU 10] TSUI E., WANG W. M., CHEUNG C. F., LAU A. S. M., « A concept-relationship acquisition and inference approach for hierarchical taxonomy construction from tags », *Information Processing and Management*, vol. 46, n° 1, p. 44–57, 2010.

[Wor 13] WORK HEALTH AND SAFETY ACT, *Code of Practice*, N° November, 2013.

[YIL 14] YILDIZ A. E., DIKMEN I., BIRGONUL M. T., « Using Expert Opinion for Risk Assessment : A Case Study of a Construction Project Utilizing a Risk Mapping Tool », *Procedia - Social and Behavioral Sciences*, vol. 119, p. 519–528, Elsevier B.V., 2014.

[ZHO 15] ZHONG B., LI Y., « An Ontological and Semantic Approach for the Construction Risk Inferring and Application », p. 449–463, 2015.