

# Sur l'évaluation de la quantité d'information d'un concept dans une taxonomie et la proposition de nouvelles mesures

Emmanuel Blanchard, Mounira Harzallah, Pascale Kuntz et 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  
prénom.nom@univ-nantes.fr  
<http://lina.atlanstic.net/>

**Résumé.** L'évaluation de la similarité entre concepts structurés dans une taxonomie connaît un réel essor lié au développement du web sémantique. En effet, les mesures sémantiques constituent une aide au développement et à l'exploitation des ontologies qui restent des tâches complexes dans un processus globale d'ingénierie des connaissances. Certaines mesures sémantiques sont basées sur la notion clef de contenu informationnel initialement proposé par Resnik (1995). Cet article expose notre vision du contenu informationnel à travers de nouveaux estimateurs indépendants de tout corpus. Nous généralisons la notion de contenu informationnel à un groupe de concepts et discutons la base du logarithme utilisée. Quelques propositions de mesures montrent l'analogie que l'on peut faire pour réutiliser des résultats bien connus sur une représentation ensembliste. Nous mettons en évidence la pertinence de notre approche par le biais de résultats statistiques.

## 1 Introduction

Le problème de l'évaluation de la similarité sémantique qui a été largement étudié en psychologie (Tversky, 1977; Rosch, 1975) connaît un nouvel essor lié au développement du web sémantique. Dans les années 70 beaucoup de recherches en classification ont été influencées par la théorie selon laquelle les classes sur un ensemble d'objets sont organisées dans une taxonomie grâce à un processus d'abstraction. La description des connaissances sur un domaine par une hiérarchie ou un graphe de concepts, appelé ontologie, est un problème central pour de nombreux systèmes développés à l'heure actuelle en ingénierie des connaissances.

Une ontologie est une spécification explicite et formelle d'une conceptualisation partagée, dont les primitives de base sont les concepts et les relations entre ces concepts (Gruber, 1993; Guarino, 1995). Associées au succès des nouveaux langages du Web sémantique (Bechhofer et al., 2004), les ontologies suscitent un intérêt croissant au sein des communautés de l'ingénierie et de la gestion des connaissances (e.g. Gruber (1993), Fürst (2004)). Cependant, malgré l'apparition d'outils d'aide à leur manipulation, leur développement et leur exploitation restent