

Fiche technique de PALM

- Accessibilité: Une plateforme web
- Interface: Basé sur interfaces hommes-machines (front-end) permettant la consultation d'une bibliothèque, la gestion d'un corpus de travail, l'annotation semi-automatique et l'export d'un corpus.
- Systèmes: Des navigateurs supportés par les système d'exploitation Windows, Linux et MacOS X essentiellement firefox et google-chrome?
- Tagueur: Utilisation des modèles Treetagger pour trois langues à savoir moyen français, moyen anglais et latin.

Remarque : Un système à base de règles NooJ pour le moyen français et un système à base de réseaux de neurones pour les trois langues est en cours d'implémentation.

- Langue(s): Moyen français, moyen anglais et latin médiéval
- Jeux d'étiquettes: Jeu d'étiquettes définit selon la langue.

PALM propose les parties du discours suivantes pour le latin:

Verbe

Adverbe

Nom commun

Nom propre

Pronom

Adjectif

Préposition

Conjonction de subordination

Conjonction of coordination

Nombre

Ponctuation

Interjection

PALM propose les parties du discours suivantes pour le moyen français:

Verbe

Adverbe

Nom commun

Nom propre

Pronom

Adjectif

Déterminant

Préposition

Conjonction of subordination

Conjonction of coordination

Nombre cardinal

Nombre ordinal

Interjection
Ponctuation

Les parties du discours suivantes ont été utilisées pour le moyen anglais:

Verbe
Adverbe
Nom commun
Nom propre
Pronom
Nom verbal
Adjectif
Déterminant
Préposition
Conjonction
Nombre cardinal
Nombre ordinal
Infinitive Marker
Verbe+Pronom
Interjection
Ponctuation

- **Corpus d'entraînement: Corpus MEDITEXT**

MEDITEXT est un corpus textuel d'abord rassemblé sous la direction de Jean-Philippe Genet et Claude Gauvard. Il a été corrigé et augmenté dans le cadre du projet du European Research Council « Signs and States » de 2010 à 2014 (voir ci-dessous, paragraphe 1.8). Il est la base de la bibliothèque interne de PALM.

MEDITEXT, et par conséquent la bibliothèque interne de PALM, rassemble essentiellement des textes « politiques » : c'est-à-dire, soit des textes ayant trait à des événements politiques identifiés (discours ; lettres ; traités ; poèmes ; sermons ; chroniques) ; soit des textes consacrés de manière générale au bon et au mauvais gouvernement. Il rassemble également des textes gouvernementaux (proclamations, ordonnances) et des textes adressés au roi par ses sujets (cahiers de doléances ; requêtes ; lettres de rémission).

Pour le moment, la bibliothèque interne de PALM regroupe des textes d'origine anglaise (en anglais, français et latin) et d'origine française (en français et en latin). Mais nous souhaitons, à l'avenir, ajouter des textes en provenance d'autres pays européens.

- **Réentraînement/personnalisation:** Il est possible que l'utilisateur constitue un corpus de travail personnalisé selon ces besoins et l'annoté semi-automatiquement. Cependant, il n'est pas encore envisageable de réentraîner le tagueur. En effet, notre démarche consiste à mettre en place une annotation automatique via plusieurs outils afin de rendre notre analyse fiable et efficace quel que soit le corpus de travail de l'utilisateur.

- **Licence:** La licence CeCILL (<http://www.cecill.info/>)

- **Export/import - format(s):** Vers plusieurs logiciels de textométrie comme Lexico3, Le Trameur, TXM et Hyperbase et vers des formats standard comme le .TXT et le XML-TEI.

- **Dernière version/date:** Les dernières modifications de la version actuelle date de Juin 2017 et il y a une nouvelle version en cours de développement qui verra le jour cette année.

- Site web: <http://palm.huma-num.fr/>
- Documentation (manuel): En Français (<http://palm.huma-num.fr/PALM/faces/images/ManuelPALMFrMai2014.pdf>) et en Anglais (<http://palm.huma-num.fr/PALM/faces/images/PALMManualEngMay2014.pdf>)
- Contacts/responsables:
aude.mairey@gmail.com & aouini.mourad@gmail.com