

Enrichissement d'une RTO par l'ajout de termes spécialisés

Soumia Lilia Berrahou^{*,**}, Ludovic Lebras^{*,**}
Patrice Buche^{*,**} Juliette Dibie-Barthélemy^{***} Mathieu Roche^{**}

^{*}INRA - UMR IATE, 2, place Pierre Viala, F-34060 Montpellier Cedex 2, France
Soumialilia.Berrahou@lirmm.fr, Patrice.Buche@supagro.inra.fr

^{**}LIRMM, CNRS, Université Montpellier 2, France
ludovic.lebras@gmail.com, Mathieu.Roche@lirmm.fr

^{***}INRA - Mét@risk-AgroParisTech, 16 rue Claude Bernard, F-75231 Paris Cedex 5, France
Juliette.Dibie@agroparistech.fr

Résumé. Nous proposons dans cet article une méthode d'enrichissement d'une Ressource Termino-Ontologique (RTO) par l'ajout de termes extraits d'un corpus de documents textuels constitué dans le domaine d'application décrit par la RTO. Une RTO est une ressource comportant une composante conceptuelle (l'ontologie) et une composante terminologique (la terminologie), dans laquelle les termes sont distingués des concepts qu'ils dénotent. Nous nous intéressons à l'enrichissement d'une RTO permettant de modéliser des relations n-aires entre des données quantitatives expérimentales, où les arguments peuvent être des concepts symboliques ou des quantités caractérisées par des unités de mesure. La méthode proposée consiste à enrichir la terminologie associée aux concepts symboliques et la terminologie associée aux unités de mesure par de nouveaux termes extraits du corpus.

1 Introduction

Le travail présenté dans cet article se situe dans le cadre de l'identification de termes candidats spécialisés à partir de données textuelles pour enrichir une Ressource Termino-Ontologique (RTO). La notion de RTO provient des travaux de (Reymonet et al., 2007), (Roche et al., 2009), (McCrae et al., 2011) et de (Cimiano et al., 2011) qui ont proposé d'associer une partie terminologique et/ou linguistique aux ontologies afin d'établir une distinction claire entre la manifestation linguistique (le terme) et la notion qu'elle dénote (le concept). Nous nous intéressons à l'enrichissement d'une RTO permettant de modéliser des relations n-aires entre des données quantitatives expérimentales (Touhami et al., 2011), où les arguments peuvent être des concepts symboliques ou des quantités caractérisées par des unités de mesure. Etant donné un domaine d'application représenté par une RTO, nous proposons une méthode générique d'enrichissement de la partie terminologique de cette RTO par l'extraction de termes, permettant de dénoter des concepts symboliques ou des unités de mesure, à partir d'un corpus de données textuelles dans le domaine d'application étudié, l'extraction étant guidée par cette RTO. La méthode proposée permet plus précisément d'identifier des termes variants candidats dans un corpus de documents textuels pour enrichir la terminologie associée aux concepts