ChêneTAL

Plateforme d'expérimentation sur des outils de traitement automatique des langues et d'intelligence artificielle

Enjeux

- → Chaînes de traitements hétérogènes et paramétrables
  - — Outils d'IA (+ opaques)
  - — Approches TAL (+ paramétrables)
- → Partage de ressources et réutilisabilité :
  - — Données d'entraînement
  - — Ressources linguistiques
  - — Outils TAL et IA
- → Reproductibilité de la recherche
- → Valorisation des résultats et des développements
- → Recherche collaborative
- → Formation et acculturation

Outils

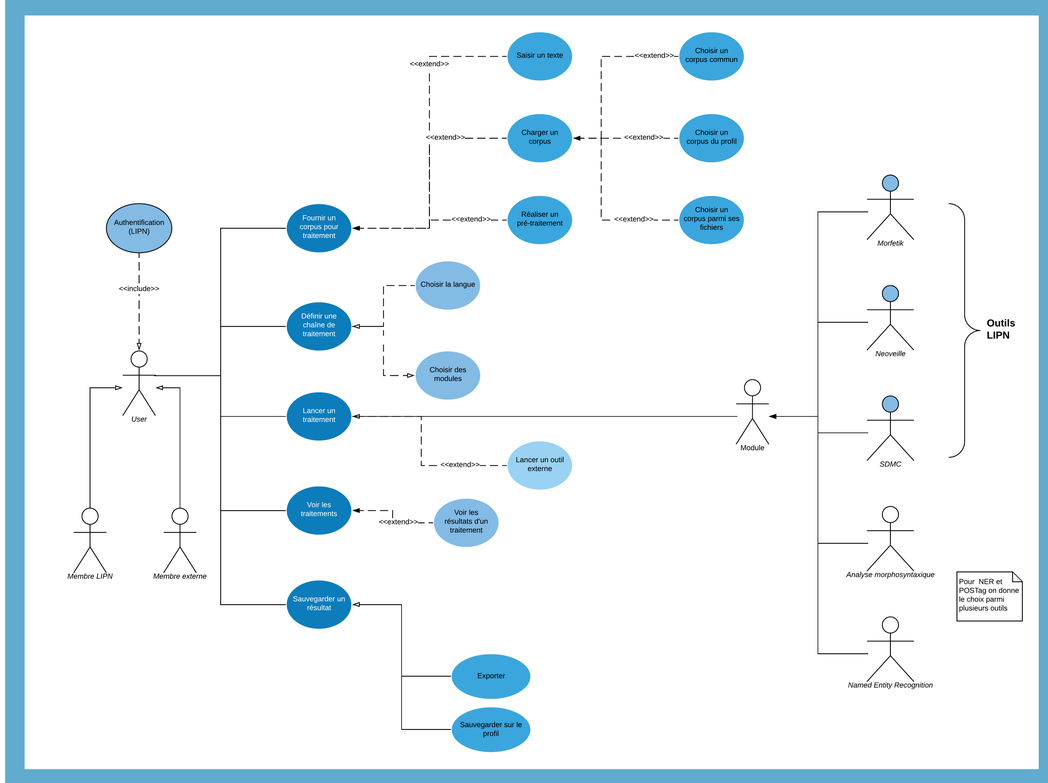
- → Entités nommées
- → Néologismes
- → Expressions figées
- → Dictionnaire morphologique
- → Fouille de motifs
- → Analyse syntaxique
- → Annotation sémantique
- → Traduction automatique
- → Résumé automatique
- → Génération de texte
- → Extraction d'information
- → ...

Usage

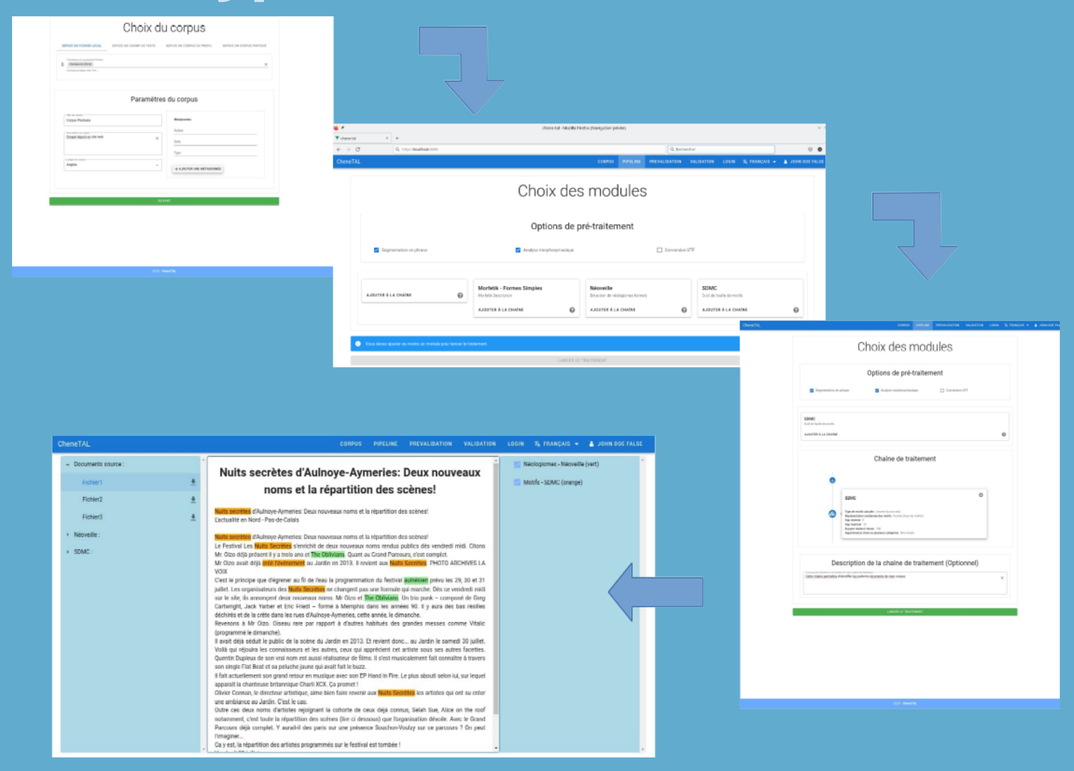
1. **Définir le traitement :**
   - — **séquence de processus**
   - — type de données
   - — paramètres d'interaction
   - — type de résultat
2. **Fournir des données**
   - — Données utilisateurs
   - — Données existantes (dépôt partagé)
3. **Lancer le traitement**
4. **Rapatrier les résultats**
   - — texte annoté au **format CoNLL-U**
   - — statistiques
   - — fichiers de sortie

Prototype fonctionnel

Cas d'utilisation



Prototype



Accélérateurs GPU pour l'IA

Espace de développement pour préparer des expériences en IA, avant d'entraîner les modèles en production sur *Jean Zay*.

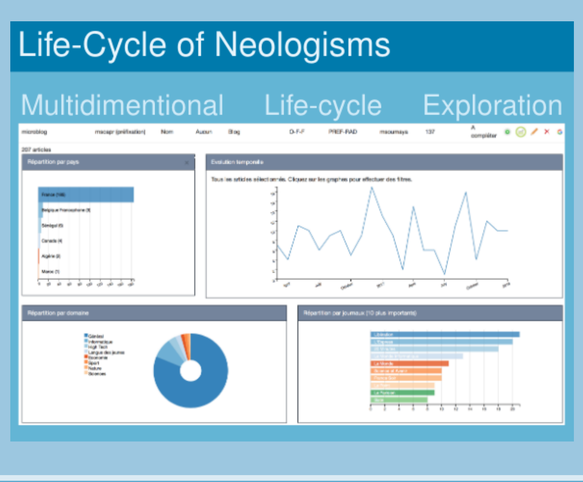
- — 11 GPUs Quadro RTX 5000,16 Go de RAM
- — 8 GPUs GEForce RTX 2080, 8 Go de RAM
- — 8 GPUs A40, 48Go de RAM

Premiers modules de TAL intégrés au prototype

Extraction de néologismes

Néoveille

https://tal.lipn.univ-paris13.fr/neoveille



Dictionnaire morphologique et extraction de motifs

https://tal.lipn.univ-paris13.fr/chenetal

Emmanuel Cartier, Thierry Charnois,
Quentin David, Jorge Garcia Flores, Aude Grezka
(équipe RCLN)