

Cosme² (Consortium Sources Médiévales 2)
Groupe de travail « Lemmes » - Atelier 4

Journée d'initiation à la lemmatisation des textes médiévaux

lundi 17 juin 2019 - 9h30-18h

Paris - IRHT - Salle Jeanne Vielliard - 40 avenue d'Iéna (métro Iéna)

<https://cosme.hypotheses.org/1127>

En plus d'être flexionnelles, les langues employées dans l'Occident médiéval (latin et langues vernaculaires) se caractérisent par une très grande variation orthographique. La lemmatisation, c'est-à-dire le regroupement des différentes formes d'un mot sous leur lemme, se trouve ainsi être une opération fondamentale aux études des textes, notamment celles utilisant des procédures de recherche formalisées et assistées informatiquement.

Dans le contexte actuel de multiplication des numérisations des textes médiévaux, plusieurs outils et paramètres de lemmatisation ont vu le jour et continuent d'être développés. De l'aide à la traduction jusqu'à la préparation aux analyses statistiques et sémantiques des larges corpus, ces outils demeurent, cependant, pas ou peu connus des médiévistes.

L'objectif de cette journée est de proposer une introduction à la lemmatisation et la découverte de quelques outils-clé à partir d'exercices pratiques. Elle vise aussi à faire se rencontrer les concepteurs d'applications et leurs utilisateurs potentiels.

Cette initiation s'adresse à tous les intéressés : étudiants, doctorants, ingénieurs, chercheurs et enseignants-chercheurs. Pour que cette initiation s'accompagne d'une mise en pratique réelle, le nombre des participants est limité à 15.

La participation est gratuite, mais devra obligatoirement faire l'objet d'une demande motivée, pour que la limite fixée à 15 participants puisse être vérifiée.

La prise en charge des frais de transport, voire de logement (une nuitée), est possible dans la mesure des crédits disponibles, à condition d'en faire la demande explicite au moment de la candidature. L'ordre des candidatures sera pris en compte en cas de besoin.

Pour faire acte de candidature, adressez par courriel une **demande motivée** à eliana.magnani@univ-paris1.fr avant le **15 mai 2019**, en précisant vos nom, prénom, institution de rattachement, thème de recherche, courriel et, le cas échéant, les besoins de prise en charge (transport, nuitée). Tous les candidats seront prévenus par courriel le 20 mai de l'état de leur candidature.

Programme mis à jour

- 9h30 - Accueil et tour de table introductif
- 10h - La lemmatisation des corpus médiévaux. Pourquoi et comment ? (Nicolas Perreaux)
- 11h45 - Pause
- 12h - Installation, prise en main et utilisation de Collatinus. Lemmatiseur et analyseur morphologique de textes latins (Yves Ouvrard, Philippe Verkerk)
- 13h15 - Buffet
- 14h30 - Phyrre : lemmatisation et correction (latin, ancien français, etc.) (Thibault Clérice, Vincent Jolivet, Ariane Pinche)
- 16h - Pause
- 16h30 - La lemmatisation avec TXM (Nicolas Perreaux)
- 17h30 - Tour de table conclusif

Organisation

Eliana Magnani (CNRS-LaMOP), Nicolas Perreaux (Université de Namur) et l'équipe administrative de l'IRHT

Comité scientifique

Paul Bertrand (Univ. Louvain)
Estelle Ingrand-Varenne (CNRS-CESCM)
Eliana Magnani (CNRS-LaMOP)
Nicolas Perreaux (SFB 1095 - Goethe-Universität)

Liens utiles

<http://cosme.hypotheses.org/>
<https://www.huma-num.fr/consortiums>
<https://www.irht.cnrs.fr/>
<https://outils.biblissima.fr/fr/collatinus-web/>
<http://www.glossaria.eu/treetagger/>
<http://palm.huma-num.fr/PALM/>
<http://textometrie.ens-lyon.fr/>