

UE ERDI - Liste des sujets d'étude

(au total 55 + x + y sujets)

Les sujets sont regroupés en fonction de leur origine, à savoir les différents parcours du master informatique. Cela ne signifie pas nécessairement qu'ils soient tous totalement en relation avec ces parcours, certains sujets pouvant être considérés comme transverses à différents parcours. Vous pouvez donc émettre des vœux pour n'importe quel sujet, du moment qu'il vous intéresse.

En fin de document, vous trouverez des éléments complémentaires et qui décrivent plus précisément certains sujets (pour les sujets étiquetés IDL, IMD et SID).

Sommaire

Page 2 : Sujets du parcours FSI

Page 3 : Sujets du parcours GIG

Page 4 : Sujets du parcours IAAA

Page 6 : Sujets du parcours IDL

Page 7 : Sujets du parcours IMD

Page 8 : Sujets du parcours SID

Page 9 : Compléments sujets du parcours IDL

Page 11 : Compléments sujets du parcours IMD

Page 13 : Compléments sujets du parcours SID

Parcours FSI (11 sujets) – Proposés par Alessia Milani

Les sujets sont rangés en 3 thèmes ci-dessous :

(1) Métiers

En prenant comme point de départ la fiche métiers du CIGREF, il faut décrire en détail les métiers suivants :

FSI-1 : Responsable Sécurité des Systèmes d'Information (RSSI)

FSI-2 : Expert en cybersécurité

FSI-3 : Auditeur SSI

FSI-4 : Délégué à la Protection des Données

Vous trouverez les fiches métiers à cette adresse :

https://www.cigref.fr/wp/wp-content/uploads/2022/08/cigref_nomenclature_rh_des_profils_metiers_du_si_version_complete_2022.3.pdf

(2) Stratégie nationale pour la sécurité du numérique

En prenant comme points de départ les liens suivants

https://www.ssi.gouv.fr/uploads/2015/10/strategie_nationale_securite_numerique_en.pdf

<https://www.ssi.gouv.fr/uploads/2022/04/anssi-avis-migration-vers-la-cryptographie-post-quantique.pdf>

<https://gdr-securite.irisa.fr/les-actualites/>

<https://www.cnrs.fr/fr/cnrsinfo/le-pepr-cybersecurite-devoile-ses-7-premiers-projets-cibles>

décrire la stratégie nationale et les recherches menées actuellement dans les thèmes suivants :

FSI-5 : Cryptographie post-quantique

FSI-6 : Protection de la vie privée

FSI-7 : Méthodes formelles pour la cyber-sécurité

FSI-8 : Sécurité des données dans le cloud

(3) Numérique, Sécurité et Société

En prenant comme points de départ les liens suivants

FSI- 9 : Expliquer les principes de base du RGPD et les outils pour la mise en conformité

<https://www.cnil.fr/professionnel>

FSI-10 : Impacts du RGPD depuis 2018 pour la divulgation de cyber-attaques

FSI-11 : Internet et désinformation

<https://news.cnrs.fr/articles/the-internet-a-disinformation-highway>

Parcours GIG (x sujets) – Proposés par Jean-Luc Mari

Il s'agit ici de réaliser un résumé (rapport écrit en français) et de présenter à l'oral (en français) un article issu de la conférence mondiale SIGGRAPH 2022 (*ACM SIG International Conference on Computer Graphics and Interactive Techniques*) qui est l'une des plus prestigieuses conférences scientifiques du domaine. Vous avez toute liberté de choisir l'un des articles que vous trouverez à partir de ce lien :

<http://kesen.realtimerendering.com/sig2022.html>

Les sujets seront numérotés par GIG-x : avec $x \geq 1$

Parcours IAAA (26 sujets) – Proposés par Alexis Nasr

(1) Calcul sur GPU / Deep learning :

IAAA-1 : GPU NVIDIA dédiés au calcul pour le deep learning

IAAA-2 : Alternatives matérielles à NVIDIA pour le deep learning (puces spécialisées, AMD...)

IAAA-3 : Méthodes et bibliothèques d'accélération des calculs matriciels (gemm, blas...)

IAAA-4 : Les offres cloud avec accélération matérielles GPU

IAAA-5 : Comparatif des toolkits de deep learning - tensorflow, pytorch, etc.

(2) Applications de l'IA

IAAA-6 : IA dans les jeux de stratégie (OpenAI vs Dota 2)

IAAA-7 : Alphago

IAAA-8 : IA en sécurité

IAAA-9 : IA sur téléphone mobile

(3) Éthique et IA

IAAA-10 : Voitures autonomes et problèmes éthiques

IAAA-11 : RGPD et la monétisation des données utilisateur

IAAA-12 : IA en médecine

(4) IA forte et faible

IAAA-13 : Tests pour déterminer qu'une IA est forte

IAAA-14 : Raisonnement automatique et deep learning

IAAA-15 : Historique des méthodes d'IA

(5) Applications de traitement du langage naturel

IAAA-16 : Plateformes pour faire des chatbots et systèmes de dialogue

IAAA-17 : Fonctionnement de la traduction automatique grand public

IAAA-18 : Génération de textes et créativité

IAAA-19 : Depuis Eliza de Joseph Weizenbaum jusqu'à Tay de Microsoft, 50 ans de chatbot : où en sommes-nous ?

IAAA-20 : Transcription automatique de la parole passée dans l'ère industriel : Un problème résolu ? Des verrous technologiques ?

(6) Modèles de traitement du langage naturel

IAAA-21 : Word embeddings et représentations sémantiques :

"Efficient Estimation of Word Representations in Vector Space" - <https://arxiv.org/abs/1301.3781>

"Understanding word vectors" - <https://gist.github.com/aparrish/2f562e3737544cf29aaf1af30362f469>

IAAA-22 - Modèles deep learning et transition énergétique

"Energy and Policy Considerations for Deep Learning in NLP" - <https://arxiv.org/abs/1906.02243>

"Evaluating the Energy Efficiency of Deep Convolutional Neural Networks on CPUs and GPUs" - <https://ieeexplore.ieee.org/document/7723730/>

IAAA-23 - Biais de genre dans les word embeddings

"Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them" - <https://arxiv.org/abs/1903.03862>

"Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings" - <https://arxiv.org/abs/1607.06520>

"Gender Bias in Contextualized Word Embeddings" - <https://www.aclweb.org/anthology/N19-1064/>

IAAA-24 - Le modèle "transformeur"

"What is a Transformer ?" - <https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04>

"The illustrated transformer" - <http://jalammar.github.io/illustrated-transformer/>

"BERT" - <https://arxiv.org/abs/1810.04805>

IAAA-25 - Où en sont les systèmes de question réponse / compréhension de textes ?

"SQuAD 2.0" - <https://arxiv.org/abs/1806.03822>

Les articles indiqués dans le "SQuAD leaderboard" - <https://rajpurkar.github.io/SQuAD-explorer/>

IAAA-26 - Peut-on utiliser le TAL pour détecter les discours haineux sur internet ?

Workshop WOA - <https://www.workshopononlineabuse.com/>

Choisir 2-3 articles parmi ceux-là <https://www.aclweb.org/anthology/volumes/W19-35/>

Parcours IDL (13 sujets) – Proposés par Jean-Luc Massat

Les sujets sont présentés avec un peu plus de détails à la fin de ce document.

IDL-1 : Évaluer la qualité du code source avec SonarQube

IDL-2 : Authentification unique des applications et sites web (SSO)

IDL-3 : Les architectures « serverless »

IDL-4 : Le *principe du* développement « *low-code* »

IDL-5 : La dette technique

IDL-6 : Les assistants d'aide à la programmation (co-pilot de github)

IDL-7 : Architecture logicielle des *applications*

IDL-8 : Architecture des *systèmes d'information*

IDL-9 : Les dernières tendances des langages informatiques et l'impact sur la productivité des développeurs

IDL-10 : Le framework Flutter

IDL-11 : Le langage rust

IDL-12 : La version 14 du framework Angular

IDL-13 : Micro-frameworks Java

Parcours IMD (y sujets) – Proposés par Kevin Perrot

Les sujets sont présentés avec un peu plus de détails à la fin de ce document.

Chaque année, une semaine de cours est donnée à de jeunes chercheurs en informatique, dans le contexte d'une école appelée EJCIM. Ces cours sont des introductions à des sujets de recherche actuels en informatique. Tous les chapitres de cours sont disponibles gratuitement au téléchargement [1], et permettent de couvrir des sujets très variés de l'informatique (algorithmique, combinatoire, géométrie, algèbre, automates...). Votre travail consiste à choisir un chapitre d'une année, à réaliser un résumé (rapport écrit), et une présentation orale.

[1] <https://www.gdr-im.fr/im-photographie/>

Les sujets seront numérotés par IMD-y : avec $y \geq 1$

Parcours SID (5 sujets) – Proposés par Sana Sellami

SID-1 : L'ère de la recommandation : Je recommande ou pas ?

SID-2 : Les métiers de la « data science »

SID-3 : Science de données (SD) et Intelligence artificielle (IA)

SID-4 : Stream processing : Traitement de données en flux

SID-5 : Apprentissage par transfert

Compléments aux sujets IDL

IDL-1 : Évaluer la qualité du code source avec SonarQube

Comment évaluer la qualité d'un code source avec l'outil SonarQube ? D'autres outils du même type pourront être étudiés. Votre présentation devra se baser sur une mise en œuvre de ces outils sur un code intéressant.

Contact : Magali Contensin (magali.contensin@univ-amu.fr)

IDL-2 : Authentification unique des applications et sites web (SSO)

Le but est de réaliser une étude sur les solutions d'authentification unique (Single Sign On). Vous effectuerez une présentation répondant aux questions suivantes :

- Qu'est-ce que l'authentification unique ?
- Quels sont ses avantages ?
- Quels sont ses inconvénients ?
- Quelles sont les différentes catégories de solutions existantes ?
- Quels sont les critères à prendre en compte pour choisir une solution ?
- Comment mettre en œuvre ces solutions dans le code ?

Contact : Magali Contensin (magali.contensin@univ-amu.fr)

IDL-3 : Les architectures « serverless »

Définition, mise en perspective, avantages et inconvénients de cette tendance qui apparaît dans le monde du développement logiciel.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-4 : Le principe du développement « low-code »

Après avoir défini le principe, vous devrez discuter des avantages et inconvénients de cette approche. Vous pourrez étudier les offres de quelques plateformes et présenter les principales fonctionnalités.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-5 : La dette technique

Après avoir défini le concept (accompagné d'exemples), vous pourrez discuter de l'impact de cette dette sur la gestion et le coût des projets informatiques.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-6 : Les assistants d'aide à la programmation (co-pilot de github)

De nouveaux outils à base d'IA sont maintenant disponibles pour assister le développeur dans son travail de conception. Il s'agit de tester un de ces outils et de donner un point de vue critique sur l'impact de ces technologies sur le développement logiciel.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-7 : Architecture logicielle des *applications*

Réaliser une étude et une présentation des principes architecturaux utilisés dans le cadre applicatif (architecture en couches, n-tiers, MVC, etc.) en détaillant les avantages et les inconvénients.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-8 : Architecture des *systèmes d'information*

Réaliser une étude et une présentation des principes architecturaux utilisés dans le cadre des systèmes d'information (architecture orientée service, micro-service, urbanisation, ESB, etc.).

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-9 : Les dernières tendances des langages informatiques et l'impact sur la productivité des développeurs

Vous réaliserez une étude permettant de mettre en avant les avantages et les inconvénients des langages les plus utilisés et leurs domaines de prédilection.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-10 : Le framework Flutter

Réaliser une présentation critique de ce framework (accompagnée d'une mise en œuvre).

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-11 : Le langage rust

Vous réaliserez une étude permettant de mettre en avant les avantages et les inconvénients de ce langage et ses domaines de prédilection. Une mise en œuvre ainsi qu'une comparaison avec le langage C seraient appréciées.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-12 : La version 14 du framework Angular

Après une présentation générale d'Angular, vous détaillerez les nouveautés de la version 14 (ou 15 si elle est disponible). Vous devrez vous appuyer sur une réalisation pratique pour justifier les nouveautés présentées.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

IDL-13 : Micro-frameworks Java

Nous vous proposons de réaliser une présentation synthétique et une comparaison de quelques micro-frameworks de développement d'applications Web en Java. Vous devrez notamment expliquer l'essor de ces architectures et montrer, sur un exemple simple à construire, les avantages et inconvénients de chaque solution.

Contact : Jean-Luc Massat (jean-luc.massat@univ-amu.fr)

Compléments sujets IMD

Sujets proposés par Kevin Perrot :

Chaque année, une semaine de cours est donnée à de jeunes chercheurs en informatique, dans le contexte d'une école appelée EJCIM. Ces cours sont des introductions à des sujets de recherche actuels en informatique. Tous les chapitres de cours sont disponibles gratuitement au téléchargement, et permettent de couvrir des sujets très variés de l'informatique (algorithmique, combinatoire, géométrie, algèbre, automates...). Votre travail consiste à choisir un chapitre d'une année, à réaliser un résumé (rapport écrit), et une présentation orale.

Toutes les ressources sont disponibles sur cette page web :

<https://www.gdr-im.fr/im-photographie/>

Excepté pour l'année 2018 (le lien ne fonctionne pas) et pour 2022 (accessible sur demande directe à K. Perrot), le livre est disponible à cette adresse :

<https://ejcim2018.sciencesconf.org/data/pages/ejcim2018.pdf>

Détail des chapitres :

EJCIM 2015

1. Algorithmes efficaces en géométrie algébrique réelle
2. Algorithmes modérément exponentiels
3. Calcul de programmes parallèles avec Coq
4. Construire et calculer dans un monde 2D
5. Contrôle, probabilités et observation partielle

EJCIM 2016

1. Petit manuel de survie en milieu digital
2. Algèbres de relations
3. Analyse de séquences
4. Modèles de Calcul Quantique
5. Les modèles topologiques

EJCIM 2017

1. Pavages et automates cellulaires
2. Une introduction aux jeux combinatoires
3. Jeux, topologie et automates
4. Une introduction à la réalisabilité classique
5. Analyses d'erreurs en arithmétique flottante

EJCIM 2018

1. Combinatoire et couplage probabiliste
2. Satisfaisabilité propositionnelle et modulo théories
3. Géométrie numérique
4. Analyse statique des réseaux booléens

5. Expérimentation mathématique avec Sage

EJCIM 2019

1. Transductions
2. Quelques méthodes pour les mots sturmiens
3. Géométrie et topologie pour les maillages 3D
4. Popopo - posets, polynômes, polytopes
5. Algorithmique distribuée : systèmes d'agents mobiles

EJCIM 2020

1. Nombre chromatique et sous-graphes induits
2. Accessibilité des systèmes d'addition de vecteurs
3. La combinatoire analytique
4. Calculer avec les nombres réels
5. Convexité combinatoire

EJCIM 2022

1. Introduction à l'IA et aux modèles génératifs
2. De l'échantillonnage adaptatif à la résolution de jeux
3. Calculabilité dans les systèmes dynamiques discrets
4. Quelques propriétés des champs de vecteurs génériques
5. Le calcul analogique
6. Le calcul numérique
7. Aléatoire algorithmique
8. Les structures algébriques de l'aléatoire

Compléments sujets SID

SID-1 : L'ère de la recommandation : Je recommande ou pas ?

La recommandation fait aujourd'hui partie de notre quotidien que ce soit pour acheter des articles sur Amazon, ou bien pour regarder des films, etc. On est tous tenté de consulter les recommandations des autres afin de prendre les bonnes décisions. Mais concrètement comment cela fonctionne ? Quels mécanismes permettent par exemple à Amazon de déterminer les articles qui pourraient correspondre à vos attentes [1] [2] ? Sur quelles données se base-t-il ?

L'objectif de ce sujet est d'essayer de répondre à toutes ces questions en identifiant plus précisément :

- - Les différentes techniques de recommandation [3]
- - Un scénario mettant en place une ou plusieurs méthodes de recommandation.

Références

[1] G. Linden ; B. Smith ; J. York: Amazon.com recommendations: item-to-item collaborative filtering. Journal IEEE Internet Computing archive Volume 7 Issue 1, January 2003 Page 76-80

[2] Brent Smith, Greg Linden: Two Decades of Recommender Systems at Amazon.com. IEEE Internet Computing 21(3): 12-18 (2017)

[3] Roshni K. Sorde, Sachin N. Deshmukh. Comparative Study on Approaches of Recommendation System. International Journal of Computer Applications (0975 – 8887) Volume 118 – No. 2, May 2015

Contact : Sana Sellami : sana.sellami@univ-amu.fr

SID-2: Les métiers de la « data science »

L'objectif de ce projet est d'étudier les métiers liés à la science des données (data scientist).

Plus précisément, il s'agit d'explorer les différents métiers en essayant de répondre aux questions suivantes :

- Quelles sont les exigences et les compétences requises ?
- Quelles formations pour quel métier ?
- Quel est le lien entre l'intelligence artificielle et les données ?
- Quels enjeux pour la recherche ?
- Quels sont les domaines d'application ?

Contact : Sana Sellami : sana.sellami@univ-amu.fr

SID-3 : Science de données (SD) et Intelligence artificielle (IA)

L'objectif est de décrire les domaines de SD et IA, leurs liens et leurs différences au niveau des concepts, approches, objectifs et applications.

Plus précisément, il s'agit de :

- Définir la science des données et l'intelligence artificielle
- Déterminer la frontière entre les deux, les principales différences et les liens qui existent.
- Identifier les techniques

Contact : Djamal Habet : Djamal.habet@univ-amu.fr

SID-4 : Stream processing : Traitement de données en flux

Objectif

Le traitement de données en flux est un paradigme de programmation « Big Data » qui manipule des objets constitués de données ou de séquence d'évènements qui arrivent en continu (temps réel). L'environnement, au sens large, d'un système de traitement de données en flux englobe : la programmation réactive, le traitement distribué des données, des systèmes de gestion de flux, des langages de requêtes, etc.

Le travail attendu pour ce sujet ERDI est d'aller à la découverte de ce domaine en déclinant les principaux composants de cet environnement.

Sont attendus en particulier les éléments suivants :

- Place des données en flux dans le traitement de la data : domaines d'application, use cases.
- Principaux algorithmes, méthodes de traitements
- Éléments d'architecture, pile logicielle
- Panorama des systèmes de stream processing.
- Illustration à travers le synopsis d'une application (un use case) o Pas de codage demandé ou alors à des fins d'illustration.

Références

1. Sherif Sakr, Tilmann Rabl, Martin Hirzel, Paris Carbone, Martin Strohbach. Dagstuhl Seminar on Big Stream Processing, ACM SIGMOD Record (SIGMOD), Volume 47, Issue 3, September 2018, pp 36–39, <https://doi.org/10.1145/3316416.3316426>
2. Timo R  th, Interactive and explorative stream processing. DEBS '22: Proceedings of the 16th ACM International Conference on Distributed and Event-Based Systems, June 2022, pp 194-197 <https://doi.org/10.1145/3524860.3543287>

Contact : Omar Boucelma, LIS-AMU, omar.boucelma@lis-lab.fr

SID-5 : Apprentissage par transfert

L'apprentissage par transfert est une m  thode qui permet de transf  rer les connaissances acquises    partir de la r  solution d'un probl  me donn   pour traiter un autre probl  me.

L'objectif de ce sujet est d'  tudier l'article de la r  f  rence [1] et de r  pondre aux questions suivantes :

- 1 Quelles sont les m  thodes permettant de r  aliser un apprentissage par transfert ?
- 2 Dans quelles applications est utilis   l'apprentissage par transfert et pour r  pondre    quel(s) besoins

Vous illustrerez   galement via un exemple d'application et un jeu de donn  es, l'utilisation de l'apprentissage par transfert et vous examinerez le code github propos   dans l'article.

R  f  rence

1. Fuzhen Zhuang and Zhiyuan Qi and Keyu Duan and Dongbo Xi and Yongchun Zhu and Hengshu Zhu and Hui Xiong and Qing He. A Comprehensive Survey on Transfer Learning. Proc. IEEE 109(1): 43-76 (2021) (<https://arxiv.org/pdf/1911.02685.pdf>)

Contact : Sana Sellami : sana.sellami@univ-amu.fr