

Analyse comparative de méthodologies et d'outils de construction automatique d'ontologies à partir de ressources textuelles

Toader Gherasim*, Mounira Harzallah*,
Giuseppe Berio**, Pascale Kuntz *

*LINA, UMR 6241 CNRS
toader.gherasim, mounira.harzallah @univ-nantes.fr, pascale.kuntz@polytech.univ-nantes.fr,
<http://www.lina.univ-nantes.fr>

**LABSTICC, UMR 3192 CNRS
giuseppe.berio@univ-ubs.fr
<http://www.lab-sticc.fr>

Résumé. Plusieurs méthodologies et outils de construction automatique des ontologies à partir de ressources textuelles ont été proposés ces dernières années. Dans cet article nous analysons quatre approches en les comparant à une approche de référence – Methontology. Dans leur sélection nous avons privilégié celles qui couvrent l'ensemble des étapes du processus de construction d'ontologies. Puis nous analysons et comparons la portée, les limites et les performances des implémentations logicielles associées aux approches analysées. Ces outils ont été testés sur un corpus de ressources textuelles, et nous avons comparé leurs résultats à ceux obtenus manuellement.

1 Introduction

Depuis les travaux fondateurs de Gruber (1993), les ontologies jouent un rôle majeur en ingénierie des connaissances, et leur essor, associé à celui du web sémantique, ne cesse de croître. Elles sont maintenant au cœur des systèmes de recherche d'information ou d'aide à la décision de multiples domaines (e.g. médical Osborne et al. (2009), juridique Bourigault et Lame (2002), etc.). Initialement de tailles restreintes et construites entièrement "à la main", elles peuvent contenir aujourd'hui plusieurs milliers de concepts (e.g. Aime et al. (2009)) et des relations variées. Et, si dans la lignée des systèmes experts, leur construction a reposé, et repose encore, sur l'expertise humaine, leur popularisation et le changement d'échelle nécessitent de plus en plus le recours massif à une forme d'automatisation qui limite la contribution des experts. L'accessibilité croissante de nombreuses ressources textuelles qui renferment des connaissances d'une grande richesse sur les concepts et les processus mis en oeuvre dans les applications rend cette aspiration de plus en plus crédible.

Des propositions à la fois méthodologiques et logicielles se sont multipliées ces dernières années pour construire automatiquement des ontologies à partir de corpus textuels (Velardi et al. (2007), Maynard et al. (2009b)). Développées souvent dans des contextes applicatifs, ce