

Titre :

Sujet de Travail d'Étude et de Recherche – Développement d'un plugin Protégé pour l'extraction, la prédiction et l'édition de métadonnées pour des ontologies.

Information :

Encadrants : Clément Jonquet (LIRMM, UM) – jonquet@lirmm.fr
Spécialités : Master DECOL, AIGLE, autres
Nombre d'étudiants : 2-3
Contexte: Projet SIFR (www.lirmm.fr/sifr) et AgroPortal (<http://agroportal.lirmm.fr>)
Ou: LIRMM, Montpellier
Quand: 2nd semestre 2018-2019

Mots clés :

Développement d'application bureau Java, Plug-in Protégé, métadonnées d'ontologies, web sémantique.

Technologies :

Java, langages du web sémantique (RDF/OWL/SKOS), OWL-API

Description :

Une ontologie est une représentation formelle des concepts, relations et règles d'un domaine. Il existe plusieurs langages du web sémantique pour encoder les ontologies : RDFS, SKOS, OWL, OBO, etc. Pour permettre d'ordonner, identifier, réutiliser les ontologies, nous avons besoins de métadonnées sur ces ontologies les plus précises possible. Ces propriétés ne sont en général pas assez remplies par les développeurs d'ontologies et notre objectif est de développer une interface dans leur outil d'édition qui leur facilite la tâche.

Ce TER consiste à concevoir et implémenter un plugin pour l'application Protégé (<https://protege.stanford.edu>) qui permettra l'édition, l'extraction et la prédiction de métadonnées. Protégé est un logiciel open source pour éditer des ontologies construit sur le modèle ouvert et extensible (type Eclipse) sous forme d'application bureau Java et de plugins additionnels. C'est le logiciel de développement d'ontologie le plus utilisé dans le monde. Chaque fois qu'une fonctionnalité doit être ajoutée (raisonnement, visualisation, connexion à des outils externes, etc.) un plugin est créé par la communauté. Dans ce TER nous nous intéressons à décrire les métadonnées des ontologies en cours d'édition dans le logiciel avec des vocabulaires standards du web sémantique.

Pour une ontologie donnée (fichier OWL, RDFS ou SKOS), le plugin devra extraire les propriétés de métadonnées de cette ontologie (nom, description, dates, contributeurs, langage, format, etc.) à partir du fichier et en prédire d'autres (langue naturelle, mot clés, etc.). De plus, l'interface graphique permettra de valider les extractions/prédictions et d'éditer à la main d'autres champs de métadonnées. Les métadonnées pourront ensuite être sérialisées dans le fichier de l'ontologie produit par Protégé ou exportées suivant un profil de métadonnées spécifique.

Au LIRMM, nous travaillons sur un profil d'application appelé « MOD » pour la description de ressources sémantiques en général (ontologies, vocabulaires, terminologies, etc.). Dans ce profil, nous avons recensé et regroupé 346 propriétés (127 une fois regroupées) pour décrire les ontologies. Nous utiliserons MOD comme schéma directeur pour ce travail.

Le plugin sera développé en Java. Le code sera hébergé et disponible sur GitHub et une page web (wiki) descriptive de documentation du plugin sera réalisée.

Références :

Portail de l'application Protégé : <https://protege.stanford.edu/>

Code source de Protégé : <https://github.com/protegeproject/protege>

Protégé plugin library: https://protegewiki.stanford.edu/wiki/Protege_Plugin_Library

Profil MOD : <https://github.com/sifrproject/MOD-Ontology> (version 1.4 bientôt disponible)

Gennari, J. H., Musen, M. A., Ferguson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., ... & Tu, S. W. (2003). The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-computer studies*, 58(1), 89-123.

Dutta, B., Toulet, A., Emonet, V., & Jonquet, C. (2017, November). New Generation Metadata vocabulary for Ontology Description and Publication. In *Research Conference on Metadata and Semantics Research* (pp. 173-185). Springer, Cham.

Jonquet, C., Toulet, A., Dutta, B., & Emonet, V. (2018). Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal. *Journal on Data Semantics*, 1-31.