

Titre :

Développement d'une application web d'extraction et prédiction de métadonnées pour des ontologies.

Information :

Encadrants : Clement Jonquet (LIRMM, UM) – jonquet@lirmm.fr
Vincent Emonet (LIRMM, UM) – emonet@lirmm.fr
Anne Toulet (LIRMM, UM) – anne.toulet@lirmm.fr
Spécialités : Master DECOL, AIGLE, autres
Nombre d'étudiants : 2-3
Contexte: Projet SIFR (Semantic Indexing of French Biomedical Data Resources)
Ou: LIRMM
Quand: 2nd semestre 2015-2016

Mots clés :

Application web, web service (REST), technologies web, ontologies, parseur XML/RDF/OWL, web sémantique.

Technologies :

Java/JEE, REST, Restful web services, XML/JSON, Tomcat, pour le Web service, RubyOnRails, web client technologies (HTML5, JavaScript, CSS), pour l'interface Web.

Résumé :

Ce TER consiste à concevoir et implémenter une application web simple pour l'extraction, la prédiction et la validation de métadonnées pour des ontologies. Pour une ontologie donnée (fichier OWL) l'application devra extraire les propriétés de cette ontologie (nom, description, dates, contributeurs, langage, format, etc.) à partir du fichier et en prédire d'autres (langue naturelle, mot clés, etc.). Différents vocabulaires du web sémantique pour décrire les métadonnées seront gérés (DublinCore, VOAF, DCAT, VOID, etc.) et mappés vers des propriétés communes. Une interface Web permettra à l'utilisateur de valider les propriétés extraites et d'éventuellement les récupérer directement dans son fichier d'origine.

Description détaillée :

Une ontologie est une représentation formelle des concepts, relations et règles d'un domaine. Il existe plusieurs langage 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 précise possible. Ces propriétés ne sont en général pas assez remplies par les développeurs d'ontologies et notre objectif est de développer une application web qui leur facilite la tâche. Quand elles sont remplies, ces ontologies peuvent utiliser différents vocabulaires que nous devons prendre en compte dans cette application.

Pour une ontologie donnée, l'application extraira les propriétés de l'ontologie (nom, description, page de documentation, contributeurs, dates, etc.) à partir du fichier. Lorsque ces informations ne seront pas là, l'application essaiera de les prédire en fonction du contenu de l'ontologie ou en utilisant des informations en ligne. Par exemple, en regardant les labels d'une ontologie, on peut en déduire la langue de l'ontologie. En cherchant dans PubMed, on peut trouver la publication de référence pour une ontologie.

L'application affichera ensuite à l'utilisateur les propriétés extraites, prédites et vides pour lui permettre de les valider, changer ou simplement remplir. Une fois cela effectué, un nouveau fichier de l'ontologie pourra être produit avec les propriétés incluses dans l'ontologie.