

Apports des ontologies à la création de bases de connaissances pour la maîtrise des conformités légales en santé et sécurité au travail

Jonathan Vigneron, Jean-Marc Rallo, Franck Guarnieri

► **To cite this version:**

Jonathan Vigneron, Jean-Marc Rallo, Franck Guarnieri. Apports des ontologies à la création de bases de connaissances pour la maîtrise des conformités légales en santé et sécurité au travail. [Rapport de recherche] CRC_WP_2014-19, MINES ParisTech. 2014, 8 p. hal-00990835

HAL Id: hal-00990835

<https://hal-mines-paristech.archives-ouvertes.fr/hal-00990835>

Submitted on 14 May 2014

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.



PAPIERS DE RECHERCHE **CRC** WORKING PAPERS SERIES

CRC_WP_2014_19

(février 2014)

APPORTS DES ONTOLOGIES À LA CRÉATION DE BASES
DE CONNAISSANCES POUR LA MAÎTRISE DES CONFORMITÉS LÉGALES
EN SANTÉ ET SÉCURITÉ AU TRAVAIL

Jonathan Vigneron, Jean-Marc Rallo, Franck Guarnieri



CENTRE DE RECHERCHE SUR LES RISQUES ET LES CRISES
MINES ParisTech
Rue Claude Daunesse CS10207
06904 Sophia Antipolis Cedex
www.crc.mines-paristech.fr

PAPIERS DE RECHERCHE DU CRC

Cette collection a pour but de rendre aisément disponible un ensemble de documents de travail et autres matériaux de discussion issus des recherches menées au CRC (CENTRE DE RECHERCHE SUR LES RISQUES ET LES CRISES).

Tous les droits afférant aux textes diffusés dans cette collection appartiennent aux auteurs.

Des versions ultérieures des papiers diffusés dans cette collection sont susceptibles de faire l'objet d'une publication. Veuillez consulter la base bibliographique des travaux du CRC pour obtenir la référence exacte d'une éventuelle version publiée.

<http://hal-ensmp.archives-ouvertes.fr>

CRC WORKING PAPERS SERIES

The aim of this collection is to make easily available a set of working papers and other materials for discussion produced at the CRC (CENTRE DE RECHERCHE SUR LES RISQUES ET LES CRISES).

The copyright of the work made available within this series remains with the authors.

Further versions of these working papers may have been submitted for publication. Please check the bibliographic database of the CRC to obtain exact references of possible published versions.

<http://hal-ensmp.archives-ouvertes.fr>

CENTRE DE RECHERCHE SUR LES RISQUES ET LES CRISES
MINES ParisTech
Rue Claude Daunesse CS 10207
06904 SOPHIA ANTIPOLIS Cedex
www.crc.mines-paristech.fr

Apports des ontologies à la création de bases de connaissances pour la maîtrise des conformités légales en santé et sécurité au travail

Jonathan Vigneron^{ab}, Jean-Marc Rallo^b, Franck Guarnieri^a

^a MINES Paristech, Centre de recherche sur les Risques et les Crises (CRC), Sophia Antipolis, France

^b Preventeo, Le Cannet, France

(Février 2014)

Pour citer ce document : Vigneron J., Rallo J.M., Guarnieri F., Apports des ontologies à la création de bases de connaissances pour la maîtrise des conformités légales en santé et sécurité au travail. Papier de recherche du Centre de recherche sur les Crises et les Crises (CRC) de MINES Paristech, février 2014.

Résumé : La "Santé-Sécurité au Travail" (SST) est au cœur des politiques d'entreprise et la judiciarisation de la société concourt à une inflation du nombre de textes réglementaires publiés chaque année. Les préventeurs sont donc confrontés au traitement croissant de données afin de rester en conformité. Ils se forment au droit et s'entourent de compétences interne (direction juridique) ou externe (bureaux de contrôle et de conseil) et se dotent de systèmes d'information. Ces derniers sont principalement des bases de données de "veille juridique" proposant des textes réglementaires accompagnés de commentaires. L'ensemble de ces données demeure complexe à interpréter et à exploiter en raison du nombre croissant de textes, de l'expertise nécessaire à fin de les comprendre et de la difficulté à s'assurer qu'ils concernent les activités de l'entreprise. Ce besoin peut être satisfait par le recours aux modèles et méthodes de l'intelligence artificielle. Parmi ceux-ci, les concepts d'ontologies et de bases de connaissances ont été retenus. Une ontologie est un modèle d'un domaine particulier de connaissances réalisé dans un but précis

Mots clefs: ontologie, base de connaissances, conformités réglementaires, santé et sécurité au travail

1. INTRODUCTION

En France, la "Santé Sécurité au Travail" (SST) est au cœur des politiques d'entreprise et la judiciarisation de la société concourt à une augmentation très significative du nombre de textes réglementaires publiés chaque année par le législateur. Les préventeurs en charge de la mise en œuvre de la politique en matière de gestion de la Santé, Sécurité au Travail (SST), tels que les animateurs Hygiène, Sécurité et Environnement ou les responsables sécurité, sont confrontés à cette situation qu'ils ont bien du mal à gérer au quotidien [1][2][3]. Ils se forment pourtant au droit du travail et s'entourent de compétences internes (direction juridique) ou externe (bureaux de contrôle et de conseil). Ils se dotent parfois de systèmes d'information (des bases de données, des solutions logiciels plus élaborées...) [4]. Toutefois, le nombre et la nature des textes publiés régulièrement se traduisent par une difficulté certaine à traiter les nouvelles informations et obligations, du fait d'un besoin accru d'expertise technique nécessaire à la compréhension et à l'application des textes aux activités de l'entreprise.

Ce constat traduit un besoin avéré d'organisation et de vulgarisation de la connaissance juridique dans le domaine de la SST. Ce besoin peut être satisfait par le recours aux modèles et méthodes de l'Intelligence Artificielle, parmi ceux-ci, les concepts d'ontologies et de bases de connaissances ont été retenus.

Dans cette optique, le Centre de recherche sur les Risques et les Crises (CRC) de MINES ParisTech, en partenariat avec l'entreprise Preventeo a lancé un projet de recherche sur les apports des ontologies pour le management des données et des connaissances à caractère juridique dans le domaine de la santé et sécurité au travail.

Cet article, composé de trois parties, décrit une méthodologie et un logiciel associé pour garantir la maîtrise des conformités légales. La première partie identifie les éléments de définition ainsi que les bénéfices avérés des ontologies et des bases de connaissances. Ensuite, l'article détaille la démarche de conception de l'ontologie et son association avec un outil d'édition de bases de connaissances spécialement développé. Enfin, un exemple concret de formalisation des connaissances réglementaires est présenté en définissant les apports et limites de l'ingénierie développée.

2. ELEMENTS DE DEFINITION

Un rapide état de l'art des pratiques en matière de systèmes d'information pour la maîtrise des conformités réglementaires en SST révèle l'usage de différentes solutions technologiques allant de la simple base de données proposant des textes légaux et le commentaire de ceux-ci jusqu'à des systèmes interactifs accessibles en ligne proposant par exemple un suivi de la veille réglementaire, un processus d'audit de la conformité et la gestion de plan d'actions. Les progrès technologiques en la matière sont évidents, mais l'essentiel reste la maîtrise du texte de loi et de son contenu. En France, le législateur à travers la base de données LEGIFRANCE publie sur Internet l'ensemble des textes du Journal Officiel de la République Française. Ce site Internet, propose, en libre accès, plusieurs types de textes : les actes à caractères normatif, les actes résultant des engagements internationaux de la France, la jurisprudence ainsi que l'ensemble des publications officielles. Dans le domaine de la SST, à ce jour ce site rassemble plus de 200 000 textes organisés par grands thèmes. Ces textes sont accessibles par un moteur de recherche. Ce site est extrêmement important pour suivre l'actualité juridique de la SST. Pourtant, l'accès à la base de données reste réservé à des initiés. Afin, de l'ouvrir au plus grand nombre des préventeurs en entreprise, il apparaît que le concept d'ontologie peut apporter une réelle valeur ajoutée.

2.1 Ontologie : définitions

Le concept d'ontologies est né de l'Intelligence Artificielle, le but est de proposer un langage commun entre un opérateur (ici le préventeur), un expert de la connaissance (un juriste) et l'outil informatique. L'ambition est de permettre l'automatisation de certaines tâches faisant appel à la représentation cognitive des connaissances d'un domaine. La définition la plus citée d'une ontologie est celle de Tom Gruber [5] : "An ontology is an explicit specification of a conceptualization". Il est à noter que de nombreux auteurs lui reprochent son caractère trop vague et générique. C'est pourquoi ce travail de recherche utilise la définition proposée par Studer (et al) [6] qui ajoute la notion de partage : "An ontology is a *formal, explicit specification of a shared conceptualization*".

Cette définition souligne quatre caractères particulièrement importants des ontologies :

- Formel : Définit une standardisation de cette représentation permettant l'utilisation dans un système informatique.
- Explicite : Les concepts ainsi que les contraintes liées sont définis de façon déclarative.
- Conceptualisation : Ceci rappelle qu'une ontologie n'est qu'une représentation de la réalité, une abstraction du monde réel. Ainsi, il est à proscrire toute ambiguïté dans la définition des termes.

- Partage : Met l'accent sur le fait qu'une ontologie permet un partage des connaissances d'un domaine donné.

Les apports des ontologies sont multiples. Elles ont pour ambition de clarifier la structure de connaissances en définissant les relations entre les concepts du domaine. Elles permettent de réduire les ambiguïtés conceptuelles terminologiques au travers de la définition des concepts proches dans un domaine donné. Les ontologies participent au partage des connaissances, à la formalisation des savoirs pour ainsi comprendre plus aisément le vocabulaire d'un domaine et les différences entre les termes.

Une ontologie est généralement composée de différentes entités : Les « concepts » représentent une abstraction de la réalité au travers d'une « classe ». Les « relations » traduisent une liaison entre deux concepts. Les « axiomes » considérés comme des informations permettent la définition d'un concept. Enfin, les « instances » sont des « individus » de l'ontologie, c'est-à-dire les membres des différentes « classes ».

2.2 Méthode retenue pour concevoir l'ontologie

Malgré une certaine maturité du concept au sein des systèmes d'informations, il n'existe aujourd'hui pas de « textbook » ou manuel de conception des ontologies. Ainsi, il est possible de lister une dizaine de méthodes de création et de suivi des ontologies. Selon le degré de formalisation nécessaire, l'opérationnalisation de l'ontologie ou encore son internationalisation, chaque approche offre des avantages et des limites souvent conditionnés par l'objectif fixé à l'ontologie à créer. La méthode retenue pour ce travail de recherche est celle proposée par l'Université de Stanford [7] proposant une approche structurée en sept étapes :

- 1 : Définir le domaine et la portée
- 2 : Réutiliser des ontologies existantes
- 3 : Définir les termes importants
- 4 : Créer les classes et la hiérarchie
- 5 : Créer les propriétés et les attributs
- 6 : Définir les facettes
- 7 : Créer les instances

Il est aussi possible d'utiliser des outils permettant la mise en œuvre des différentes méthodes. L'outil retenu est Protégé® développé par l'Université de Stanford. Le choix de cet outil s'appuie sur la taille de la communauté d'utilisateurs ainsi que les possibilités d'adaptations de l'outil selon les différents langages de formalisation. Les différents outils utilisent des langages parfois exclusifs mais le langage qui semble le plus adéquate est le langage OWL (Ontology Web Language) qui, comme son nom l'indique se positionne dans une optique "web" de la formalisation des connaissances. Ce langage correspond à un « mélange » entre le langage RDF (Ressource Description Format) et le langage XML (eXtensible Markup Language), tous les deux sont très largement utilisés au sein des applications Web.

3. PROPOSITION D'UNE ONTOLOGIE COUPLEE A UN SYSTEME D'AUDIT DE LA CONFORMITE REGLEMENTAIRE

La création d'une ontologie de la réglementation SST, outre l'effort d'organisation de la connaissance, se révèle particulièrement bénéfique dans le cadre d'un couplage avec un outil d'édition de bases de connaissances lui-même relié à un outil de gestion des audits de conformité réglementaire. La suite de l'article, détaille les relations entre ces trois composants

intégrés au sein d'un système d'information de management global des risques dénommé Preventeo.

3.1 Présentation de l'ontologie SST

L'ontologie créée relève du domaine de la réglementation française en santé et sécurité au travail. Elle traite plus particulièrement des obligations des « agents », à savoir les employeurs, les chefs d'établissements et plus largement les délégués de pouvoir. Des recherches complémentaires ont permis de repérer une ontologie de niveau « supérieur », c'est-à-dire classant des concepts de haut niveau, utile à la définition de la base de connaissances. L'ontologie LKIF (Legal Knowledge Interchange Format) a donc été retenue. LKIF se compose de 155 classes, 97 propriétés de classes, 266 annotations et sa portée se concentre sur les concepts primitifs du droit. Elle ne propose aucun individu, aucune instance (ontologie dite de « top-level »). Un travail a donc été conduit afin de constituer un corpus traduisant la réglementation française. Les termes candidats ont été extraits à l'aide de la méthode Yatea [8] ainsi que par la conduite de nombreuses interviews d'experts du domaine. La construction de l'ontologie SST s'est réalisée sur une profondeur allant de trois à dix niveaux, ce qui a conduit à la définition de 200 termes organisés en deux grandes catégories : « Agent » et « Action ». Les figures 1 et 2 illustrent deux extractions de l'ontologie centrées sur les concepts « Agent » et « Action ». Ces deux branches de l'ontologie étant utilisées principalement pour la définition des exigences réglementaires.



Fig 1 : Présentation d'un extrait de l'ontologie réalisée : Agent

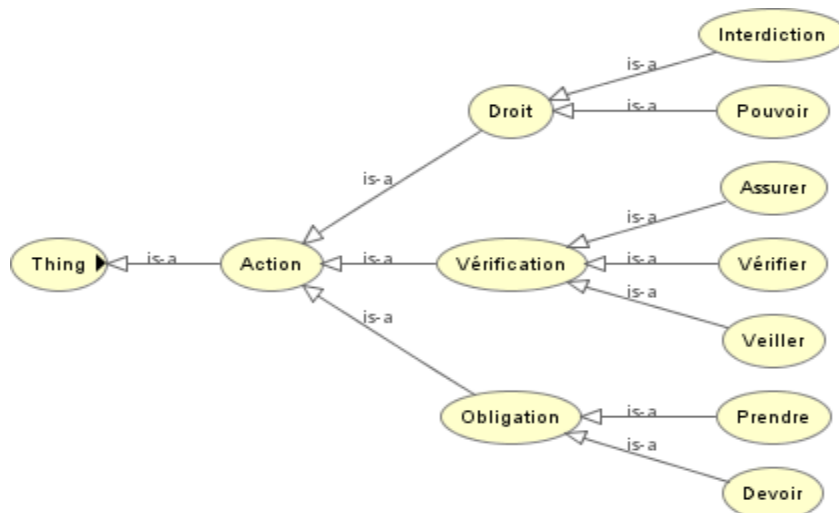


Fig 2 : Présentation d'un extrait de l'ontologie réalisée : Action

A noter que dans ces extractions, les instances de classes ne sont pas représentées.

3.2 Présentation de l'éditeur de base de connaissances : Cogniteo

Afin d'exploiter l'ontologie SST, un outil d'édition de bases de connaissances a été conçu pour la formalisation et l'usage de questionnaires d'audit réglementaire : l'outil Cogniteo. Cogniteo est un outil accessible en ligne. Il guide un utilisateur selon un parcours en plusieurs étapes permettant de créer des questionnaires utilisables en ligne :

- La première étape consiste à organiser un corpus de textes réglementaires au sein d'une base de données. Chaque texte est décrit par un ensemble de descripteurs (titre, date, objet...).
- Chaque texte est ensuite scindé en une série d'articles, chaque article correspondant à un point clef du texte réglementaire (étape 2). Les articles sont à leur tour stockés et décrits dans une base de données.
- Les articles sont « réduits » (étape 3) à un « atome » élémentaire de connaissance l'exigence (l'exigence est représentée sous la forme d'un verbe complété d'une précision).
- Les exigences permettent de construire les questionnaires d'audit.

L'usage de Cogniteo est particulièrement intuitif. Il offre une série de fonctions qui permettent d'organiser les questionnaires en thèmes et sous-thèmes et de revoir aisément l'ordre des questions. Les mises à jour sont possibles. Chaque question est accompagnée de « sa » réponse et une aide question en facilite la compréhension.

3.3 Réalisation d'un prototype d'outil pour repérer automatiquement des exigences réglementaires

Le couplage entre l'ontologie et l'éditeur de bases de connaissances Cogniteo est réalisé à l'aide d'un prototype d'outil de repérage automatique des exigences réglementaires présentes dans les textes réglementaires. Le prototype vise à améliorer la productivité et la qualité du processus de repérage d'exigences réglementaires qui se révèlent particulièrement coûteux.

Dans un premier temps, l'outil « découpe » le texte en phrases, il réalise ensuite diverses requêtes sur l'ontologie SST afin de récupérer les termes utiles considérés comme générateur d'exigences (Ontologie : Agent) et vérifie la présence des termes liés aux obligations

(Ontologie : Action). Pour chaque phrase du texte, si un couple "Agent", ainsi qu'un des termes "Action" qui lui est lié est découvert, la phrase est détectée et sauvegardée en base de données. Une fois la lecture automatique du texte effectuée, la liste des propositions d'exigences est notifiée à l'utilisateur. Les termes candidats sont classés selon trois catégories : Direction, Personnel ou Action. Il en est de même pour les exigences qui ont été repérées.

Cette opération réalisée, la liste est reprise par l'outil Cogniteo afin qu'un expert juridique accepte, précise et complète les propositions faites.

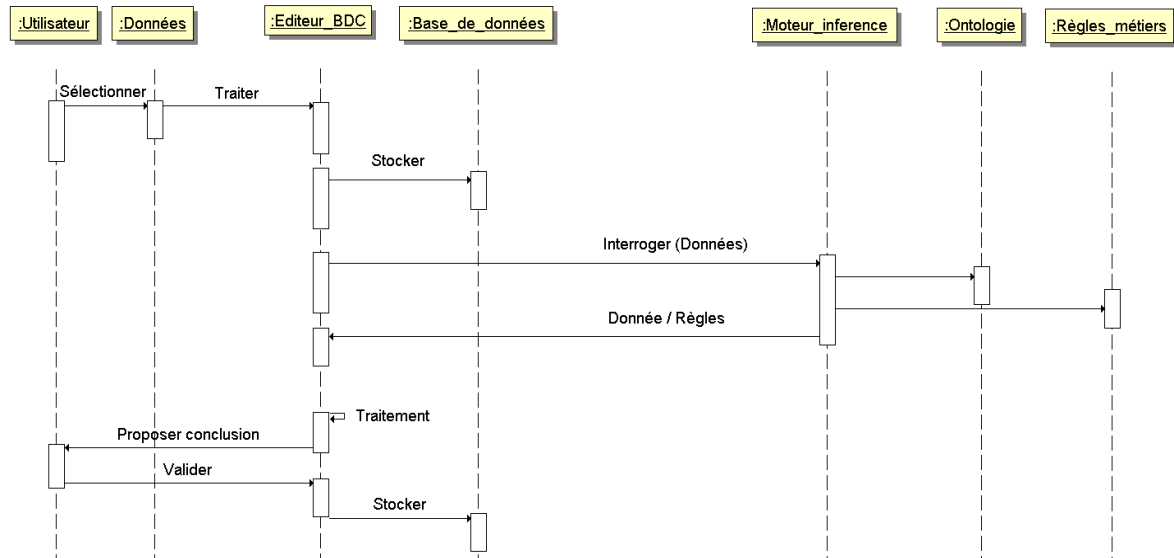


Fig 3 : Processus de création d'un questionnaire semi-automatisé

4. CONDUITE D'UNE EXPERIMENTATION ET DISCUSSION DES RESULTATS

Afin de valider le couplage entre l'ontologie et l'éditeur de bases de connaissances, une expérimentation a été réalisée. Elle repose sur un principe simple, un texte juridique (« Violence et Harcèlement au travail ») [9] a été « simultanément » analysé par un expert juridique et le prototype de repérage automatique des exigences réglementaires.

4.1 Premiers résultats

Le texte « Violence et Harcèlement au travail » correspond aux articles 3 à 5 de l'accord n° 2010-01. Il est constitué de neuf pages. Après lecture, l'expert juridique a repéré un total de vingt six exigences pour l'ensemble des deux textes. De son côté, le prototype logiciel a dénombré vingt exigences. L'expert juridique a réalisé la tâche en 50 minutes. Il aura fallu 5 secondes au prototype pour livrer son résultat. Notons cependant que l'expert livre une exigence formulée avec exactitude, alors que le prototype propose un extrait de phrase « candidat » qui sera forcément repris et précisé par le juriste afin de devenir à terme une exigence mise en base de données.

4.2 Axes de progrès

Les premiers résultats de l'expérimentation sont encourageants, mais ils sont encore loin d'offrir une réelle assistance aux juristes chargés de concevoir les questionnaires d'audit.

Deux axes principaux de progrès sont à considérer.

Le premier vise à amélioration et à l'élargissement du domaine couvert par l'ontologie. L'ontologie réalisée à ce jour est limitée à un sous domaine de la SST, dans le cas présent, la maîtrise des conformités réglementaires. Ce sous domaine définit l'ensemble des actions obligatoires que doivent réaliser les « agents ». Le caractère incomplet de l'ontologie conduit naturellement à occulter des exigences « candidates » qui seraient en lien avec d'autres domaines. Il conviendrait, afin de proposer une reconnaissance plus étendue, de traiter différents domaines dans l'ontologie tels que le risque chimique et la pénibilité au travail par exemple.

Le second axe d'amélioration porte sur la performance du prototype d'outil de repérage automatique des exigences. Au-delà de la qualité de l'ontologie, certaines exigences pourraient voir leur précision accrue via une meilleure analyse. Le prototype se limite à une proposition d'un extrait candidat, ainsi, certaines exigences peuvent être « oubliées » étant donné le formalisme d'un extrait : les tableaux ainsi que les schémas ne peuvent actuellement être exploités par le prototype. Si les schémas, ou images, requièrent un traitement complexe et difficile à mettre en œuvre dans le cadre de ce travail, il est possible de mettre en place certaines fonctionnalités permettant de gérer les tableaux. Les textes sont insérés en base de données au format HTML, des travaux en vue de modifier le découpage proposé sont en cours afin de prendre en compte ce type de format.

5. CONCLUSION

Les travaux sont encore au stade de prototype, mais les premiers résultats sont encourageants. Le logiciel de repérage automatique se doit d'être encore très largement perfectionné. Ce travail est étroitement relié à l'enrichissement de l'ontologie de la SST. Il est aussi prévu d'élargir l'usage du concept d'ontologie à d'autres domaines de la maîtrise des risques. Ainsi, des travaux conduits dans le champ de la sécurité industrielle [10] pourraient être avantageusement mobilisés. L'éditeur de base de connaissances Cogniteo est quant à lui utilisé au quotidien par les juristes et les ingénieurs hygiène, sécurité et environnement de la société Preventeo dans le cadre de la mise à jour et de la conception de questionnaires d'audit [11]. Dans le domaine de la SST, il permet ainsi aujourd'hui de gérer plus 4 000 exigences réglementaires et chaque mois près de deux cents sont rajoutées.

References

[1] Audiffren T., Rallo JM., Guarnieri F., Martin C. "A quantitative analysis of health, safety and environment policy in France, ESREL Conference 2013, Amsterdam, Netherlands, September 2013.

[2] Guarnieri F, Miotti H, Martin C, Besnard D, Rallo JM, Occupational safety and health in France : Practitioners and policy, Edited by AFNOR Group, November 2010.

[3] Audiffren T, Rallo, JM, Guarnieri F, The contribution of case law to compliance management in Occupational Health and Safety (OHS) in France, ESREL Conference 2012, Helsinki, Finland, June 2012.

- [4] Cambon J, Guarnieri F, Groeneweg J, Towards a new tool for measuring Safety Management Systems performance, Proceedings 2nd Symposium on Resilience Engineering, 2006, 2nd Symposium on Resilience Engineering (2006-11-08), Edited by Presses des Mines (France).
- [5] Gruber T. A translation approach to portable ontologies. "Knowledge Acquisition", 5(2):199-220, 1993.
- [6] Studer, Rudi, V. Richard Benjamins, and Dieter Fensel. "Knowledge Engineering : Principles and Methods." *Data & Knowledge Engineering* 25 (1-2) (March) : 161-197.
- [7] Natalya F. Noy , Deborah L. McGuinness : *Ontology Development - 101: A Guide to Creating Your First Ontology*, 2001
- [8] Sophie Aubin, Thierry Hamon : *Improving term extraction with terminological resources*, 2006
- [9] Accord n° 2010-01 du 26 mars 2010 sur le harcèlement et la violence au travail
- [10] Bourreau L, Audiffren T, Rallo JM, Guarnieri F, *Industrial maintenance in the gas sector : the contribution of knowledge bases to compliance evaluation*, ESREL Conference 2012, Helsinki, Finland, June 2012.
- [11] Vigneron J. *Contribution des ontologies à la création de bases de connaissances pour la maîtrise des conformités réglementaires en santé, sécurité au travail et environnement*. Ecole Nationale Supérieure des Mines de Paris, Décembre 2013.



APPORTS DES ONTOLOGIES À LA CRÉATION DE BASES DE CONNAISSANCES POUR LA MAÎTRISE DES CONFORMITÉS LÉGALES EN SANTÉ ET SÉCURITÉ AU TRAVAIL

Mots-clés : ontologie, base de connaissances, conformités réglementaires, santé et sécurité au travail

Résumé

La "Santé-Sécurité au Travail" (SST) est au cœur des politiques d'entreprise et la judiciarisation de la société concourt à une inflation du nombre de textes réglementaires publiés chaque année. Les préventeurs sont donc confrontés au traitement croissant de données afin de rester en conformité. Ils se forment au droit et s'entourent de compétences interne (direction juridique) ou externe (bureaux de contrôle et de conseil) et se dotent de systèmes d'information. Ces derniers sont principalement des bases de données de «veille juridique» proposant des textes réglementaires accompagnés de commentaires. L'ensemble de ces données demeure complexe à interpréter et à exploiter en raison du nombre croissant de textes, de l'expertise nécessaire à fin de les comprendre et de la difficulté à s'assurer qu'ils concernent les activités de l'entreprise. Ce besoin peut être satisfait par le recours aux modèles et méthodes de l'intelligence artificielle. Parmi ceux-ci, les concepts d'ontologies et de bases de connaissances ont été retenus. Une ontologie est un modèle d'un domaine particulier de connaissances réalisé dans un but précis.

Jonathan VIGNERON
MINES ParisTech
PREVENTEO

Jean-Marc RALLO
PREVENTEO

Franck GUARNIERI
MINES ParisTech
PREVENTEO

PREVENTEO
6 rue de l'Abreuvement
06110 Le Cannet, France

MINES ParisTech
CRC - Centre de recherche sur les Risques et les Crises
rue Claude Daunesse, CS 10207
06904 Sophia Antipolis Cedex, France

