

Modélisation d'une ressource termino-ontologique de domaine pour l'annotation sémantique de tableaux

Patrice Buche*, Juliette Dibie-Barthélemy**
Liliana Ibanescu**, Abir Saïd**

*INRA - UMR IATE, 2, place Pierre Viala,
F-34060 Montpellier Cedex 2, France
LIRMM, CNRS-UM2, F-34392 Montpellier, France
Patrice.Buche@supagro.inra.fr

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

Résumé. Nous proposons dans cet article une modélisation d'une ressource termino-ontologique (RTO) de domaine, guidée par la tâche d'annotation sémantique de tableaux. L'annotation d'un tableau consiste à annoter ses cellules, pour pouvoir ensuite identifier les concepts représentés par ses colonnes et enfin identifier la ou les relations n -aires qu'il représente. La RTO proposée permet d'une part de modéliser dans sa composante lexicale les termes utilisés pour l'annotation des cellules en intégrant la gestion des synonymes et du multilingue, et, d'autre part, de modéliser dans sa composante conceptuelle les concepts symboliques, les concepts numériques et les relations n -aires, qui sont propres au domaine étudié.

1 Introduction

L'intégration de données permet d'accéder de manière unifiée à des sources multiples, hétérogènes en syntaxe, schéma ou sémantique. Le but de l'intégration de données est de faciliter l'accès et la réutilisation d'un ensemble de sources. La notion centrale sur laquelle reposent les recherches actuelles en intégration sémantique de données est la notion d'ontologie, une ontologie étant une spécification formelle et explicite d'une conceptualisation partagée (Studer et al., 1998). Dans Reymonet et al. (2006), trois paramètres influencent la modélisation d'une ontologie : la tâche à réaliser, le domaine d'intérêt et l'application. Dans notre cas, l'application est la construction d'entrepôts de données thématiques ouverts sur le Web pour l'aide à la décision. Notre domaine d'intérêt est la sécurité alimentaire. Nous aurions pu, pour modéliser notre ontologie, reprendre le thesaurus AGROVOC¹ conçu et maintenu par la FAO et qui définit un vocabulaire contrôlé, multilingue et structuré couvrant tous les domaines de l'agriculture, de la pêche et de l'alimentation. Mais ce thesaurus, à vocation très généraliste, n'est pas assez spécifique pour couvrir le sous-domaine de l'alimentation qui nous intéresse (moins

1. <http://aims.fao.org/website/AGROVOC-Thesaurus/sub>