

L'analyse des dommages de voitures à l'aide de la reconnaissance des entités nommées et de l'ontologie

Hamid Ahaggach^{*,**}, Lylia Abrouk^{*}, Eric Lebon^{**}

* Laboratoire d'informatique de Bourgogne
Université de Bourgogne Franche-Comté, France
prenom.nom@u-bourgogne.fr

** Syartec, Aix en Provence, France
elebon@syartec.com

La gestion des transports de voiture est une tâche complexe dans le domaine de vente d'automobile. Différents processus doivent être mis en place pour le suivi et le partage de dommages causés lors du transport tels que la prise de photos, la rédaction de rapports. La majorité des rapports d'assurance existants sont des documents textuels non structurés, parfois manuscrits et non numériques. De plus, il n'y a pas de norme pour décrire les dommages, ce qui rend le partage des rapports d'assurance entre les concessionnaires automobiles et les compagnies d'assurance difficile. Dans ce contexte, les ontologies sont utilisées pour la modélisation des dommages dans différents domaines (Rachman et Chandima Ratnayake, 2018; Hamdan et al., 2019; Everett et al., 2002). Dans le domaine automobile, les travaux existants (Barrachina et al., 2012) (Dardailler, 2012) visent à proposer des ontologies pour la modélisation des accidents de la circulation pour décrire les circonstances, la localisation, les causes et les effets de l'accident et ne s'intéressent pas à la modélisation des dommages causés par l'accident sur les voitures. Dans cet article, nous proposons une approche pour la construction et le peuplement d'une ontologie de domaine. Les contributions de notre travail peuvent être résumées comme suit : (1) La construction d'une ontologie pour modéliser la description des dommages en se basant sur les connaissances des experts et des rapports d'assurance. (2) La proposition d'un modèle d'extraction d'information basée sur l'approche neuronale pour la reconnaissance d'entités nommées en utilisant des données textuelles étiquetées. (3) Le peuplement de l'ontologie de domaine avec les informations extraites. (4) L'évaluation de notre modèle sur un jeu de données réel.

Construction d'ontologie : Nous proposons une ontologie de domaine basée sur les connaissances des experts et leurs rapports d'assurance pour modéliser les avaries des voitures. Dans cette ontologie, la classe *Voiture* a plusieurs sous-classes qui contiennent les composantes de la voiture. Il existe des composantes qui ont également des sous-classes, par exemple, une roue de voiture contient plusieurs sous-classes (roulement de roue, jantes, pneus, fixations de roue). Ces composantes sont liées avec la classe *Avarie* qui définit le type et la sévérité de l'avarie, ainsi que de la partie de la voiture endommagée. Nous avons défini avec les experts du domaine 9 types d'avaries (Bosse, Casse, Coincement, Déchirure, Écrasement, Enfoncement, Manque, Perforation, Rayure). L'ontologie a été validée par les experts, et nous avons vérifié sa consistance et sa cohérence en utilisant les raisonneurs Pellet et HermiT.

Extraction d'information : Nous proposons une approche d'extraction d'information composée de quatre modules principaux : (1) le module de prétraitement des données pour la vérification d'orthographe. (2) le module de traitement automatique du langage (TAL) en appliquant les techniques suivantes : la lemmatisation, la tokenisation et la suppression des mots vides. (3) le module de l'extraction de l'information où nous avons utilisé la reconnaissance des entités nommées (REN) pour extraire les entités (concept) et les attributs (propriété de données). (4) le module de peuplement de l'ontologie en utilisant les packages *Owlready* et *Python-skos* (Lamy, 2017), qui fournissent une grande variété de méthodes pour traiter les ontologies, en particulier pour l'insertion d'instances dans l'ontologie.

Notre expérimentation est réalisée sur un jeu de données étiqueté qui contient 500 rapports pour extraire 4 types d'entités (Composante, Avarie, Marque, Modèle) avec apprentissage par transfert des modèles REN basés sur les réseaux de neurones les plus utilisés (Conditional Random Fields, Long Short-Term Memory Bidirectional, Bidirectional Encoder Representations from Transformers, et le modèle REN de SpaCy). Pour évaluer et comparer ces modèles, nous avons utilisé les mesures d'évaluation Rappel, Précision et F1-Score. Les résultats obtenus par la plupart des modèles sont satisfaisants. BILSTM-CRF donne de meilleurs résultats dans la reconnaissance des avaries, tandis que CRF est puissant pour reconnaître les entités *Marque*, *Modèle* et *Composante*. Ces résultats s'expliquent par le fait que les entités ne dépendent pas du contexte. Les autres modèles basés sur les transformateurs comme BERT et le modèle REN de SpaCy donnent des résultats acceptables, mais moins performants que les CRF.

Références

- Barrachina, J., P. Garrido, M. Fogue, F. J. Martinez, J.-C. Cano, C. T. Calafate, et P. Manzoni (2012). Caova : A car accident ontology for vanets. In *2012 IEEE wireless communications and networking conference (WCNC)*, pp. 1864–1869. Ieee.
- Dardailler, D. (2012). Road accident ontology. <https://www.w3.org/2012/06/rao.html>. Accessed : 2022-12-05.
- Everett, J. O., D. G. Bobrow, R. Stolle, R. Crouch, V. de Paiva, C. Condoravdi, M. van den Berg, et L. Polanyi (2002). Making ontologies work for resolving redundancies across documents. *Communications of the ACM* 45(2), 55–60.
- Hamdan, A.-H., M. Bonduel, et R. J. Scherer (2019). An ontological model for the representation of damage to constructions. In *CEUR Workshop Proceedings*, Volume 2389, pp. 64–77. CEUR Workshop Proceedings.
- Lamy, J.-B. (2017). Owlready : Ontology-oriented programming in python with automatic classification and high level constructs for biomedical ontologies. *Artificial intelligence in medicine* 80, 11–28.
- Rachman, A. et R. Chandima Ratnayake (2018). Ontology-based semantic modeling for automated identification of damage mechanisms in process plants. In *Working Conference on Virtual Enterprises*, pp. 457–466. Springer.