

Analyse sémantique des recommandations médicales pour la génération des règles SWRL

Samia Sbissi*, Mariem Mahfoudh**,***
Said Gattoufi*

*Laboratoire SMART, Université de Tunis, Tunisie
samia.sbissi@gmail.com, algattoufi@yahoo.com

**Laboratoire MIRACL, Université de Sfax, Tunisie

***Université de Kairouan, Kairouan, Tunisie
mariem.mahfoudh@gmail.com

Résumé. Les guides de bonnes pratiques sont des documents de référence en évolution qui contiennent des recommandations et des règles visant à aider les professionnels à maîtriser un domaine médical. Dans notre travail, nous sommes intéressés par l'utilisation de ces documents afin d'aider les médecins cardiologues à prendre des décisions sur les soins de santé appropriés pour les patients à risque de maladie cardiovasculaire. Plus précisément, notre papier propose une approche automatique qui analyse et transforme le texte (recommandations médicales) en OWL DL (Web Ontology Language - Description Logic) et des règles SWRL (Semantic Web Rule Language). Pour analyser le texte, nous avons utilisé une ontologie du domaine cardiovasculaire et des outils de traitement du langage naturel (TAL). Notre travail est original en ce qu'il propose une transformation automatique des textes en règles SWRL alors que les travaux associés portent uniquement sur la transformation de texte en des axiomes OWL légers.

1 Introduction

Les documents et les ressources médicales connaissent, de plus en plus, une croissance exponentielle et des changements simultanés. Ceci complique la tâche des experts en médecine dans leur suivi des nouveautés médicales pour la prise de bonnes décisions. Ainsi, et dans le but d'automatiser le processus d'analyse de ces documents et les transformer en connaissances, les techniques de la fouille de textes et d'analyse sémantique sont fortement nécessaires. Noter que l'analyse sémantique est un processus permettant de produire une représentation formelle d'un texte écrit en langage naturel (Kiryakov et al., 2004). Cette formalisation puisse être utilisée par exemple dans le raisonnement et l'inférence ou également pour déterminer si deux textes sont en relation d'implication (Gyawali et al., 2017). Dans notre travail qui se place dans le cadre d'une coopération entre l'hôpital La Rabta de Tunis et le laboratoire SMART (Strategies for Modelling

and ARTificial inTelligence), nous sommes intéressés à assister les médecins cardiologues dans la prise des décisions au niveau des traitements des patients qui souffrent d'une dissection aortique. Une dissection aortique est une déchirure partielle de la paroi de l'aorte pouvant à tout moment évoluer vers la rupture complète, avec comme conséquences le décès (Criado, 2011). Il s'agit donc d'une urgence absolue dans son diagnostic et dans sa prise en charge. Alors, afin d'élaborer le système d'aide à la décision, nous avons besoin, bien entendu, de modéliser les connaissances relatives à cette maladie. Ces connaissances sont généralement décrites dans des documents spécialisés appelés les guides de bonnes pratiques. Elles sont constituées d'un ensemble de règles et de recommandations (ex. *In patients with abdominal aortic diameter of 25-29 mm, new ultrasound imaging should be considered 4 years later*). L'objectif de cet article étant donc de présenter notamment notre travail d'analyse sémantique et la transformation des recommandations en règles SWRL (semantic web rule language). Ex. $Patient(?p) \wedge hasAbdominalDiameter(?p, ?d) \wedge swrlb : greaterThan(?d, 25) \wedge swrlb : lessThan(?d, 29) \rightarrow recommendedDiagnosis(?p, "ultrasoundImaging")$). SWRL est un langage de règles intégré directement dans OWL (Ontology Web Language). Il permet de définir des règles sous forme d'implications logiques entre conditions et conclusions. Nous pensons que l'utilisation des techniques de web sémantique et des mécanismes d'inférence des ontologies pourrait être une bonne aide pour bien élaborer notre système d'assistance médical. Ainsi, afin d'analyser le texte, nous avons utilisé une ontologie existante du domaine cardiovasculaire et des outils de traitement du langage naturel. En effet, l'analyse sémantique des textes a été bien étudiée dans la littérature [(Gyawali et al., 2017), (Petrucci et al., 2016)]. Cependant, l'ensemble de ces travaux se sont focalisés sur une transformation en formules logiques simples ou en des ontologies légères. Nous nous intéressons plutôt à produire des règles SWRL.

Le reste de l'article sera comme suit : la section 2 présente quelques travaux connexes. La section 3 détaille notre approche proposée. La section 4 discute les résultats obtenus. La section 5 conclut ce travail et donne quelques perspectives.

2 État de l'art

L'extraction des connaissances à partir du texte s'est trouvée au coeur de plusieurs domaines de recherche vu l'augmentation vertigineuse, d'année en année, du volume de données disponibles sous forme de corpus de textes (Upadhyay et Fujii, 2016; Ristoski et Paulheim, 2016). Une étude comparative des approches et d'outils d'extraction des connaissances est présentée dans (Gangemi, 2013). Nous nous intéressons dans cette section à présenter essentiellement des travaux qui ont étudié la transformation du texte en formules logiques (analyse sémantique).

(Park et Lee, 2007) ont proposé une méthode semi-automatique pour extraire des règles de documents Web. Leur méthode nécessite, en entrée, une ontologie de domaine existante et une sélection manuelle des pages Web pertinentes. L'approche n'utilise que des techniques simples de TAL, ce qui l'empêche de gérer des textes complexes. (Gangemi et al., 2017) ont proposé un système appelé FRED qui extrait les relations n-aires sur la base de structures de représentation du discours et les transforme ensuite en représentation RDF (Resource Description Framework). Une méthode présentée

par (Petrucci et al., 2016), permettant de convertir des textes sous forme d’axiome de logique de description. (Gyawali et al., 2017) ont proposé une méthode qui convertit des documents techniques d’Airbus en des axiomes DL à l’aide des grammaires d’arbres adjoints.

Dans notre travail, nous nous intéressons non seulement à la transformation de texte en logique de description (axiomes OWL) mais également à la génération automatique de règles SWRL. Nous pensons que l’analyse d’un texte composé d’un ensemble de recommandations sous forme de règles SWRL est plus utile pour élaborer notre système d’aide à la décision.

3 Approche proposée

Notre approche d’analyse sémantique des recommandations médicales est constituée de trois grandes étapes détaillées ci-dessous.

1. Pré-traitement. Le pré-traitement de notre texte repose sur les méthodes de tokenisation, la lemmatisation, l’analyse syntaxique, etc. Cette phase vise à nettoyer notre corpus de texte des mots inutiles (de bruit). Notre corpus (voir tableau 1) est composé d’un ensemble de recommandations médicales définies par la Société européenne de cardiologie ESC.

In complicated Type B AD, surgery may be considered.
In patients with acute contained rupture of TAA, urgent repair is recommended.
In complicated Type B AD, TEVAR is recommended.

TAB. 1 – *Quelques règles du corpus.*

2. Annotation sémantique. Il s’agit d’annoter l’ensemble de recommandations par l’ontologie de domaine cardiovasculaire CVDO¹. Le choix de l’ontologie est fait après une recherche sur les ontologies existantes (essentiellement via le portail bioportal²) qui nous a amené à conclure que l’ontologie CVDO est la plus adaptée à nos besoins et couvre le plus notre domaine étudié (maladies cardiovasculaires). Ceci est dit, l’ontologie CVDO manque de relations et sera par la suite enrichie par le résultat de l’analyse sémantique. Ainsi, le processus d’annotation sémantique prend le texte pré-traité et l’ontologie CVDO, puis essaie de trouver les correspondances entre eux. Chaque terme du texte peut être associé à un ou plusieurs concepts de l’ontologie (classes, propriétés d’objet, propriétés de données, etc.). Dans notre travail, nous nous sommes intéressés à trois types de correspondances (Mahfoudh et al., 2016) : 1) correspondance syntaxique en utilisant la distance Levenshtein ; 2) correspondance morphologique en utilisant la lemmatisation et 3) correspondance sémantique en utilisant l’ontologie WordNet.

1. <http://purl.bioontology.org/ontology/CVDO>
2. <https://bioportal.bioontology.org>

3. Analyse sémantique. Pour la transformation du texte en règles SWRL, nous avons défini un ensemble de patrons selon la nature de la recommandation médicale. La transformation est basée sur le résultat de l'annotation sémantique. Nous identifions pour chaque ligne de recommandation le pattern ayant la même structure et les classer ensemble. Un exemple de pattern est présenté ci dessous.

Patron1 : règles avec expressions de classe.

if the patient has a disease related to heart disease.

↓

Patron1 : Patient(?p) ∧ (hasRelatedDisease min 1 RelatedDiseases)(?p)
 -> RelatedDiseaseHistory (?p , true)

Chaque règle SWRL est composée de deux parties : antécédent et conséquence. Pour analyser le corpus des recommandations, nous commençons par la segmentation (chunking en anglais). Il s'agit de grouper les mots selon leur étiquetage morpho-syntaxique comme illustré dans la figure 1.

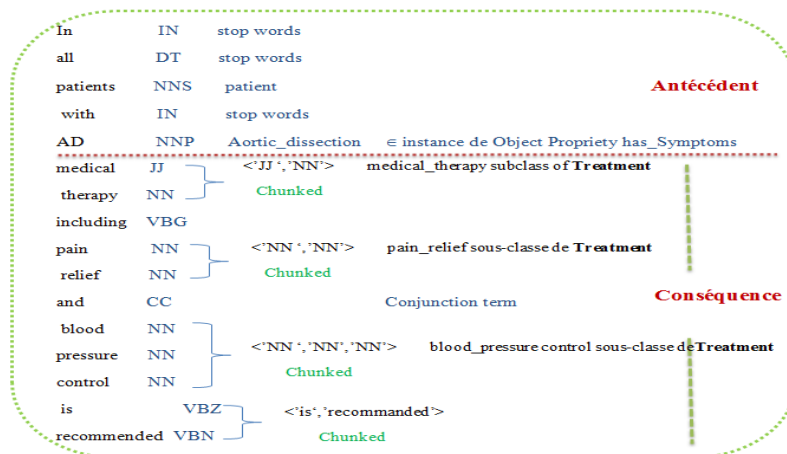
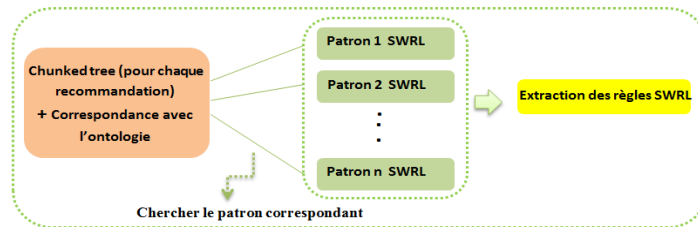


FIG. 1 – Correspondances entre composants.

Ensuite, il faut générer l'arbre correspondante de la recommandation qui sera automatiquement convertie en règle SWRL en cherchant sa correspondance avec les patrons prédéfinis (figure 2).

4 Expérimentations

Afin d'établir une correspondance entre l'ontologie et le texte, nous avons développé un prototype avec le langage Java. Les correspondances morphologiques et la lemmatisation sont assurées par l'API Stanford CoreNLP. La bibliothèque NLTK de

FIG. 2 – *Extraction des règles SWRL.*

Python est utilisée pour le pré-traitement du texte. Nous avons obtenu 30% de correspondance entre l'ontologie CVDO et les textes de recommandations médicales. Ceci reflète que l'ontologie CVDO couvre une partie de texte mais elle manque aussi d'autres qui doivent être ajoutés à la fin de l'analyse sémantique. Pour la construction des patterns et la transformation des règles, nous avons utilisé le framework Jena et l'API SWRLAPI. 60 % des règles SWRL sont correctement extraites du corpus de recommandations. Dans certains cas, une recommandation peut correspondre à un ou deux patterns simultanément, ce qui a réduit le taux de précision.

5 Conclusion

Nous avons présenté, dans cet article, une méthode d'analyse sémantique des recommandations médicales avec une ontologie de domaine. Notre approche consiste à transformer ces recommandations en règles SWRL. L'analyse sémantique est menée par une annotation sémantique du texte de recommandation par les concepts de l'ontologie. Nous avons combiné l'utilisation de l'ontologie OWL et des techniques de TAL pour capturer des connaissances et les représenter sous forme de règles dans le langage SWRL.

Nous rappelons que notre objectif était d'élaborer un système d'aide médicale permettant aux médecins de prendre des décisions concernant les maladies aortiques. Nous prévoyons, dans un futur travail, d'incorporer le résultat de l'analyse sémantique dans l'ontologie CVDO afin de l'enrichir et l'exploiter dans des tâches d'inférence. Nous nous baserons sur un travail précédent sur l'enrichissement et le peuplement d'ontologie (Mahfoudh et al., 2015). Une fois enrichie, il faut interroger cette ontologie afin de répondre, via des interfaces IHM adaptées, aux requêtes des médecins et leur assister dans leur prise de décision.

Références

- Criado, F. J. (2011). Aortic dissection : a 250-year perspective. *Texas Heart Institute Journal* 38(6), 694.
- Gangemi, A. (2013). A comparison of knowledge extraction tools for the semantic web. pp. 351–366.

- Gangemi, A., V. Presutti, D. Reforgiato Recupero, A. G. Nuzzolese, F. Draicchio, et M. Mongiovì (2017). Semantic web machine reading with fred. *Semantic Web* 8(6), 873–893.
- Gyawali, B., A. Shimorina, C. Gardent, S. Cruz-Lara, et M. Mahfoudh (2017). Mapping natural language to description logic. In *European Semantic Web Conference*, pp. 273–288. Springer.
- Kiryakov, A., B. Popov, I. Terziev, D. Manov, et D. Ognyanoff (2004). Semantic annotation, indexing, and retrieval. *Web Semantics : Science, Services and Agents on the World Wide Web* 2(1), 49–79.
- Mahfoudh, M., G. Forestier, et M. Hassenforder (2016). A benchmark for ontologies merging assessment. In *International Conference on Knowledge Science, Engineering and Management*, pp. 555–566. Springer.
- Mahfoudh, M., G. Forestier, L. Thiry, et M. Hassenforder (2015). Algebraic graph transformations for formalizing ontology changes and evolving ontologies. *Knowledge-Based Systems* 73, 212–226.
- Park, S. et J. K. Lee (2007). Rule identification using ontology while acquiring rules from web pages. *International Journal of Human-Computer Studies* 65(7), 659–673.
- Petrucci, G., C. Ghidini, et M. Rospocher (2016). Ontology learning in the deep. In *European Knowledge Acquisition Workshop*, pp. 480–495. Springer.
- Ristoski, P. et H. Paulheim (2016). Semantic web in data mining and knowledge discovery : A comprehensive survey. *Web semantics : science, services and agents on the World Wide Web* 36, 1–22.
- Upadhyay, R. et A. Fujii (2016). Semantic knowledge extraction from research documents. In *Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on*, pp. 439–445. IEEE.

Summary

Clinical practice guideline is an evolving reference document that contains recommendations and knowledge which aims to assist professionals to master a medical domain. In our work, we are interested in using these documents in order to assist doctors to make decisions about appropriate health care for patients who are at risk of cardiovascular disease. More precisely, our paper proposes an automatic approach that parses and transforms text (clinical practice guideline) into OWL DL (Ontology Language Web Description Logic) axioms and SWRL (semantic web rule language) rules. To parse the text, we have used an existing ontology of cardiovascular domain and natural language processing tools (NLP). Our work is original in that studies the mapping between text and SWRL rules while the related work are focused only on mapping text on OWL lightweight axioms.