

# Détection de citations

**Atila Kaan Alkan<sup>\*,†</sup>, Faïza Belbachir<sup>\*</sup>, Ben Kabongo<sup>⊗</sup>**

<sup>\*</sup> Université Paris-Saclay, CNRS, Laboratoire interdisciplinaire des sciences du numérique, 91405 Orsay, France

<sup>†</sup> IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

<sup>\*</sup> Institut Polytechnique des Sciences Avancées (IPSA), Ivry-sur-Seine

<sup>⊗</sup> Sorbonne Université

HackaTAL 2023  
(CORIA-TALN)

# Plan

1. Introduction
2. Détection de citations
3. Conclusion et perspectives

# Plan

1. Introduction
2. Détection de citations
3. Conclusion et perspectives

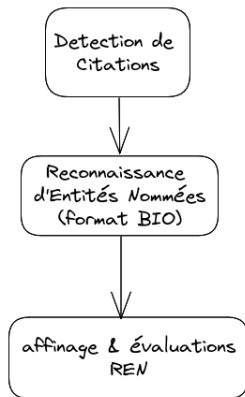


Figure: Approche proposée pour la détection de citations.

# Expériences et résultats

Modèle	Jeu de test		
	P	R	F
PerceptronTagger	0,590	0,370	0,440 (NA)
CRF	0,270	0,160	0,190 (NA)
SciBERT	0,394	0,550	<b>0,460</b> (0.014)
astroBERT	0,350	0,470	0,390 (0.02)

**Table:** Impact du modèle d'apprentissage sur une tâche de détection de citations (ramenée à une tâche de détection d'entités nommées). Pour chaque configuration, nous affichons la moyenne (et l'écart type) des 5 affinages réalisés. Pour des raisons de visibilité, nous indiquons l'écart type uniquement pour la F-mesure.

# La distribution des classes peut expliquer les faibles performances

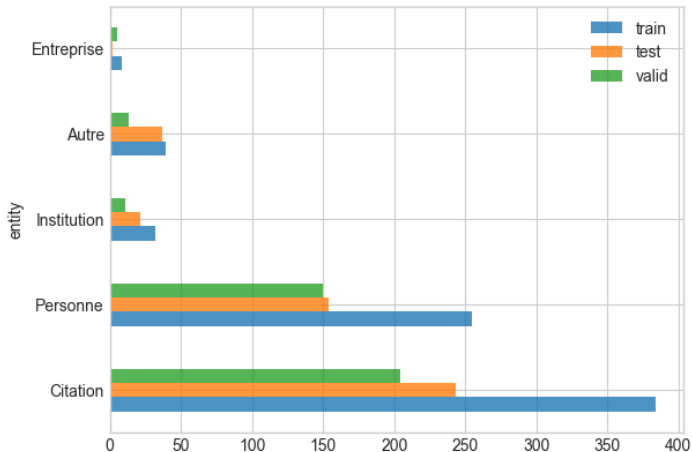


Figure: Distribution des classes du corpus.

# Plan

1. Introduction
2. Détection de citations
3. Conclusion et perspectives

# Conclusion et perspectives

- **Contributions :**
  - **Détection de citations :**
    - Comparaison de différents modèles d'apprentissage ;
    - Le corpus de pré-entraînement semble avoir peu d'influence (dans nos expériences) ;
- **Perspectives :**
  - **Détection de citations et extraction de relations :**
    - Distinction discours direct et indirect ;
    - Ajout de tokens spéciaux autour des portions annotées pour aider le modèle ;
    - Explorer des approches hybrides (règles et apprentissage) ;
    - Arbres de dépendances syntaxiques.
  - **Coréférences :**
    - Si plusieurs auteurs et citations, repérer les mentions faisant références à une même entité peut s'avérer utile ;
  - **Analyse et méthodes d'évaluation :**
    - Analyse plus fine des erreurs avec des métriques adaptées : *Slot Error Rate* (SER).