

Une nouvelle mesure sémantique pour le calcul de la similarité entre deux concepts d'une même ontologie

Emmanuel Blanchard, Mounira Harzallah
Pascale Kuntz, Henri Briand

Laboratoire d'informatique de Nantes Atlantique
Site École polytechnique de l'université de Nantes
rue Christian Pauc
BP 50609 - 44306 Nantes Cedex 3
emmanuel.blanchard@univ-nantes.fr

Résumé. Les ontologies sont au coeur du processus de gestion des connaissances. Différentes mesures sémantiques ont été proposées dans la littérature pour évaluer quantitativement l'importance de la liaison sémantique entre paires de concepts. Cet article propose une synthèse analytique des principales mesures sémantiques basées sur une ontologie modélisée par un graphe et restreinte ici aux liens hiérarchiques is-a. Après avoir mis en évidence différentes limites des mesures actuelles, nous en proposons une nouvelle, la PSS (Proportion of Shared Specificity), qui sans corpus externe, tient compte de la densité des liens dans le graphe reliant deux concepts.

1 Introduction

Associées notamment au succès des nouveaux langages du Web sémantique, les ontologies suscitent un intérêt croissant au sein des communautés de l'ingénierie et de la gestion des connaissances (Gruber, 1993; Fürst, 2004). Cependant, malgré le développement d'outils d'aide à leur manipulation, le développement et l'exploitation des ontologies restent des phases complexes dans un processus global de gestion de connaissances. En amont, une des difficultés majeures concerne la structuration des ensembles de concepts dont la taille ne cesse de croître. Et en aval, le problème consiste à rechercher efficacement des sous-ensembles de concepts à la fois en temps de calcul et en pertinence sémantique des résultats.

Pour faciliter ces tâches, le recours à des mesures sémantiques semble judicieux ; il permet de constituer une « connaissance heuristique » directement exploitable. De façon générale, une mesure sémantique est une application de l'ensemble $\mathcal{C} \times \mathcal{C}$ des paires de concepts d'une ontologie dans \mathbb{R}^+ qui permet d'évaluer quantitativement la proximité ou l'éloignement sémantique de deux concepts. Quelque soit le domaine applicatif, la pertinence de la mesure utilisée est étroitement associée à l'efficacité des algorithmes qui l'intègrent. Cependant, son choix reste un problème délicat. Pour comparer les mesures existantes, plusieurs approches complémentaires sont envisageables (Budanitsky, 1999). L'analyse formelle vise à étudier précisément leurs propriétés à la fois algorithmiques et statistiques. La comparaison avec le jugement humain analyse la corrélation entre les valeurs des mesures et les évaluations subjectives de sujets