

Acquisition de la théorie ontologique d'un système d'extraction d'information

Alain-Pierre Manine*

*LIPN, Université Paris 13/CNRS UMR7030
99 ave. Jean-Baptiste Clément
F93430 Villetaneuse
manine@lipn.univ-paris13.fr

Résumé. La conception de systèmes d'Extraction d'Information (EI) destinés à extraire les réseaux d'interactions géniques décrits dans la littérature scientifique est un enjeu important. De tels systèmes nécessitent des représentations sophistiquées, s'appuyant sur des ontologies, afin de définir différentes relations biologiques, ainsi que les dépendances récursives qu'elles présentent entre elles. Cependant, l'acquisition de ces dépendances n'est pas possible avec les techniques d'apprentissage automatique actuellement employées en EI, car ces dernières ne gèrent pas la récursivité. Afin de palier ces limitations, nous présentons une application à l'EI de la Programmation Logique Inductive, en mode multi-predicats. Nos expérimentations, effectuées sur un corpus bactérien, conduisent à un rappel global de 67.7% pour une précision de 75.5%.

1 Introduction

La modélisation des interactions géniques présente un considérable intérêt scientifique pour les biologistes ; pourtant, la majeure partie de la connaissance la concernant n'est pas présente dans des banques de données génomiques, mais dans la littérature scientifique. De fait, de nombreux travaux (ex. Craven et Kumlien (1999); Krallinger et al. (2007)) ont été entrepris afin de concevoir des systèmes d'Extraction d'Information (EI) visant à extraire un réseau d'interactions géniques à partir de la bibliographie. Dans la plupart de ces systèmes, des patrons d'extraction permettent l'extraction d'une *unique* relation d'interaction binaire (ex. Saric et al. (2005)). De tels modèles ne rendent pas compte de la complexité des données biologiques, telles que les voies métaboliques. En effet, l'EI nécessite des représentations complexes, fondées sur des ontologies, et impliquant de multiples relations interdépendantes (Berardi et Malerba (2006)), éventuellement récursives.

Afin de modéliser ce type de connaissances, Manine et al. (2008) ont récemment introduit une architecture dans laquelle l'EI est considérée comme une tâche de *population d'ontologie*¹. Dans ce contexte, la théorie logique de l'ontologie subsume les patrons d'extraction, et le problème de l'apprentissage de patrons devient alors une tâche d'*apprentissage d'ontologie*².

¹Ontology Population

²Ontology Learning