

Hackatal 2023 : Détection de citations

Équipe FlowBERT
Kévin Deturck, Qi WANG

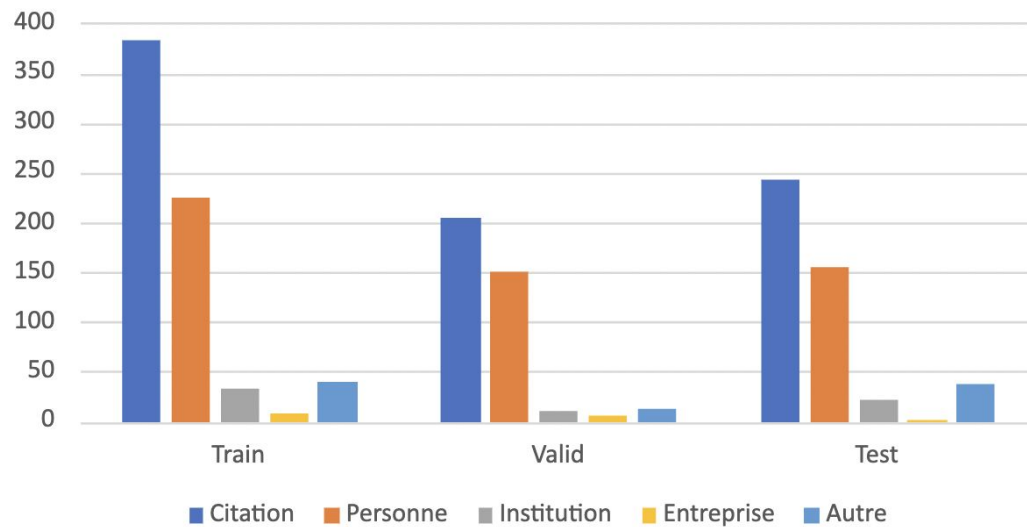


Tâches

- Détection de citations (discours rapportés direct et indirect)
- Détection de source (entité qui a produit initialement le discours)
- Détection du couple (discours, source) comme relation entre les deux détections précédentes

Données

Data distribution





Méthode

1. Détection des entités : segment + catégorie

1. **Modèle de langue** “transformer” *flaubert-large-cased*
2. **Affinage** (« Fine-tuning ») sur « **train** » + « **dev** »
3. **Application** sur “**test**”



Méthode

2. Détection des relations citation-source

→ Par dépendances syntaxiques

1. liste des **têtes syntaxiques** des tokens **d'une citation**
2. séq. de tokens "**source**" > **parcours** de **l'arbre syntaxique**
3. voir si **dépendance** à une tête syntaxique de la **citation**
4. si **dépendance** > **relation** citation-source



Résultats

Version	F-Mesure	Précision	Rappel
Que les entités avec une relation	0,58	0,85	0,44
Idem + Toutes citat.	0,66	0,84	0,54

Merci pour votre attention !