

BELUGA : un outil pour l'analyse dynamique des connaissances de la littérature scientifique d'un domaine. Première application au cas des maladies à prions.

Nicolas Turenne*, Marc Barbier**

*Biométrie et Intelligence Artificielle (BIA), INRA
78 352 Jouy-en-Josas
turenne@jouy.inra.fr

**Systèmes Agraires et Développement (SAD), INRA
78 850 Thiverval-Grignon
barbier@grignon.inra.fr

Résumé. Un projet ciblé sur l'étude du domaine des maladies à prions a permis de formaliser une méthodologie commune, sociologique et informatique, de compréhension de sa dynamique par l'analyse thématique. Nous avons créé une plate forme d'indexation de notices bibliographiques dont le but est d'extraire des associations évoluant à travers des intervalles de temps. Beluga propose une chaîne de traitement basée sur l'indexation des documents en unités de base: références, auteurs, termes simples et composés, organismes. L'outil est fondé sur une double approche d'apprentissage et de visualisation qui automatise les processus d'extraction de groupes d'auteurs et de termes, et permet à l'utilisateur de revenir aux données documentaires sources. L'analyse diachronique de corpus de documents électroniques nous permet d'analyser comment la terminologie est structurée en thématiques en nous ramenant au problème de détection des tendances thématiques émergentes.

Mots-clés: fouille de textes; extraction terminologique, traitement de corpus; extraction de connaissances à partir de corpus; détection de thématiques émergentes; maladies à prions ; scientométrie.

1. Introduction

L'utilisation par les chercheurs en sciences sociales d'outil de fouille, d'extraction de connaissances et d'aide à l'interprétation est d'actualité pour caractériser les dynamiques d'un certain nombre « d'affaires », dont le dossier maladies à prions [Chateauraynaud et Torny, 1999]. Notre travail s'inscrit dans cette perspective¹. Dans le dossier des maladies à prions, la complexité des dynamiques sociales est forte puisque de nombreux mondes professionnels sont impliqués et en interactions. L'une des « couches » d'information importante est bien sûr celle constituée par les bases de données scientifiques.

Concernant les maladies à prions, les travaux scientifiques sur la dynamique de ce domaine sont, en France, finalement peu nombreux au regard de l'importance des crises sanitaires liées à la vache-folle. [Barbier et De Looze, 1999], [Maunoury et al., 1999]. Le

¹ *L'expertise du Comité interministériel sur les ESST à la frontière entre recherche et décision publique* (resp. scientifique: Marc Barbier, INRA), GIS Prions AO 2001.