

Outils d'interrogation corpus



ECTS
3 crédits



Composante
UFR Études
Interculturelles
de Langues
Appliquées



Volume horaire
24h



Période de
l'année
Semestre 1

En bref

- > Langue(s) d'enseignement: Français
- > Ouvert aux étudiants en échange: Non

Présentation

DESCRIPTION

Programme :

Familiarisation avec la construction et l'exploitation des corpus

1. Construction et compilation d'un corpus à utiliser pour l'activité d'analyse terminologique et pour la traduction spécialisée
2. Familiarisation avec les formats de fichiers
3. Utilisation des concordanciers pour l'interrogation des corpus
4. Introduction aux expressions régulières
5. Introduction à l'annotation des corpus

6. Connaissances de base sur les outils d'analyse automatique: tokenisation, étiquetage morphosyntaxique

OBJECTIFS

Compétences visées

- Connaître les différents types de corpus et d'outils d'annotation et d'exploitation des corpus.
- Connaître les principaux formats de fichiers impliqués dans la constitution des corpus (.txt, .pdf, .html, .xml) et les principaux langages utilisés pour l'interrogation des corpus : les expressions régulières, le langage de requête CQP

HEURES D'ENSEIGNEMENT

Outils d'interrogation corpus	Travaux Dirigés	24h
-------------------------------	-----------------	-----

PRÉ-REQUIS NÉCESSAIRES

Connaissances bureautiques de base

Pour en savoir plus, rendez-vous sur > u-paris.fr/choisir-sa-formation