

---

# Une approche symbolique pour l'extraction d'informations dans un corpus restreint de textes médicaux

Nicolas Hiot\*<sup>1</sup>

<sup>1</sup>Laboratoire d'Informatique Fondamentale d'Orléans – Université d'Orléans : EA4022, Institut National des Sciences Appliquées - Centre Val de Loire : EA4022, Institut National des Sciences Appliquées – France

## Résumé

Le domaine médical représente un challenge important pour la tâche d'extraction d'informations. Il existe un grand nombre de documents textuels comme des comptes rendus ou des descriptions d'effets secondaires. Il est très difficile pour les membres du corps médical et pour les chercheurs de trouver facilement l'information souhaitée. Nous travaillons sur l'extraction automatique des informations dans ces textes dans le but de construire des bases de connaissances plus faciles à interroger et sur lesquelles il sera possible de raisonner. Ces textes diffèrent des tâches d'extraction habituelles par leur richesse en informations et leur langage technique et propre au domaine. À l'heure où l'état de l'art se concentre sur l'apprentissage automatique et les réseaux de neurones, nous proposons une approche symbolique qui nous permet d'obtenir des résultats corrects pour un corpus restreint. En effet, il est très difficile de constituer un corpus de grande taille du domaine médical à cause de la diversité des textes et de leur disponibilité réduite causée par les données sensibles et personnelles qu'ils contiennent.

---

\*Intervenant