

5.0 crédits	30.0 h + 15.0 h	1q
-------------	-----------------	----

Enseignants:	Dupont Pierre ; Fairon Cédric ;
Langue d'enseignement:	Anglais
Lieu du cours	Louvain-la-Neuve
Ressources en ligne:	http://www.icampus.ucl.ac.be/claroline/course/index.php?cid=INGI2263
Thèmes abordés :	<p>--</p> <p>Concepts de base en phonologie, morphologie, syntaxe et sémantique</p> <p>--</p> <p>Ressources linguistiques</p> <p>--</p> <p>Etiquetage en partie du discours</p> <p>--</p> <p>Modélisation statistique de la langue (N-grams et modèles de Markov cachés)</p> <p>--</p> <p>Algorithmes d'analyse robuste, grammaires hors-contexte probabilistes</p> <p>--</p> <p>Applications de l'ingénierie linguistique telles que les logiciels de correction orthographique ou grammaticale, l'étiquetage en partie du discours ,l'indexation automatique de documents ou la catégorisation de textes</p>
Acquis d'apprentissage	<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>--</p> <p>INFO1.1-3</p> <p>--</p> <p>INFO2.3-4</p> <p>--</p> <p>INFO5.3-5</p> <p>--</p> <p>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>--</p> <p>SINF1.M4</p> <p>--</p> <p>SINF2.3-4</p> <p>--</p> <p>SINF5.3-5</p> <p>--</p> <p>SINF6.1, SINF6.4</p> <p>Les étudiants ayant suivi avec fruit ce cours seront capables de</p> <p>--</p> <p>décrire les concepts fondamentaux de la modélisation du langage naturel</p> <p>--</p> <p>maîtriser la méthodologie de l'utilisation de ressources linguistiques (corpus, dictionnaires, réseaux sémantiques, etc) et faire un choix argumenté entre les différentes ressources linguistiques</p> <p>--</p> <p>appliquer de manière pertinente les techniques statistiques de modélisation du langage</p> <p>--</p> <p>développer des applications en ingénierie linguistique</p> <p>Les étudiants auront développé des compétences méthodologiques et opérationnelles. En particulier, ils auront développé leur capacité à</p> <p>--</p> <p>s'intégrer dans une approche pluridisciplinaire à la frontière entre l'informatique et la linguistique, en utilisant à bon escient la terminologie et les outils de l'une ou l'autre discipline,</p> <p>--</p> <p>gérer au mieux le temps disponible pour mener à bien des mini-projets,</p> <p>--</p> <p>manipuler et exploiter de grandes quantités de données.</p>

	<i>La contribution de cette UE au développement et à la maîtrise des compétences et acquis du (des) programme(s) est accessible à la fin de cette fiche, dans la partie « Programmes/formations proposant cette unité d'enseignement (UE) ».</i>
Modes d'évaluation des acquis des étudiants :	25 % pour les travaux pratiques + 75 % examen final (livre fermé) Pas de possibilité de refaire les travaux en seconde session
Méthodes d'enseignement :	-- 12 cours magistraux -- 3 miniprojets -- séance de discussion sur les corrections des miniprojets
Contenu :	-- Bases en linguistique : morphologie, partie du discours, structures de phrase, sémantique et pragmatique -- Rappels mathématiques : langages formels et éléments de la théorie de l'information -- Analyse de corpus : formatage, tokenization, morphologie, marquage des données -- N-grams : estimation par maximum de vraisemblance et lissage -- Modèles de Markov Cachés : définitions, algorithmes de Baum-Welch et de Viterbi -- Étiquetage en parties du discours -- Grammaires hors-contexte probabilistes : estimation des paramètres et algorithmes d'analyse, utilisation de treebanks -- Traduction automatique: méthodes classiques et statistiques (modèles IBM, modèles basés sur la phrase), évaluation -- Applications : prédicteur de mots pour SMS, marquage de parties de texte, outils d'extraction d'information http://www.icampus.ucl.ac.be/claroline/course/index.php?cid=INGI2263
Bibliographie :	Slides obligatoires disponibles sur le site : http://www.icampus.ucl.ac.be/claroline/course/index.php?cid=INGI2263 1 textbook conseillé : -- Speech and Language Processing (2nd Edition), D. Jurafsky and J.H. Martin, Prentice Hall, 2009.
Autres infos :	Préalables: -- LSINF1121 : Algorithmique et structure de données
Cycle et année d'étude :	> Master [120] en statistiques, orientation générale > Master [120] en sciences informatiques > Master [120] : ingénieur civil en informatique > Master [120] en linguistique
Faculté ou entité en charge:	INFO