

# Analyse en rôles sémantiques pour le résumé automatique

Elyase Lassouli\*, Yasmine Mesbahi\*  
Camille Pradel\* Damien Sileo\*,\*\*

\*Synapse Développement  
5 rue du Moulin Bayard, 31 000 Toulouse  
<http://www.synapse-developpement.fr/>  
\*\*IRIT, Université Toulouse 3  
118 Route de Narbonne, 31 062 Toulouse  
<https://www.irit.fr>

**Résumé.** Cet article présente une approche visant à extraire les informations exprimées dans un corpus de textes et en produire un résumé. Plusieurs variantes de méthodes extractives de résumé de texte ont été implémentées et évaluées. Leur principale originalité réside dans l'exploitation de structures appelées CDS (pour *Clause Description Structure*) issues d'un composant d'annotation en rôles sémantiques et non directement des phrases composant les textes. Le résumé obtenu est un sous-ensemble des CDS issus du corpus d'origine ; ce format permettra dans la suite la détection d'incohérences textuelles. Dans ce travail, nous retransformons les CDS résumés en texte pour permettre la comparaison de notre approche avec celles de la littérature. Les premiers résultats sont très encourageants : les variantes que nous proposons obtiennent généralement de meilleurs scores que des implémentations de méthodes de référence.

## 1 Introduction

Sur le Web, l'utilisateur est démuné face à de très grands volumes de documents de qualité assez inégale et d'une fiabilité parfois douteuse. Le résumé automatique peut permettre aux humains de mieux appréhender ces données surabondantes. Notre objectif est d'utiliser le résumer automatique de corpus pour construire une base de connaissances fiable permettant l'identification d'incohérences dans un nouveau texte. Cette approche est illustrée dans la figure 1 : elle consiste à extraire d'une grande quantité de textes les faits redondants en se basant sur l'idée qu'ils représentent des connaissances consensuelles et donc aptes à permettre de façon fiable l'identification d'incohérences dans un nouveau texte. En plus d'un début de modèle de détection d'incohérence, ce travail peut être vu comme une contribution à la tâche de résumé automatique de corpus de textes ; dans cet article, nous le présentons et l'évaluons comme tel.

La principale originalité de l'approche réside dans l'exploitation de structures appelées CDS (pour *Clause Description Structure*) issues d'un composant propriétaire produisant des

---

Ce travail a été financé dans le cadre du projet DGA-RAPID *Détection d'Incohérences Textuelles* n162906126