

# Une bibliothèque pour la manipulation d'ontologies formelles

Document de travail

23 février 2018

Ce document présente les fonctionnalités à implémenter dans la bibliothèque, ainsi que des pistes de développement orientées vers le *data mining*. Bien que ce projet soit avant tout un projet scolaire, les contributions extérieures sont les bienvenues.

## 1 Introduction

La représentation des connaissances est un domaine de l'Intelligence Artificielle visant à structurer des informations sur le monde pouvant être utilisées dans des applications de haut niveau. Ces représentations peuvent ensuite être visualisées, comparées entre elles ou encore utilisées pour le raisonnement automatique (BRACHMAN & LEVESQUE, 1985).

SPIVAK & KENT (2012) introduisent les «*ontology logs*» (ou «*ologs*»), une nouvelle représentation des connaissances qui se base sur la théorie des catégories pour structurer les ontologies. Celles-ci sont décrites sous forme de types (objets) et d'aspects (flèches entre les objets) et font intervenir les constructions usuelles des catégories : diagrammes commutatifs (égalité entre deux aspects), produits, etc.

La présente bibliothèque vise à proposer une implémentation simple des ologs qui sera par la suite appliquée à la fouille de données.

## 2 Fonctionnalités

La bibliothèque devra proposer les fonctionnalités suivantes :

- représentation, enregistrement et chargement d'ologs ;
- fonctions génériques permettant la création et la modification d'ontologies en OCaml aussi simplement que possible ;
- population d'ontologies simples et vérification automatisée du respect des « faits » déclarés par l'olog ;

- affichage des ontologies dans un format graphique à définir ;

Par la suite, des fonctionnalités avancées pourront être implémentées, telles que l'export vers SQL, la génération de code OCaml (façon *type providers*) pour la « programmation orientée ontologies » ou l'intégration d'un prouveur.

### 3 Rappel sur les ologs

## 4 Application au *data mining*

### 4.1 Littérature

Dans un second temps, la bibliothèque sera appliquée à des problèmes de fouille de données (*data mining*). Bien que leur rôle premier soit de représenter des connaissances structurées par l'utilisateur même, les ontologies ont déjà été appliquées par plusieurs auteurs à différentes tâches de la fouille de données telles la recherche de motifs ou le clustering.

Une application récurrente des ontologies au clustering consiste à raffiner les distances et méthodes de pondération utilisées pour tenir compte d'informations sémantiques et améliorer les performances du clustering, ou au contraire tenir compte de la subjectivité des utilisateurs. Par exemple en clustering de documents textes, HOTH0 et al. (2002) utilisent un graphe d'ontologies (une hétérarchie plutôt qu'une hiérarchie) afin de trouver des concepts adaptés à la description des différents clusters.

## Références

- BRACHMAN, R. J. & LEVESQUE, H. J. (1985). *Readings in knowledge representation*. Morgan Kaufmann Publishers Inc.
- HOTH0, A., MAEDCHE, A. & STAAB, S. (2002). Ontology-based text document clustering. In : *KI* 16.4, p. 48–54.
- SPIVAK, D. I. & KENT, R. E. (2012). Ologs : a categorical framework for knowledge representation. In : *PLoS One* 7.1, e24274.