

# Traçabilité de l'information, de l'extraction à l'exploitation

Claire Laudy\*, Charlotte Jacobé De Naurois\*, Bénédicte Goujon\*

\*THALES, 1 avenue Agustin Fresnel, 91767 Palaiseau, France  
<prénom>.<nom>@thalesgroup.com

## 1 Une chaîne intégrée de l'extraction à l'exploitation

Afin de fournir un support à la prise de décision, nous proposons une chaîne fonctionnelle permettant d'extraire des informations à partir de textes et de les agréger au sein d'un réseau d'informations sémantiques. Pour illustrer notre approche, nous proposons un exemple d'extraction d'informations précises (noms de composants chimiques et valeurs sur les propriétés associées) à partir d'articles scientifiques. Dans cet exemple, nous nous concentrons sur l'extraction et la fusion d'information concernant la molécule ABS/ZnO et ses propriétés.

**Phrase1** : *The tensile strength for ABS/ZnO line samples were 23.3, 24.19, and 28.24 MPa for the infill density of 50%, 75%, and 100%, respectively.*

**Phrase2** : *The tensile strength for ABS/ZnO rectilinear samples were 20.21, 20.32, and 22.19 MPa for the infill density of 50%, 75%, and 100%.*

La première étape d'extraction d'informations est réalisée par le module d'annotation de la plateforme STRASS (fig. 1). La plateforme STRASS est centrée sur l'apprentissage de patrons linguistiques à partir de textes annotés manuellement par un expert métier. Elle vise l'annotation automatique de textes et l'export des informations extraites (Goujon (2021)). En sortie de ce module, des graphes composés de nœuds entités isolés ou liés par une relation sont générés.



FIG. 1 – Annotation de la Phrase1 de notre exemple illustratif.

Cet ensemble de petits graphes est agrégé par fusion avec InSyTo. InSyTo est une bibliothèque d'algorithmes de manipulation de graphes conceptuels qui peuvent être combinés afin de fournir des fonctions avancées (Laudy (2017)). En sortie de ce second module, l'ensemble des informations extraites des textes initiaux est agrégé au sein d'un graphe d'informations.