

Sujet Stage Master 2

Titre : Enrichissement semi-automatique d'ontologie à partir de textes basé sur la découverte de patrons: Application au domaine de la pêche.

Mots clés : Enrichissement d'ontologie à partir de textes, Construction semi-automatique, Ontologie Core, TAL, Apprentissage automatique, Text2onto, Domaine de la pêche.

Encadrants :

Mounira HARZALLAH, Equipe Duke, LINA, Nantes

Giuseppe BERIO, Equipe Expression, IRISA, Vannes

Durée : 6 mois

Indemnité : 508,2€/mois

Compétences requises. Des compétences en TAL, en apprentissage automatique ou en web sémantique sont souhaitées

Lieu du stage : LINA, Nantes

Poursuite en thèse en relation avec le sujet est envisageable

Contact. Pour candidater, il faut envoyer votre CV et les relevés des notes des deux dernières années à Mounira Harzallah mounira.harzallah@univ-nantes.fr

Contexte

L'objectif du projet KIFANLO (Projet Financé par la fondation de France) est d'analyser des stratégies de pêche des marins. Ces stratégies, exprimées dans des vocabulaires différents, sont à définir à l'aide des cartes cognitives et à expliciter leur sémantique à l'aide d'une ontologie de la pêche que nous avons à développer. Le domaine de la pêche comme plusieurs autres domaines, se caractérise par un besoin prépondérant d'exploiter une masse importante d'information produite par plusieurs experts. Nous avons donc opté pour une approche semi-automatique pour la construction de cette ontologie. Actuellement, nous avons développé une « ontologie core » de la pêche qui représente les sous-domaines principaux de la pêche et les relations entre eux. Nous avons aussi adapté l'outil Text2Onto à la construction d'ontologie à partir de textes en français. Nous avons ensuite utilisé cet outil pour extraire des concepts, des relations de spécialisation ou des relations ad-hoc que nous avons intégrés à la main à l'ontologie core de la pêche. A ce stade, l'ontologie obtenue est le résultat d'une première itération de l'enrichissement de l'ontologie core. Plusieurs itérations sont à réaliser à partir d'autres documents. Cependant, l'enrichissement manuel n'est plus envisageable car il serait trop coûteux, long et fastidieux.

Des méthodes et outils ont été développés afin de pouvoir extraire et organiser les connaissances explicites et/ou implicites, contenues dans le texte. Ils utilisent généralement des techniques spécifiques issues de domaines de recherches différents, notamment le TAL, l'apprentissage automatique, le text mining. Cependant, il n'y a pas encore une méthode/un outil qui a prouvé son efficacité pour l'enrichissement semi-automatique d'ontologie à partir de textes.

Objectif.

Le sujet de stage porte sur le développement d'une démarche/outil pour l'enrichissement semi-automatique d'ontologie à partir des ressources textuelles. Nous nous intéressons à l'utilisation des techniques d'analyse syntaxique profonde de texte et des techniques d'apprentissage pour découvrir des nouveaux patrons d'extraction (d'un texte) des relations déjà présentes dans une ontologie existante afin de l'enrichir avec des nouveaux concepts.

Travail à réaliser

Dans ce stage, il s'agit :

1. de faire une étude bibliographique sur les méthode et outils existants pour l'apprentissage de patrons à partir de textes.
2. de proposer une méthode utilisant une technique d'analyse syntaxique profonde et une approche par apprentissage pour découvrir des patrons pour l'extraction à partir de textes, de relations (core) de l'ontologie existante et l'utilisation de ces relations pour l'enrichissement de cette ontologie.
3. d'implémenter cette méthode en intégrant un analyseur syntaxique profond, un outil d'apprentissage et l'outil Text2Onto de construction d'ontologie à partir de textes.
4. d'appliquer le résultat pour l'enrichissement de l'ontologie de la pêche dans le projet KIFANLO

Références

- Buitelaar, P., Cimiano, P., and Magnini, B. (2005). *Ontology Learning from Text: Methods, Applications and Evaluation*, chapter *Ontology learning from text: an overview*, pages 3–12. IOS Press.
- Chulyadyo R., M. Harzallah and G. Berio, *Core Ontology Based Approach for Treating the Flatness of Automatically Built Ontology*, KEOD 2013, Vilamoura, Portugal.
- Cimiano P. & Volker J. (2005). *Text2onto - a framework for ontology learning and data-driven change discovery*. In A. MONTOMOY, R. MUNOZ & E. MÉTAIS, Eds., 2nd Eur. Semantic Web Conf., volume 3513, p. 227–238.
- Deprès, S., and Szulman S. (2007) *Merging of legal micro-ontologies from European Directives*. In *Artificial Intelligence and Law 15*, 187–200.

- Ghadfi S., N. Béchet, G. Berio Building Ontologies from Textual Resources: A Pattern Based Improvement Using Deep Linguistic Information. In Proceedings of the 5th Workshop on Ontology and Semantic Web Patterns (WOP2014), Riva del Garda, Italy, October 19, 2014. Edited by: Victor de Boer, Aldo Gangemi, Krzysztof Janowicz, Agnieszka Lawrynowicz. CEUR-WS Voll1302, November 2014, pp14-25.
- Gherasim, T., M. Harzallah, G. Berio, P. Kuntz (2013). Methods and Tools for Automatic Construction of Ontologies from Textual Resources: A Framework for Comparison and Its Application. *Advances in Knowledge Discovery and Management* 3, 177-201.
- Lehmann J., J. Völker; Perspectives on Ontology Learning. Volume 18 of [Studies on the Semantic Web](#). IOS Press, 2014.
- Snow R., D. Jurafsky, Andrew Y. Ng. 2004. "*Learning syntactic patterns for automatic hypernym discovery*". In NIPS 2004.
- Zouaq A., D. Gasevic, M. Hatala. 2011. "*Towards open ontology learning and filtering*". *Information Systems*, v.36 n.7, p.1064-1081, November, 2011.