

Master 2 M2 Apprentissage et Traitement Automatique de la Langue (ATAL)

Année universitaire 2020-2021

Information générale

| | |
|--|---|
| Objectifs | |
| Responsable(s) | MORIN EMMANUEL QUINIOU SOLEN |
| Mention(s) incluant ce parcours | master Informatique |
| Lieu d'enseignement | Le Mans/Nantes |
| Langues / mobilité internationale | |
| Stage / alternance | |
| Poursuite d'études / débouchés | |
| Autres renseignements | |
| Conditions d'obtention de l'année | L'année est validée si la partie théorique est validée en première ou deuxième session (moyenne supérieure ou égale à 10/20) et si l'UE correspondant au stage est également validée avec une note supérieure ou égale à 10/20. |

Programme

| 1 ^{er} SEMESTRE | Code | ECTS | CM | CI | TD | TP | Distanciel | Total |
|---|--------------|------|-------|----|-------|------|------------|---------------|
| Groupe d'UE : UE à la charge de l'Université du Maine (16 ECTS) | | | | | | | | |
| Fouille de textes | X3ITM10 | 3 | 16 | 0 | 8 | 0 | 0 | 24 |
| Traduction automatique | X3ITM20 | 3 | 16 | 0 | 8 | 0 | 0 | 24 |
| Reconnaissance du locuteur | X3ITM30 | 3 | 16 | 0 | 8 | 0 | 0 | 24 |
| Reconnaissances et synthèse du langage | X3ITM40 | 3 | 24 | 0 | 12 | 0 | 0 | 36 |
| Apprentissage automatique en langue | X3IT050 | 4 | 16 | 0 | 8 | 12 | 0 | 36 |
| Groupe d'UE : UE à la charge de l'Université de Nantes (14 ECTS) | | | | | | | | |
| Développement logiciel en industrie de la langue | X3IT010 | 3 | 5.33 | 0 | 13.33 | 2.67 | 2.67 | 24 |
| Insertion professionnelle | X3IT020 | 2 | 5.33 | 0 | 13.33 | 2.67 | 2.67 | 24 |
| Algorithmique et alignement de chaînes | X3IT030 | 2 | 10.67 | 0 | 10.66 | 0 | 2.67 | 24 |
| Statistical and symbolic language modeling | X3IT040 | 3 | 16 | 0 | 5.33 | 0 | 2.67 | 24 |
| Conférences et cours invités | X3ITM50 | 1 | 0 | 0 | 6 | 0 | 0 | 6 |
| Corpus | X3IT060 | 1 | 10.67 | 0 | 0 | 0 | 1.33 | 12 |
| Projet | X3IT070 | 2 | 0 | 0 | 11 | 0 | 0 | 11 |
| | Total | 30 | | | | | 12.01 | 269.00 |

| 2 ^{ème} SEMESTRE | Code | ECTS | CM | CI | TD | TP | Distanciel | Total |
|--|--------------|------|----|----|----|----|------------|-------------|
| Groupe d'UE : Stage (30 ECTS) | | | | | | | | |
| Période de suivi en formation par alternance | X4II030 | 30 | 0 | 0 | 0 | 0 | 0 | 0 |
| Stage | X4II010 | 30 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Total | 30 | | | | | 0.00 | 0.00 |

Modalités d'évaluation

Mention Master 2ème année

Parcours : M2 Apprentissage et Traitement Automatique de la Langue (ATAL)

Année universitaire 2020-2021

Responsable(s) : MORIN EMMANUEL, QUINIOU SOLEN

REGIME ORDINAIRE

| | | | | | PREMIERE SESSION | | | | | | | DEUXIEME SESSION | | | | | | | TOTAL | |
|---|----------|--|---|-------------|------------------|-------|------|--------|-------|------|-------|------------------|-------|------|--------|-------|------|-------|--------|------|
| | | | | | Contrôle continu | | | Examen | | | | Contrôle continu | | | Examen | | | | Coeff. | ECTS |
| CODE UE | INTITULE | UE non dipl. | | | écrit | prat. | oral | écrit | prat. | oral | durée | écrit | prat. | oral | écrit | prat. | oral | durée | | |
| Groupe d'UE : UE à la charge de l'Université du Maine | | | | | | | | | | | | | | | | | | | | |
| 3 | X3ITM10 | Fouille de textes | N | obligatoire | 3 | | | | | | | 1.5 | | | 1.5 | | | | 3 | 3 |
| 3 | X3ITM20 | Traduction automatique | N | obligatoire | 3 | | | | | | | 1.5 | | | 1.5 | | | | 3 | 3 |
| 3 | X3ITM30 | Reconnaissance du locuteur | N | obligatoire | 3 | | | | | | | 1.5 | | | 1.5 | | | | 3 | 3 |
| 3 | X3ITM40 | Reconnaissances et synthèse du langage | N | obligatoire | 3 | | | | | | | 1.5 | | | 1.5 | | | | 3 | 3 |
| 3 | X3IT050 | Apprentissage automatique en langue | N | obligatoire | 4 | | | | | | | 2 | | | 2 | | | | 4 | 4 |
| Groupe d'UE : UE à la charge de l'Université de Nantes | | | | | | | | | | | | | | | | | | | | |
| 3 | X3IT010 | Développement logiciel en industrie de la langue | N | obligatoire | 1.5 | | 1.5 | | | | | 0.75 | | 0.75 | 1.5 | | | | 3 | 3 |
| 3 | X3IT020 | Insertion professionnelle | N | obligatoire | | | 2 | | | | | | | 1 | 1 | | | | 2 | 2 |
| 3 | X3IT030 | Algorithmique et alignement de chaînes | N | obligatoire | 2 | | | | | | | 1 | | | 1 | | | | 2 | 2 |
| 3 | X3IT040 | Statistical and symbolic language modeling | N | obligatoire | 3 | | | | | | | 1.5 | | | 1.5 | | | | 3 | 3 |
| 3 | X3ITM50 | Conférences et cours invités | N | obligatoire | | | | | | | | | | | | | | | 1 | 1 |
| 3 | X3IT060 | Corpus | N | obligatoire | 1 | | | | | | | 0.5 | | | 0.5 | | | | 1 | 1 |
| 3 | X3IT070 | Projet | N | obligatoire | | | 2 | | | | | | | 1 | | | 1 | | 2 | 2 |
| Groupe d'UE : Stage | | | | | | | | | | | | | | | | | | | | |
| 4 | X4II030 | Période de suivi en formation par alternance | N | optionnelle | 10 | 10 | 10 | | | | | 10 | 10 | 10 | | | | | 30 | 30 |
| 4 | X4II010 | Stage | N | optionnelle | 10 | 10 | 10 | | | | | 10 | 10 | 10 | | | | | 30 | 30 |
| TOTAL | | | | | | | | | | | | | | | | | | 60 | 60 | |

A la seconde session, les notes de contrôle continu correspondent à un report des notes de CC de la première session.

DISPENSE D'ASSIDUITE

| | | | | PREMIERE SESSION | | | | | | | | DEUXIEME SESSION | | | | | | | | TOTAL | |
|---|----------|--|---|------------------|-------|------|-------|--------|------|-------|-------|------------------|------|-------|-------|--------|-------|----|----|--------|------|
| | | | | Contrôle continu | | | | Examen | | | | Contrôle continu | | | | Examen | | | | Coeff. | ECTS |
| CODE UE | INTITULE | UE non dipl. | | écrit | prat. | oral | écrit | prat. | oral | durée | écrit | prat. | oral | écrit | prat. | oral | durée | | | | |
| Groupe d'UE : UE à la charge de l'Université du Maine | | | | | | | | | | | | | | | | | | | | | |
| 3 | X3ITM10 | Fouille de textes | N | obligatoire | | | | 3 | | | | | | | | | | 3 | 3 | | |
| 3 | X3ITM20 | Traduction automatique | N | obligatoire | | | | 3 | | | | | | | | | | 3 | 3 | | |
| 3 | X3ITM30 | Reconnaissance du locuteur | N | obligatoire | | | | 3 | | | | | | | | | | 3 | 3 | | |
| 3 | X3ITM40 | Reconnaissances et synthèse du langage | N | obligatoire | | | | 3 | | | | | | | | | | 3 | 3 | | |
| 3 | X3IT050 | Apprentissage automatique en langue | N | obligatoire | | | | 4 | | | | | | | | | | 4 | 4 | | |
| Groupe d'UE : UE à la charge de l'Université de Nantes | | | | | | | | | | | | | | | | | | | | | |
| 3 | X3IT010 | Développement logiciel en industrie de la langue | N | obligatoire | | | | 1.5 | | 1.5 | | | | | | | | 3 | 3 | | |
| 3 | X3IT020 | Insertion professionnelle | N | obligatoire | | | | | | 2 | | | | | | | | 2 | 2 | | |
| 3 | X3IT030 | Algorithmique et alignement de chaînes | N | obligatoire | | | | 2 | | | | | | | | | | 2 | 2 | | |
| 3 | X3IT040 | Statistical and symbolic language modeling | N | obligatoire | | | | 3 | | | | | | | | | | 3 | 3 | | |
| 3 | X3ITM50 | Conférences et cours invités | N | obligatoire | | | | | | | | | | | | | | | 1 | 1 | |
| 3 | X3IT060 | Corpus | N | obligatoire | | | | 1 | | | | | | | | | | 1 | 1 | | |
| 3 | X3IT070 | Projet | N | obligatoire | | | | | | 2 | | | | | | | 2 | 2 | 2 | | |
| Groupe d'UE : Stage | | | | | | | | | | | | | | | | | | | | | |
| 4 | X4II030 | Période de suivi en formation par alternance | N | optionnelle | | | | | | | | | | | | | | | 30 | 30 | |
| 4 | X4II010 | Stage | N | optionnelle | | | | | | | | | | | | | | | 30 | 30 | |
| TOTAL | | | | | | | | | | | | | | | | | | 60 | 60 | | |

A la seconde session, les notes de contrôle continu correspondent à un report des notes de CC de la première session.

Description des UE

| X3ITM10 | Fouille de textes |
|---------------------------------------|--|
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 24h Répartition : CM : 16h TD : 8h CI : 0h TP : 0h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requise(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Fouille de textes 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p><i>Savoir mettre en oeuvre et maîtriser une chaîne complète de fouille de données sur diverses tâches spécifiques à la fouille de textes : de la constitution pertinente du corpus à l'analyse argumentée des résultats en passant par la maîtrise des prétraitements et des algorithmes spécifiques au texte.</i></p> <ul style="list-style-type: none"> • Manipulation de corpus de textes : <ul style="list-style-type: none"> - Spécificités - Ensemble des pré-traitements linguistiques - Représentations vectorielles - Sélection et création de descripteurs pertinents - Ressources complémentaires • Algorithmes supplémentaires : <ul style="list-style-type: none"> - Arbre de décisions sémantiques - Conditional Random Field • Tâche de fouille de textes : <ul style="list-style-type: none"> - Catégorisation de texte - Reconnaissance d'entités nommées - Détection d'opinions dans un cadre supervisé |
| Méthodes d'enseignement | |
| Langue d'enseignement | Français |
| Bibliographie | <p>« Data Mining et Statistique décisionnelle », S. Tufféry, [Ed. Technip, 2009. « Data Mining : Practical machine learning tools and techniques with Java implementations », 3e Ed., I. Witten and E. Frank, Morgan Kaufman Pub., 2011. □ « Expérimentations et évaluations en fouille de textes : un panorama des campagnes DEFT », C. Grouin et D. Forest, Hermes Sciences Publications / Lavoisier, 2012. J.Ross Quinlan, <i>Machine Learning</i>, 1986, «Induction of Decision Trees», p.81-106 □ http://www.kdnuggets.com http://data.mining.free.fr https://deft.limsi.fr</p> |

| X3ITM20 | Traduction automatique |
|---------------------|------------------------|
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |

| | |
|---------------------------------------|--|
| Volume horaire total | TOTAL : 24h Répartition : CM : 16h TD : 8h CI : 0h TP : 0h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requise(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Traduction automatique 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p><i>Appréhender la complexité de la traduction automatique, de la traduction assistée par ordinateur et de son évaluation</i> <i>Comprendre la topologie des modèles de traduction automatique (statistique et neuronale)</i> <i>Comprendre les modèles de la traduction assistée par ordinateur</i> <i>Mettre en œuvre les méthodes étudiées sur des applications simples avec des outils open source</i></p> <ul style="list-style-type: none"> • Formulation statistique de la traduction automatique • Évaluation humaine et automatique • Phrase-based SMT • Hierarchical MT • Traduction automatique neuronale • Adaptation des systèmes de traduction • Post-édition et Computer Assisted Translation (CAT) |
| Méthodes d'enseignement | |
| Langue d'enseignement | Anglais |
| Bibliographie | |

| | |
|---------------------------------------|--|
| X3ITM30 | Reconnaissance du locuteur |
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 24h Répartition : CM : 16h TD : 8h CI : 0h TP : 0h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requise(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Reconnaissance du locuteur 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |

| | |
|-------------------------|---|
| Contenu | <p><i>Savoir définir ce qu'est la parole : comment elle est produite, comment elle perçue. Connaître les spécificités de la parole comme support du langage naturel. Connaître les différents traitements informatiques qui ont pour cible la parole, dans sa dimension linguistique mais aussi paralinguistique.</i></p> <ul style="list-style-type: none"> • Appareil productif et appareil perceptif humains de la parole • Codage de la parole : PCM, loi mu, loi A, container (format WAV et autres), compression psychoacoustique (MP3 et autres) • Notions de théorie de l'information, entropie, entropie conditionnelle, information mutuelle • Notions de traitement du signal audio : définition d'un son, calcul et interprétation d'un spectre sonore (transformation de Fourier, transformation cosinus discrète), calcul et interprétation d'un spectrogramme, cepstre, bancs de filtres, extraction de la fréquence fondamentale • Paramétrisations acoustiques : MFCC, LPC, PLP,... • Phonétique et Phonologie : phonèmes, phones, voyelles, consonnes, coarticulation, prosodie ... • Mots et unités sous-lexicales : parole continue et segmentation en mots, phonétisation automatique, liaisons, assimilations, ... • Focus sur les informations non linguistiques portées par la parole : émotions, langue, identité, âge, sexe, pathologie, ... • Modélisation : modélisation acoustique pour la reconnaissance de la parole (HMM/GMM et HMM/DNN), modélisation du locuteur, modélisation des langues. Introduction à l'analyse factorielle, description de l'algorithme espérance/maximisation (EM), ... • Syntaxe : spécificités du langage parlé, gestion des agrammaticalités et des disfluences |
| Méthodes d'enseignement | Hybride : présentiel (Le Mans) et distanciel (Nantes) synchrone Méthodologie : mix entre cours classique et classe inversée, création et utilisation de ressources numériques, et mini projet sous forme de TP en autonomie avec rencontres bilans lors des séances de TD. |
| Langue d'enseignement | Français |
| Bibliographie | Springer handbook of speech processing. Benesty, J., Sondhi, M. M., & Huang, Y. (Eds.). (2007). Springer Science & Business Media. Reconnaissance Automatique de la Parole Du signal à son interprétation. Jean-Paul Haton, Christophe Cerisara, Dominique Fohr, Yves Laprie, Kamel Smaïli. DUNOD, 2006, UniverSciences (Paris) Statistical methods for speech recognition. Frederick Jelinek. 1997. MIT press. |

| X3ITM40 | Reconnaissances et synthèse du langage |
|---------------------------------------|---|
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 36h Répartition : CM : 24h TD : 12h CI : 0h TP : 0h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requise(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Reconnaissances et synthèse du langage 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |

| | |
|-------------------------|--|
| Contenu | <p><i>Maitrise du développement d'un système de reconnaissance d'écriture.</i> <i>Maitrise du développement d'un système de transcription.</i> <i>Maitrise du développement d'un système de reconnaissance du locuteur.</i></p> <p>Reconnaissance de l'écriture : (7h CM, 4h TD) les différentes étapes d'un système complet</p> <ul style="list-style-type: none"> • Prétraitements du signal manuscrit • Construction d'un système : segmentation, treillis de segmentation, extraction de caractéristiques, classification en contexte (rejet des mauvaises segmentations), intégration de modèles de langage. • Evaluation et métrique <p>Reconnaissance de la parole : (8h CM, 4h TD)</p> <ul style="list-style-type: none"> • Bref historique : des HMM ou BNN • Construction d'un système : les données, alignement du signal et du texte, phonétisation, apprentissage des modèles • Evaluation et métrique <p>Reconnaissance du locuteur : (7h CM, 4h TD)</p> <ul style="list-style-type: none"> • Définition des tâches en reconnaissance de locuteur : vérification, identification, segmentation et regroupement • Construction d'un système de vérification : les données, apprentissage des modèles, normalisation et décision • Evaluation et métrique <p>Système multimodal : (2h)</p> <ul style="list-style-type: none"> • reconnaissance d'équation écrite et lue • reconnaissance de personne : transcription + reconnaissance du locuteur + reconnaissance de visage |
| Méthodes d'enseignement | CM, mini projet avec un suivi lors de TD |
| Langue d'enseignement | Français |
| Bibliographie | <ul style="list-style-type: none"> • Prince, Simon JD; « Computer vision: models, learning, and inference », 2012, Cambridge University Press • Richard O. Duda, Peter E. Hart, David G. Stork, « Pattern Classification, 2nd Edition », 2000, Wiley, ISBN: 978-0-471-05669-0 • Pierre-Michel Bousquet, Bénéfices et limites des représentations en facteur de variabilité totale pour la reconnaissance du locuteur, Université d'Avignon. • F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska, et D. A. Reynolds, "A tutorial on text-independent speaker verification," EURASIP Journal on Audio, Speech, and Music Processing, vol. 4, pp. 430-451, 2004. • S. Tranter et D. Reynolds, "An overview of automatic speaker diarization systems," IEEE Transactions on Audio, Speech and Language Processing, vol. 14, no. 5, pp. 1557-1565, Septembre 2006. • Harold Mouchère, Richard Zanibbi, Utpal Garain, Christian Viard-Gaudin, "Advancing the state of the art for handwritten math recognition: the CROHME competitions, 2011-2014", Intern. Journal on Document Analysis and Recognition, pp. 173-189, 2016. • Sofiane Medjkoune, Harold Mouchère, Simon Petitrenaud, Christian Viard-Gaudin, "Approches multimodales pour la reconnaissance d'expressions mathématiques » RSTI, Document numérique, pp. 97-122, 2014. |

| | |
|-----------------------------------|--|
| X3IT050 | Apprentissage automatique en langue |
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 36h Répartition : CM : 16h TD : 8h CI : 0h TP : 12h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requise(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Apprentissage automatique en langue 100% |
| Obtention de l'UE | |

| Programme | |
|---------------------------------------|---|
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p><i>Présenter les méthodes d'apprentissage automatique utilisées actuellement et former les étudiants à l'utilisation des méthodes actuelles en apprentissage automatique.</i></p> <ul style="list-style-type: none"> • SVM • Mélanges de modèles gaussiens • Modèles de Markov cachés • Réseaux de neurones (architectures Deep Learning) • Régression linéaire, régression logistique • Méta-classifieurs (boosting, bagging, forêts aléatoires...) • Mise en place d'expérimentation : découpage des données en ensembles d'apprentissage/validation/test + tests de significativité • Utilisation de langages pour le calcul scientifique en TP (Octave, R, Matlab, Python) |
| Méthodes d'enseignement | Cours magistraux, travaux dirigés et projet transverse avec les UE mettant en application des méthodes d'apprentissage automatique. |
| Langue d'enseignement | Français |
| Bibliographie | <p>Antoine Cornuéjols, Laurent Miclet : Apprentissage artificiel. Concepts et algorithmes. Eyrolles, 2010 (2e édition)</p> <p>Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining. Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2011 (3rd edition)</p> <p>Yann LeCun, Yoshua Bengio, Geoffrey Hinton. Deep Learning. Nature, 521, pp436-444, 2015.</p> |

| X3IT010 | Développement logiciel en industrie de la langue |
|---------------------------------------|---|
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 24h Répartition : CM : 5.33h TD : 13.33h CI : 0h TP : 2.67h EAD : 2.67h |
| Place de l'enseignement | |
| UE pré-requise(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Développement logiciel en industrie de la langue 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p>Sensibiliser aux problèmes/solutions d'analyse de contenus textuels non structurés multilingues et multimodaux</p> <p>Sensibiliser aux problèmes/solutions d'interopérabilité technique et sémantique de composants de TALP</p> <p>Sensibiliser aux problèmes/solutions de traitement et stockage de données distribués et échelonnables (scalables)</p> <p>Présentation de cadres de développement de chaînes de traitement intégrant des composants hétérogènes et distribués en favorisant la maintenance et la réutilisation de ressources</p> <p>Donner un aperçu des outils/instruments/framework existants pour les différentes activités du TALP et permettre la prise en main de quelques uns</p> |
| Méthodes d'enseignement | |
| Langue d'enseignement | Français |

| | |
|---------------|--|
| Bibliographie | <ul style="list-style-type: none"> • Tika a content analysis toolkit https://tika.apache.org/ • Hadoop, a distributed computing platform http://hadoop.apache.org/ • Storm a distributed realtime computation system http://storm.apache.org/ • Mahout an environment for quickly creating scalable performant machine learning applications https://mahout.apache.org/ • Elasticsearch is a distributed, open source search and analytics engine, designed for horizontal scalability, reliability, and easy management. https://www.elastic.co/fr/downloads/elasticsearch • "Unstructured Information Management Architecture (UIMA)", Version 1.0, OASIS Standard, 2 March 2009, http://docs.oasis-open.org/uima • UIMA framework for the development and deployment of multi-modal analytics for the analysis of unstructured information content such as text, audio and video https://uima.apache.org/ • Jimmy Lin and Chris Dyer. Data-Intensive Text Processing with MapReduce. Morgan & Claypool Publishers, 2013. http://lntool.github.io/MapReduceAlgorithms/ • Torque distributed resource manager providing control over batch jobs and distributed compute nodes http://www.adaptivecomputing.com/products/open-source/torque/ • The Voice Browser Working Group's mission is to support browsing the web by voice https://www.w3.org/Voice/ • Media Resource Control Protocol https://fr.wikipedia.org/wiki/MRCP • DKPro, a community of projects focussing on re-usable Natural Language Processing software https://dkpro.github.io • Platform for building Python programs to work with human language data http://www.nlTK.org/ • Stanford CoreNLP - a suite of core NLP tools http://stanfordnlp.github.io/CoreNLP/ |
|---------------|--|

| X3IT020 | Insertion professionnelle |
|---------------------------------------|--|
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 24h Répartition : CM : 5.33h TD : 13.33h CI : 0h TP : 2.67h EAD : 2.67h |
| Place de l'enseignement | |
| UE pré-requis(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Insertion professionnelle 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p><i>Savoir gérer un projet Logiciel intégrant plusieurs plusieurs niveaux de participants (client, développeur, designer, ...).</i> <i>Savoir esquisser un Business Model à partir d'un logiciel/d'une innovation.</i> <i>Pouvoir pitcher son idée / son Business Model dans un temps contraint.</i></p> <ul style="list-style-type: none"> • Ingénierie logicielle pour l'industrialisation de prototypes en lien avec l'apprentissage automatique et le traitement des langues • Concevoir une architecture technique complexe impliquant des algorithmes TAL ou ML • Gestionnaire de calcul distribué (torque/maui par exemple), Streaming, Cloud computing... • Retours d'expérience sur l'architecture logicielle de Dictanova et Voxolab/Voxygen • Méthodes agiles pour le développement Logiciel • Modèles économiques du Logiciel : business canvas, empathy map... • Entraînement au pitch |
| Méthodes d'enseignement | |
| Langue d'enseignement | Français |
| Bibliographie | |

| X3IT030 | Algorithmique et alignement de chaînes |
|---------------------------------------|--|
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | FERTIN GUILLAUME |
| Volume horaire total | TOTAL : 24h Répartition : CM : 10.67h TD : 10.66h CI : 0h TP : 0h EAD : 2.67h |
| Place de l'enseignement | |
| UE pré-requis(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Algorithmique et alignement de chaînes 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p><i>Connaissance et acquisition des principales techniques de recherche de motifs dans les textes de grande taille :</i></p> <ul style="list-style-type: none"> • sans pré-traitement, avec pré-traitement sur le texte et/ou le motif • motif exact ou approché <p><i>Connaissance et acquisition des algorithmes associés et de leur complexité.</i></p> <p><i>Comparaison des avantages/inconvénients des techniques présentées, en fonction du contexte d'utilisation.</i></p> <p><i>Connaissances des principaux algorithmes d'alignement bilingue à partir de corpus.</i></p> <ol style="list-style-type: none"> 1. Recherche exacte de motifs <ul style="list-style-type: none"> • Recherche de motifs par fenêtre glissante • Prétraitement du motif • Algorithme de Knuth-Morris-Pratt et variantes • Algorithme de Boyer-Moore et variantes • Correction et évaluation du temps d'exécution • Recherche de motifs dans des textes statiques • Arbres et arbres compacts des suffixes : définition, construction • Automates des suffixes : construction par minimisation de l'arbre des suffixes • Recherche de motifs, de répétitions, de marqueurs dans un ou plusieurs textes 2. Algorithmes d'alignement et Recherche approchée de motifs <ul style="list-style-type: none"> • Transformations : substitution, effacement, insertion • Alignement local, alignement global, alignement pondéré • Distance de Hamming, distance de Levenshtein, distance d'édition pondérée • Algorithme d'alignement de Needleman-Wunch, et sa variante pondérée • Recherche de mot avec k différences. 3. Alignement bilingue <ul style="list-style-type: none"> • Alignement en corpus parallèle • Alignement en corpus comparable |
| Méthodes d'enseignement | Méthodes d'enseignement |
| Langue d'enseignement | Français |
| Bibliographie | <p>M. Crochemore, C. Hancart, T. Lecroq - Algorithms on strings, Cambridge University Press 2007.</p> <p>M. Crochemore, W. Rytter - Jewels of Stringology, World Scientific Publishing, 2002</p> <p>D. Gusfield - Algorithms on Strings, Trees, and Sequences, Cambridge University Press, 1997.</p> <p>Pierre Zweigenbaum et Benoit Habert (2006). Faire se rencontrer les parallèles : regards croisés sur l'acquisition lexicale monolingue et multilingue. GLOTTOPOL :8, 22-44.</p> <p>Gale, William A.; Church, Kenneth W. (1993), "A Program for Aligning Sentences in Bilingual Corpora" (PDF), <i>Computational Linguistics</i>, 19 (1): 75-102</p> |

| X3IT040 | Statistical and symbolic language modeling |
|---------------------|--|
| Lieu d'enseignement | Le Mans/Nantes |

| | |
|---------------------------------------|---|
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | DE LA HIGUERA COLIN |
| Volume horaire total | TOTAL : 24h Répartition : CM : 16h TD : 5.33h CI : 0h TP : 0h EAD : 2.67h |
| Place de l'enseignement | |
| UE pré-requise(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Statistical and symbolic language modeling 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p><i>Savoir mettre en oeuvre et maîtriser différents modèles de langages tant du point de vue de sa définition (puissance d'un modèle), de son utilisation effective (prédiction, lissage) que de son acquisition (apprentissage).</i></p> <ul style="list-style-type: none"> • Les modèles <ul style="list-style-type: none"> - Modèles numériques (à base de vecteurs, essentiellement, donc bag of words et embeddings) - Modèles probabilistes (n-grams & co) - Modèles grammaticaux "syntaxiques" (à noter que les automates et transducteurs probabilistes sont là) - Modèles grammaticaux lexicaux • Utilisation des modèles : <i>parsing</i>, prédiction, techniques associées (lissage). Questions algorithmiques, probabilistes, expérimentales. • Apprentissage des modèles : techniques d'inférence grammaticale, méthodes spectrales, méthodes d'estimation, et réseaux de neurones récurrents. |
| Méthodes d'enseignement | Cours magistraux, travaux dirigés et mini projets |
| Langue d'enseignement | Anglais |
| Bibliographie | Jeffrey Heinz, Colin de la Higuera, Menno van Zaanen: Grammatical Inference for Computational Linguistics. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers 2015 |

| | |
|-----------------------------------|---|
| X3ITM50 | Conférences et cours invités |
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 6h Répartition : CM : 0h TD : 6h CI : 0h TP : 0h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requise(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Conférences et cours invités 100% |
| Obtention de l'UE | Les étudiants doivent obligatoirement assister aux conférences pour valider l'UE. |

| Programme | |
|---------------------------------------|---|
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p><i>Ouverture à des problématiques connexes. Connaître les métiers en lien avec le Master ATAL. Découverte du tissu industriel et des problématiques de recherche du domaine.</i></p> <ul style="list-style-type: none"> • Interventions d'industriels, chercheurs et développeurs, partage d'expérience, description des problématiques en lien avec le traitement du langage |
| Méthodes d'enseignement | |
| Langue d'enseignement | Français |
| Bibliographie | |

| X3IT060 | Corpus |
|---------------------------------------|---|
| Lieu d'enseignement | Le Mans/Nantes |
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 12h Répartition : CM : 10.67h TD : 0h CI : 0h TP : 0h EAD : 1.33h |
| Place de l'enseignement | |
| UE pré-requis(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Corpus 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | <p><i>Constitution d'un corpus : 1) metadonnées, 2) critères de sélection des textes et 3) normes linguistiques d'annotation.</i></p> <ul style="list-style-type: none"> • Définition d'un corpus électronique, • Typologies des corpus : monolingues, bilingues, etc. , • Corpus de références, • Contraintes de constitution d'un corpus : domaines, genres, situation de discours • Metadonnées : DublinCore CES TEI OLAP • Mesures d'évaluation intrinsèque et extrinsèque de la qualité du corpus • Annotation : mesurer la complexité de l'annotation • Codage de la parole, préparée ou conversationnelle • Appréhender les outils d'annotation (transcriber, Webanno) • Guide d'annotations |
| Méthodes d'enseignement | |
| Langue d'enseignement | Français |
| Bibliographie | |

| X3IT070 | Projet |
|---------------------|--------|
| Lieu d'enseignement | |

| | |
|---------------------------------------|--|
| Niveau | Master |
| Semestre | 3 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 11h Répartition : CM : 0h TD : 11h CI : 0h TP : 0h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requis(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Projet 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | |
| Méthodes d'enseignement | |
| Langue d'enseignement | Français |
| Bibliographie | |

| | |
|---------------------------------------|--|
| X4II030 | Période de suivi en formation par alternance |
| Lieu d'enseignement | |
| Niveau | Master |
| Semestre | 4 |
| Responsable de l'UE | |
| Volume horaire total | TOTAL : 0h Répartition : CM : 0h TD : 0h CI : 0h TP : 0h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requis(s) | |
| Parcours d'études comprenant l'UE | M2 Apprentissage et Traitement Automatique de la Langue (ATAL) |
| Evaluation | |
| Pondération pour chaque matière | Période de suivi en formation par alternance 100% |
| Obtention de l'UE | |
| Programme | |
| Objectifs (résultats d'apprentissage) | |
| Contenu | |
| Méthodes d'enseignement | |
| Langue d'enseignement | Français |
| Bibliographie | |

| X4II010 | Stage |
|---------------------------------------|---|
| Lieu d'enseignement | Soutenance à la faculté des sciences et techniques |
| Niveau | Master |
| Semestre | 4 |
| Responsable de l'UE | MOLLI HALA |
| Volume horaire total | TOTAL : 0h Répartition : CM : 0h TD : 0h CI : 0h TP : 0h EAD : 0h |
| Place de l'enseignement | |
| UE pré-requis(s) | |
| Parcours d'études comprenant l'UE | M2 Optimisation en Recherche Opérationnelle (ORO),M2 Architectures Logicielles (ALMA),M2 Apprentissage et Traitement Automatique de la Langue (ATAL),M2 CMI-OPTIM |
| Evaluation | |
| Pondération pour chaque matière | Stage 100% |
| Obtention de l'UE | Il n'y a pas de seconde session pour le stage. Les dispenses d'assiduité ne sont pas autorisées. |
| Programme | |
| Objectifs (résultats d'apprentissage) | <ul style="list-style-type: none"> - S'intégrer dans le milieu professionnel de son stage, mettre en oeuvre sa capacité d'analyse sur le sujet proposé, et être force de proposition. - Synthétiser le travail effectué dans un compte rendu de manière concise, correcte, et complète. - Préparer des supports puis présenter oralement le travail effectué de manière dynamique et convaincante. |
| Contenu | |
| Méthodes d'enseignement | |
| Langue d'enseignement | Anglais |
| Bibliographie | |

Dernière modification par ISABELLE BEAUDET, le 2020-07-02 00:08:58