

Mention et/ou parcours dont relève cette UE : DEG_Master Traitement Automatique des Langues (TAL)

Numéro de l'UE :

Nom complet de l'UE : UE 901 Apprentissage profond et fouille de données

Composante de rattachement : BH0

Nom du responsable de l'UE et adresse électronique : Emmanuel Vincent
emmanuel.vincent@inria.fr

Semestre : 9

Volume horaire enseigné : 72h, Nombre de crédits ECTS : 6

Volume horaire travail personnel de l'étudiant : 108h

Langue d'enseignement de l'UE : Anglais

Enseignements composants l'UE	CNU	CM	TD	TPL	EqT D
Mat1 Réseaux de neurones	2700	9	9		22,5
MAT2 Réseaux de neurones pour le TAL	2700	5	4		11,5
Mat1 Fouille de données	2700	9	9		22,5
MAT2 Fouille de données pour le TAL	2700	5	4		11,5
EC3 Systèmes intelligents et de recommandation	2700	9	9		22,5

Descriptif

L'objectif de cette UE est d'acquérir des outils d'apprentissage automatique, notamment les réseaux de neurones profonds et les factorisations matricielles, de les manipuler sur des cas d'usage pratiques (tweets, traces d'e-learning), ainsi que d'approfondir les connaissances en web sémantique et les extensions de l'analyse formelle de concepts pour le traitement de données textuelles et relationnelles.

EC1 Réseaux de neurones (composé des Mat1 et Mat2)

L'EC fournit un panorama de l'apprentissage automatique par réseaux de neurones, des perceptrons multicouches aux architectures avancées utilisées aujourd'hui (réseau déterministe/probabiliste, réseaux convolutifs, réseaux récurrents). Il revient sur les bases algorithmiques relatives aux réseaux de neurones, ainsi que la présentation des propriétés permettant la sélection des architectures de réseaux les plus appropriées selon la tâche en pratique.

EC2 Fouille de données complexes et de texte (composé des Mat1 et Mat2)

Cet EC a pour objectif de présenter les outils et techniques de découverte de connaissances pour le traitement de données complexes et textuels. Il présente des méthodes et des outils de fouille de données (séquences, graphes) pour la classification et fouille de textes (pattern structures, analyse relationnelle de concepts, fouille de données exploratoire, skylines). Il présente aussi des techniques d'extraction d'information qui font intervenir des méthodes de classification non-supervisée (comme les k-means) ainsi que d'analyse sémantique latente dans des textes (LSA).

EC3 Filtrage collaboratif

Ce cours aborde un des aspects de l'apprentissage automatique en traitant le cas particulier des systèmes de recommandation qui s'appuient notamment sur les traces d'usage collectées dans les réseaux sociaux. L'objectif de l'EC est de mettre en place un dispositif d'évaluation de systèmes de recommandation. Ce cours s'appuie sur une étude de cas, matérialisée par un ensemble de données et par une boîte mise à outils logicielle à disposition des étudiants pour mener des expérimentations en autonomie.

Pré-requis

UE801

Acquis d'apprentissage

EC1-1 maîtriser les théories sur les réseaux de neurones

EC1-2 appliquer ces théories pour obtenir une solution

EC2-1 utilisation de la sémantique latente pour la fouille de textes

EC2-2 extraire des informations par des méthodes de classification

EC3-1 identification des traces d'utilisation d'un système

EC3-2 connaissance des mécanismes de recommandation

Compétences visées

1. Analyser un problème devant être traité informatiquement sur des données issues de l'oral et /ou de l'écrit.

4. Savoir mettre en œuvre des techniques algorithmiques, d'analyse linguistique, de statistique et de traitement des connaissances.

Mention et/ou parcours dont relève cette UE : DEG_Master Traitement Automatique des Langues (TAL)

Numéro de l'UE :

Nom complet de l'UE : UE 902 Traitement de l'écrit et de l'oral

Composante de rattachement : BH0

Nom du responsable de l'UE et adresse électronique : Claire Gardent claire.gardent@loria.fr

Semestre : 9

Volume horaire enseigné : 72h, Nombre de crédits ECTS : 6

Volume horaire travail personnel de l'étudiant : 108h

Langue d'enseignement de l'UE : Anglais

Enseignements composants l'UE	CNU	CM	TD	TPL	EqT D
EC1 Traitement de l'oral	2700	12	12		30
EC2 Traitements de l'écrit	2700	12	12		30
EC3 Terminologie et ontologies	2700	12	12		30

Descriptif

Le traitement automatique des textes et de la parole met en jeu différentes méthodes d'apprentissage automatique. Ce cours vise à présenter ces méthodes et à illustrer leur utilisation par des exemples concrets d'applications ou d'outils développés pour le traitement de l'écrit et de l'oral.

EC1 Traitement de l'oral

Cet EC a pour objectif d'enseigner les connaissances, les outils de modélisation statistique et neuromimétique et les logiciels utilisés dans le domaine du traitement automatique de la parole. Le cours introduit d'abord les concepts de traitement du signal pour la parole. Il présente ensuite les approches par modèles de Markov cachés, les approches par réseaux de neurones profonds apparues récemment et finalement les logiciels utilisés pour la mise en œuvre d'un système réel en synthèse et en reconnaissance automatique.

EC2 Traitement de l'écrit

Cet EC présente les méthodes d'apprentissage et les bibliothèques spécialisées (NLTK, Gizza, etc.) couramment exploitées pour modéliser différentes sous-tâches (étiquetage et analyse syntaxiques, reconnaissance d'entités nommées, etc.) et applications du traitement de l'écrit. Un focus sur la traduction automatique permettra d'illustrer ces étapes sur des corpus parallèles et de re-visiter l'algorithme Expectation-Maximisation.

EC3 Terminologie et ontologie

Cet EC étudie le lien entre le niveau lexical et le niveau ontologique (lié à la connaissance du domaine des textes). En effet, dans les domaines spécialisés, les termes sont souvent des unités poly-lexicales et leur traitement est une étape essentielle pour la compréhension des textes. Nous expérimenterons le logiciel TTC TermSuite et en analyserons les résultats. Les outils de classification étudiés auparavant seront utilisés pour organiser ces termes. Nous verrons aussi des cas pratiques de ressources terminologiques, de ressources lexicales transdisciplinaires et ferons le lien avec les Linked Open Data.

Pré-requis

UE 704, UE 803 et UE 804

Acquis d'apprentissage

EC1-1 identifier et appliquer les outils pour traiter les données issues de l'oral

EC1-2 utiliser les outils de reconnaissance de la parole

EC2-1 utiliser les outils du TAL pour les tâches classiques

EC2-2 mettre en œuvre les techniques pour la traduction automatique

EC3-1 construire des ontologies de domaines

EC3-2 manipulation des ressources terminologiques

Compétences visées

1. Analyser un problème devant être traité informatiquement sur des données issues de l'oral et /ou de l'écrit.

3. Mettre en synergie des connaissances pluridisciplinaires au sein de solutions innovantes.

4. Savoir mettre en œuvre des techniques algorithmiques, d'analyse linguistique, de statistique et de traitement des connaissances.

Mention et/ou parcours dont relève cette UE : DEG_Master Traitement Automatique des Langues (TAL)

Numéro de l'UE :

Nom complet de l'UE : UE 903 Traitement des textes et du discours

Composante de rattachement : BH0

Nom du responsable de l'UE et adresse électronique : Maxime Amblard-Virollaud
maxime.amblard@univ-lorraine.fr

Semestre : 9

Volume horaire enseigné : 62h, Nombre de crédits ECTS : 6

Volume horaire travail personnel de l'étudiant : 98h

Langue d'enseignement de l'UE : Anglais

Enseignements composants l'UE	CNU	CM	TD	TPL	EqT D
EC1 Applications aux textes	2700	13	13		32,5
EC2 Sémantique Computationnelle	2700	9	9		22,5
EC3 Discours et modèle de discours	2700	9	9		22,5

Descriptif

Cette UE rassemble les cours traitant des niveaux sémantique et discursif.

EC1 Applications au texte

1/ application à la génération de texte en introduisant deux thématiques : génération à partir de données (provenant de bases de données ou de connaissances) et génération à partir de texte (simplification et résumé automatique). Le cours introduira les différents types d'approches proposées (symboliques, probabilistes, statistiques).

2/ analyse en sentiment et opinions multilingues : notion de corpus comparables, méthodes état de l'art d'alignement entre documents (via dictionnaires bilingues et approche sémantique latente LSA), détection d'opinions et de sentiments dans les textes (différentes typologies, algorithmes d'apprentissage appliqués à cette détection).

EC2 Sémantique computationnelle

L'objectif de cet EC est d'introduire les différents moyens de représenter le sens d'expressions complexes en se limitant à la phrase. Une partie présente les représentations du sens à l'aide du lambda-calcul et de la logique d'ordre supérieur, et une autre présente la manière de construire le sens d'expressions complexes à partir du sens des éléments qui les composent. Nous décrivons l'association systématique de sens aux mots d'une expression pour construire le sens complet de l'expression en fonction des règles syntaxiques utilisées.

EC3 Discours et modèle du discours

L'objectif de cet EC est d'exposer les phénomènes linguistiques qui apparaissent lorsque l'on ne considère plus seulement des phrases, mais des séquences de phrases (discours), ainsi que de présenter leur modélisation. Le sens de certains éléments (tels que les pronoms) ne peut être calculé que relativement à un contexte donné. D'autres nécessitent la prise en

compte de la structure hiérarchique des textes et de l'articulation entre paragraphes. Cet EC introduit les formalismes utilisés pour prendre en compte ces phénomènes, notamment la DRT, la DPL, la SDRT, et la sémantique des continuations.

Pré-requis

UE 701, UE 704 ;
UE 801, UE 802, UE 804 ;
UE 902

Acquis d'apprentissage

EC1-1 appliquer les principes de la génération sur les textes
EC1-2 connaître les principes du multilinguisme

EC2-1 connaître des théories de la représentation de la sémantique et du discours
EC2-2 appliquer le lambda-calcul et la logique à la sémantique

EC3-1 appliquer les théories de la représentation du discours
EC3-2 résoudre des problèmes linguistiques au niveau du discours

Compétences visées

1. Analyser un problème devant être traité informatiquement sur des données issues de l'oral et /ou de l'écrit.
2. Savoir recueillir, structurer et représenter des données (son, texte, image, ...).
4. Savoir mettre en œuvre des techniques algorithmiques, d'analyse linguistique, de statistique et de traitement des connaissances.

Mention et/ou parcours dont relève cette UE : DEG_Master Traitement Automatique des Langues (TAL)

Numéro de l'UE :

Nom complet de l'UE : UE 904 Lexiques et grammaires pour le TAL

Composante de rattachement : BH0

Nom du responsable de l'UE et adresse électronique : Bruno Guillaume
bruno.guillaume@univ-lorraine.fr

Semestre : 9

Volume horaire enseigné : 72h, Nombre de crédits ECTS : 6

Volume horaire travail personnel de l'étudiant : 108h

Langue d'enseignement de l'UE : Anglais

Enseignements composants l'UE	CNU	CM	TD	TPL	EqT D
UE704.1 EC1 Lexicologie Diachronique et Synchronique	0700	12	12		30
EC2 Ressources lexicales	2700	12	12		30
EC3 Modèles syntaxiques	2700	12	12		30

Descriptif

L'UE 904 présente des outils avancés pour la modélisation informatique de différents types d'informations linguistiques qui décrivent les unités lexicales (ressources lexicales) ou les règles de formation des unités plus larges (grammaires).

EC1 Lexicologie diachronique et synchronique

Ce cours constitue une introduction complète à la lexicologie et la lexicographie, que ce soit synchronique (description d'un mot à un moment précis) ou diachronique (description de l'évolution d'un mot dans le temps). Les étudiants devront prendre conscience des divergences entre l'analyse lexicologique et la description lexicographique, en particulier dans les domaines de la sémantique et l'étymologie. Ils devront être capable de transposer les résultats des recherches lexicologiques dans une description lexicographique pour un dictionnaire, en utilisant les méthodologies étudiées en cours.

EC2 Ressources lexicales

Ce cours présente les principaux types de ressources lexicales informatisées : dictionnaires électroniques, bases de données lexicales, ontologiques lexicales, etc. Il s'appuie sur les notions de base en lexicologie introduites dans l'EC "Lexicologie : unités lexicales et phraséologie » (UE804).

EC3 Modèles syntaxiques

Le but de ce cours est de présenter les concepts fondamentaux des formalismes grammaticaux symboliques pour décrire la syntaxe et l'interface syntaxe-sémantique de la langue naturelle. Pour quelques formalismes, par exemple les différentes présentations des grammaires catégorielles, les grammaires d'arbres adjoints (TAG) ou les Grammaires fonctionnelles lexicales (LFG). On étudiera leurs aspects mathématiques et on verra comment ils peuvent être mis en œuvre dans des systèmes informatiques qui passent à

l'échelle.

On étudiera notamment comment certains phénomènes linguistiques complexes (comme les verbes à montée, les verbes à contrôle ou les dépendances à longue distance) peuvent être représentés.

Pré-requis

UE 704, UE802 et UE804

Acquis d'apprentissage

EC1-1 situer un mot par rapport aux autres mots de la langue et dans son contexte

EC1-2 assurer la rédaction d'une notive de dictionnaire

EC2-1 Connaître les ressources lexicales existances

EC2-2 Construire des lexiques spécialisés

EC3-1 analyser un ensemble d'exemples linguistiques pour en extraire une généralisation

EC3-2 implémenter dans un formalisme existant une modélisation d'un phénomène linguistique

EC3-3 identifier les défauts et limites d'une formalisation

Compétences visées

1. Analyser un problème devant être traité informatiquement sur des données issues de l'oral et /ou de l'écrit.

3. Mettre en synergie des connaissances pluridisciplinaires au sein de solutions innovantes

Mention et/ou parcours dont relève cette UE : DEG_Master Traitement Automatique des Langues (TAL)

Numéro de l'UE :

Nom complet de l'UE : UE 905 Projet et cours de langues

Composante de rattachement : BH0

Nom du responsable de l'UE et adresse électronique : Manuel Rebuschi
manuel.rebuschi@univ-lorraine.fr

Semestre : 9

Volume horaire enseigné : 72h, Nombre de crédits ECTS : 6

Volume horaire travail personnel de l'étudiant : 108h

Langue d'enseignement de l'UE : Anglais

Enseignements composants l'UE	CNU	CM	TD	TPL	EqT D
EC1 Semaine de démarrage et remédiation	2700	6	6		15
EC2 Projet logiciel	2700	8	8		20
EC3 Droit et éthique des données	2700	5	5		12,5
EC4-A Méthodes bibliographiques	0000	5	5		12,5
EC4-B Insertion en entreprise	0000	5	5		12,5
EC5 Français	0900			24	24
EC5 Anglais	1100			24	24

Descriptif

Cette UE rassemble plusieurs enseignements dont le projet et les cours de langue.

EC1 Semaine de démarrage et de remédiation
remise à niveau en informatique, mathématique et linguistique

EC2 Projet logiciel
Il s'agit de conduire les étudiants à évaluer un de leurs travaux réalisés dans d'autres UE, relativement à l'acquisition des compétences. On privilégiera le choix de travaux pluridisciplinaires mobilisant notamment des compétences en programmation.

EC3 Droit et éthique des données
Une particularité de la formation est de massivement travailler sur des données, souvent personnelles. Ce cours a pour objectif de former les étudiants à respecter les droits des sujets qui participent à une expérimentation vis-à-vis de la collecte, du stockage et de la publication de leurs données.

Afin d'accompagner les étudiants dans leur insertion, ils choisissent l'un des deux EC :
EC4-A Méthodes bibliographiques

Le cours comporte deux aspects : une première partie présentant le système de recherche français (organisation en EPST, EPSCP, etc ; modes de financements, où trouver une thèse,

débouchés après thèse, etc.) et une seconde partie portant à proprement parler sur la recherche bibliographique (où et comment chercher, règles de citation, sensibilisation aux risques du plagiat) et la rédaction de rapports scientifiques.

EC4-B Insertion en entreprise

Cet UE présente aux étudiants la manière d'appréhender le monde de l'entreprise depuis leur cursus spécifique. Il s'agira de connaître les différents attendus et de préparer les documents afférents à une candidature en entreprise.

EC de cours de langue

Langue : anglais scientifique pour les non-anglophones

Langue : français quotidien pour les non-francophones

Pré-requis

UE 705 et UE 805

Acquis d'apprentissage

EC1 connaître tous les prérequis et identifier les ressources pour consolider les acquis

EC2 conduire un projet de grande envergure en groupe

EC3 connaître les différentes problématiques éthiques sur la constitution, l'utilisation et la diffusion de corpus

EC4-A utiliser les outils pour la recherche bibliographique

EC4-B maîtriser la communication en et vers les entreprises

EC5-1 consolidation de la connaissance de l'anglais scientifique

EC5-2 maîtrise du français pour la communication quotidienne

Compétences visées

1. Analyser un problème devant être traité informatiquement sur des données issues de l'oral et /ou de l'écrit.
2. Savoir recueillir, structurer et représenter des données (son, texte, image, ...).
3. Mettre en synergie des connaissances pluridisciplinaires au sein de solutions innovantes.
4. Savoir mettre en œuvre des techniques algorithmiques, d'analyse linguistique, de statistique et de traitement des connaissances.
5. Mener un projet d'envergure de manière autonome et en équipe.
6. Expliquer et documenter la mise en œuvre d'une solution technique.
7. Développer une argumentation avec esprit critique.
8. Faire preuve d'autonomie et d'initiative.