

## **mycoMINE : un système de fouille de texte sémantique dédié à la recherche de biocarburants de seconde génération**

Marie-Jean Meurs, Bahar Sateli, Caitlin Murphy, René Witte, Greg Butler, Justin Powlowski and Adrian Tsang  
Concordia University

80ème Congrès de l'Association francophone pour le savoir (ACFAS)  
7-11 Mai 2012, Montréal, QC, Canada

Dans la course pour remplacer les combustibles fossiles par des biocarburants renouvelables, la découverte d'enzymes dégradant efficacement la lignocellulose est un défi majeur. Pour identifier ces enzymes, les biologistes rassemblent les résultats expérimentaux publiés dans une littérature scientifique dont volume et diversité des sources sont en expansion constante. Analyser la littérature nécessite une lecture approfondie de toutes les publications pertinentes mais cette tâche est difficilement exhaustive et des informations cruciales peuvent échapper au lecteur.

Combinant outils de fouille de texte sémantique et ressources ontologiques, notre système mycoMINE extrait automatiquement de la littérature les connaissances concernant les enzymes fongiques. mycoMINE est intégré dans les outils des utilisateurs (navigateurs, éditeurs de textes, wikis) via la plateforme Semantic Assistants. Les performances de mycoMINE sont évaluées sur un corpus, librement accessible, d'articles scientifiques manuellement annotés par des biologistes experts. Les évaluations menées sur les quatre entités les plus fréquentes - organisme, enzyme, pH et température - extraites automatiquement montrent que plus de 90% des entités présentes sont trouvées par mycoMINE.

mycoMINE, développé pour le projet Genozymes, est le premier système de fouille de texte sémantique dédié à la recherche d'enzymes fongiques impliquées dans le développement des biocarburants de deuxième génération.