






5.00 crédits	30.0 h + 15.0 h	Q1
--------------	-----------------	----

Enseignants	Dupont Pierre ;
Langue d'enseignement	Anglais > Facilités pour suivre le cours en français
Lieu du cours	Louvain-la-Neuve
Thèmes abordés	<ul style="list-style-type: none"> • Différents niveaux d'analyses linguistiques • Traitements de corpus • Etiquetage en parties du discours • Modélisation probabiliste du langage (Ngrams et Modèles de Markov Cachés) • Grammaires formelles et algorithmes d'analyse syntaxique • Traduction automatique, apprentissage profond • Applications en ingénierie linguistique telles que les logiciels de complétion automatique, d'étiquetage automatique, d'analyse ou de traduction automatique
Acquis d'apprentissage	<p>A la fin de cette unité d'enseignement, l'étudiant est capable de :</p> <p>Eu égard au référentiel AA du programme « Master ingénieur civil en informatique », ce cours contribue au développement, à l'acquisition et à l'évaluation des acquis d'apprentissage suivants :</p> <p>INFO1.1-3 INFO2.3-4 INFO5.3-5 INFO6.1, INFO6.4</p> <p>Eu égard au référentiel AA du programme « Master [120] en sciences informatiques », ce cours contribue au développement, à l'acquisition et à l'évaluation des acquis d'apprentissage suivants :</p> <p>SINF1.M4 SINF2.3-4 SINF5.3-5 SINF6.1, SINF6.4</p> <p>Les étudiants ayant suivi avec fruit ce cours seront capables de</p> <ol style="list-style-type: none"> 1 <ul style="list-style-type: none"> • décrire les concepts fondamentaux de la modélisation du langage naturel • maîtriser la méthodologie de l'utilisation de ressources linguistiques, en particulier sous la forme de corpus à grande échelle, éventuellement annotés ou structurés • appliquer de manière pertinente les techniques statistiques de modélisation du langage • mettre en oeuvre des méthodes récentes relevant de l'apprentissage automatique en lien avec le traitement du langage • développer des applications en ingénierie linguistique <p>Les étudiants auront développé des compétences méthodologiques et opérationnelles. En particulier, ils auront développé leur capacité à</p> <ul style="list-style-type: none"> • s'intégrer dans une approche pluridisciplinaire entre l'informatique et la linguistique, en utilisant à bon escient la terminologie, les outils et méthodes existants, • gérer au mieux le temps disponible pour mener à bien des projets de moyenne ampleur, • manipuler et exploiter de grandes quantités de données.

Modes d'évaluation des acquis des étudiants	<h3>Calcul de la note globale pour le cours</h3> <p>Les projets valent pour 30 % de la note finale, 70 % pour l'examen final (à livre fermé). Les projets ne peuvent pas être refaits en seconde session. La note globale des projets est fixée à la fin du quadrimestre et reprise telle quelle dans la note finale en seconde session. L'examen final est, par défaut, un écrit (sur papier ou, le cas échéant, sur un ordinateur).</p> <h3>Règles concernant les collaborations entre étudiants et l'usage de ressources externes</h3> <p>Une étude collaborative entre les étudiant.e.s est encouragée lors des séances de suivi de projets et via un forum d'échanges sur Moodle. Chaque étudiant.e doit soumettre une solution personnelle pour chaque projet. Le recours à des ressources publiques (p.ex. stackoverflow.com), ce-inclus des IA génératives (p.ex. chatGPT) est autorisé pour autant que chaque (fragment de) code soumis par l'étudiant.e mentionne toutes les ressources utilisées. La diffusion ou l'échange entre étudiants de (fragments de) code ne sont pas autorisés par quelque moyen que ce soit (GitHub, Facebook, Discord, ...) et ce même après la date limite de remise des projets. Le non-respect de ces règles pour n'importe quel projet peut entraîner une note globale de 0 pour l'ensemble des projets. L'examen final sur ordinateur doit se faire sans accès à aucune ressource externe. Ces règles sont exposées et détaillées lors du premier cours (voir site Moodle du cours).</p>
Méthodes d'enseignement	<ul style="list-style-type: none"> • Cours magistraux • Projets pratiques implémentés en Python sur le serveur <i>Inginious</i>
Contenu	<ul style="list-style-type: none"> • Différents niveaux d'analyse linguistique • Traitement (automatisé) de corpus : formatage, tokenization, marquage des données • Modélisation probabilistes du langage : N-grams, HMMs • Etiquetage en parties du discours • Introduction à l'apprentissage profond • Réponse aux questions • Traduction automatique • Applications représentatives telles que la génération automatique de texte, le marquage automatique de parties de texte, la traduction automatique ou agents conversationnels
Ressources en ligne	moodle.uclouvain.be/course/view.php?id=1182
Bibliographie	One recommended textbook - un ouvrage conseillé : <ul style="list-style-type: none"> • Speech and Language Processing, D. Jurafsky and J.H. Martin, Prentice Hall.
Faculté ou entité en charge:	INFO

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en science des données, orientation statistique	DATS2M	5		
Master [120] en linguistique	LING2M	5		
Master [120] : ingénieur civil en informatique	INFO2M	5		
Master [120] en sciences informatiques	SINF2M	5		
Master [120] : ingénieur civil en science des données	DATE2M	5		
Master [120] en science des données, orientation technologies de l'information	DATI2M	5		