

**JOURNÉE DU LIFO - 2023**  
**27 juin 2023**  
**LIFO - Bâtiment IIIA - Amphi H**

**PROGRAMME**

9:00. Accueil

9:20-9:30. Ouverture

9:30-10:50. Le LIFO et les perspectives de recherche autour de la santé

**DOING : données intelligentes** (Mirian Halfeld Ferrari)

Le projet APR-IA DOING vise à développer des méthodes et des outils pour, dans un premier temps, extraire des informations de données textuelles en les structurant dans une base de données graphe, puis pour manipuler de façon intelligente ce graphe de connaissance. Le domaine d'application choisi est le domaine de la santé, avec en premier lieu l'utilisation de données disponibles librement (tels que des cas cliniques). DOING vise à concevoir des requêtes *data science*, c'est-à-dire, une nouvelle forme de requêtes déclaratives, pouvant intégrer des analyses, qui guideront les spécialistes du domaine de la santé dans leur prise de décision. DOING est conçu sur une réelle collaboration interdisciplinaire (Traitement Automatique des Langues, Bases de Données et Intelligence Artificielle) pour transformer des données en information puis en connaissance. Le but du projet est de concrétiser des propositions émanant de DOING groupe de travail du RTR-DIAMS et de DOING action du GDR-MADICS.

**BioSkel : Framework HPC pour la biologie** (Sophie Robert)

A partir d'une IHM définie par les biologistes et en accord avec leurs habitudes de travail, l'objectif du framework BioSkel est de générer une parallélisation des traitements de données pour une architecture cible et qui tient compte également du volume de données impliquées. Ce framework s'appuiera sur différents squelettes de programmation en fonction de classes de traitement, classes qui restent à définir. Les cas d'études concerneront l'alignement de séquences et l'analyse de métabolites. De plus ce projet utilisera une étude clinique du MIT "GALIG-CBD" concernant la recherche de certaines propriétés de phytocannabinoïdes produits par des fleurs de cannabis.

**Panorama de travaux CA liés au domaine de la santé** (Thi-Bich-Hanh Dao, Anais Lefevre-Halftermeyer, Vincent Nguyen et Christel Vrain)

Nous présentons des travaux dans l'équipe CA qui ont un lien avec le domaine de la santé. Ces travaux sont développés dans différents cadres de collaboration : ICVL, RTR Diams, ou projet

ANR, etc. et interviennent du côté praticien et/ou patient. Ils traitent différents types de données: données longitudinales, images, textes, etc.

### **SIOMediC: Sécurité de l'Internet des Objets Médicaux Connectés (Frédéric Louergue)**

La santé digitale et en particulier les Dispositifs Médicaux Connectés (DMC) connaissent un essor fulgurant. Malheureusement, les nombreuses et récurrentes vulnérabilités de ces dispositifs induisent des problèmes de confidentialité majeurs pour les données médicales et des dangers importants pour l'intégrité physique des personnes. Nos travaux visent à rendre les DMC plus sûrs (exempts de bogues internes) et plus sécurisés (résistants aux attaques externes). Pour protéger les DMC des attaques extérieures, nous souhaitons proposer de nouveaux protocoles d'interconnexion des DMC dans leur environnement (hospitalier ou domestique) et des modèles de contrôles d'accès, adaptés à ces différents usages (souvent sans fil). Nous souhaitons également pouvoir apporter des garanties que les codes (programmes) embarqués sur les DMC ne présentent pas de vulnérabilités, ni les protocoles et mécanismes de protection que nous y auront adjoints.

10:50- 11:00. Pause

11:00 -12:00. Le LIFO et les perspectives de recherche autour de la sécurité

### **Activités LMV liées à la cyber-sécurité (Frédéric Louergue)**

La garantie d'absence de bugs dans les programmes vérifiés contribue à la sécurité par la diminution des vulnérabilités exploitables. Cet exposé présentera un aperçu des activités et projets LMV liés à la sécurité, notamment co-financés par l'agence de l'innovation de défense (AID) dans le cadre de son programme cyber-sécurité.

### **Panorama des Nouveaux Projets SDS (Pascal Berthomé)**

L'équipe SDS est impliquée dans divers projets structurels autour de la cybersécurité. Ce panorama reviendra sur le contexte général de l'INSA, du département STI et SDS et de la cybersécurité au niveau régional et national. Il listera les actions dans lesquelles l'équipe est engagée sur cette thématique. Nous nous attarderons sur le projet emblématique cyberINSA.

### **Une micro-introduction à la cryptographie quantique (Ioan Todinca)**

Si l'ordinateur quantique est loin d'être au point, la cryptographie quantique est déjà opérationnelle, et de façon raisonnablement convaincante. Cet exposé présentera brièvement le protocole cryptographique de Bennet et Brassard (BB84), dont le but est de permettre à Alice et Bob d'établir une clé commune, tout en détectant s'ils se font espionner. Il se base sur deux propriétés des qubits : (1) leur aspect « vectoriel », qui permet de mesurer un qubit dans des

bases différentes, et (2) la propriété de non-clonage, qui fait que si un intrus « lit » un qubit il lui est impossible de le remettre dans le circuit à l'identique.

Malgré l'aspect contre-intuitif de ces propriétés, le protocole reste très facile à comprendre.

12:00 - 12:40. Discussions

12:40 - 12:50. Une vision de l'informatique (Sylvie Billot)

12:50 - 13:00. Conclusion de la matinée

13:00 - 14:15 : Déjeuner (buffet)

14:15 -15:30. **POSTERS et DÉMOS postdoc/doctorants/master (programme à part)**

15:30- 15:45. Le LIFO et les relations Internationales (Jérôme Durand-Lose)

15:45- 16:00. Comité à l'Intégrité scientifique et à la Déontologie de la Recherche (Jérôme Durand-Lose)

16:00- 16:15. Le LIFO et la sobriété énergétique (Mostafa Bamha)

16:15 - 16:30. ICVL: nouvelles (Guillaume Cleuziou)

16:30 - 16:35. Clôture

**JOURNÉE LIFO****POSTERS**

<b>AUTEUR</b>	<b>Titre</b>	<b>Résumé</b>	<b>Équipe</b>
Olivier Gracianne	Des clusters de tweets aux tags de descriptions : présentation d'un évènement par la caractérisation de ses manifestations	Notre travail aborde le problème de caractérisation de sous-événements dans des tweets par la description de leur groupement. Nous nous appuyons sur la représentation vectorielle de tweets dans un espace de plongement et les clusterisons. Nous proposons deux méthodes pour sélectionner des candidats parmi les mots des tweets pour construire un ensemble de tags de description. À partir de ces tags, nous proposons de construire une description par cluster en utilisant un modèle déclaratif en programmation linéaire en nombres entiers. Les expérimentations sur un jeu de données réelles montrent l'intérêt de notre approche.	CA
Sofiane Elguendouze	Explainable Image Captioning through Latent Space Perturbation	Most vision-language tasks, such as image captioning, require complex deep learning (DL) models/architectures. However, the high complexity of such models often leads their inner functioning to be regarded as black-boxes, preventing humans from understanding their decision process. Representation space is a key element in DL models, understanding this space and how information is encoded constitute an interesting approach for interpretable DL. Indeed, as far as we are aware, our work is the first presenting a novel method based on perturbation principle of representation space, to study the components that influence DL architectures used in image captioning. The core idea is to isolate and identify the importance of each component (element of the architecture) involved in the captioning pipeline by perturbing, by means of Gaussian functions, the representation space rather than the original space of inputs. We experimentally demonstrate that those components differ in their influence and relevance. The results show that the visual modality would constitute a critical explanation target in captioning models in contrast to language modality, thus leading to more fine-grained explanations. We also propose MSICE, an automatic evaluation metric for image captioning which addresses two important yet overlooked linguistic aspects, morphology and semantic.	CA
Thibaut Martinet	A two levels-model for learning representations of programs	In NLP (Natural Language Processing), most language models are based on the distributional hypothesis, derived from linguistics and according to which two words appearing in similar contexts are semantically close. In this work we are interested in adapting these models to the analysis of programming languages, whose structure is slightly different from that of natural language. Our contribution therefore includes a reformulation of this distributional hypothesis according to two levels of granularity: tokens and statements, followed by the implementation of a representation learning model that takes account of these two levels of analysis.	CA

AUTEUR	Titre	Résumé	Équipe
Shahrzad Heydarshahi	DNA Tile Self-assembly for 3D Surfaces : Genus Identification	We introduce a new DNA tile self-assembly model: the Surface Flexible Tile Assembly Model (SFTAM), where 2D tiles are placed on host 3D surfaces made of axis-parallel unit cubes glued together by their faces, called polycubes. The bonds are flexible, so that the assembly can bind on the edges of the polycube. We are interested in the study of SFTAM self-assemblies on 3D surfaces which are not always embeddable in the Euclidean plane, in order to compare their different behaviors and to compute the topological properties of the host surfaces. We focus on a family of polycubes called cuboids. Order-0 cuboids are polycubes that have six rectangular faces, and order-1 cuboids are made from two order-0 cuboids by subtracting one from the other. Thus, order-1 cuboids can be of genus 0 or of genus 1 (then they contain a tunnel). We are interested in the genus of these structures, and we present a SFTAM tile assembly system that determines the genus of a given order-1 cuboid. The SFTAM tile assembly system which we design, contains a specific set Y of tile types with the following properties. If the assembly is made on a host order-1 cuboid C of genus 0, no tile of Y appears in any producible assembly, but if C has genus 1, every terminal assembly contains at least one tile of Y. Thus, we are able to distinguish the host surfaces according to their genus, by the tiles used in the assembly. This system is specific to order-1 cuboids but the techniques we use should be generalizable to other families of shapes.	GAMOC
Darine Rammal	Synchronous cooperative threading model in MSSL	We propose a synchronous cooperative threading model for a programming language called MSSL. In MSSL, threads are executed cooperatively by a round-robin scheduler. Meanwhile, shared data between threads does not require the locking primitives for protection. This result is achieved by introducing a new type of smart pointers, named Trc, that combines sharing and mutability. Furthermore, using the notion of logical time, called instants, threads can progress synchronously with a consistent view of the state of signals. Signals are powerful, flexible, and have compelling means of communication. In this paper, we present a subset of the formal semantics of MSSL in the format of structural operational semantics.	LMV
Nicolas Hiot	Database with incomplete information: construction from unstructured text and incremental updates	La pharmacovigilance vise à identifier les effets secondaires indésirables des médicaments. Les entreprises pharmaceutiques souhaitent automatiser la transcription des déclarations de cas dans une base de données. Nous considérons les étapes pour passer des textes aux graphes, en nous basant sur des travaux d'extraction d'entités et de relations. Il est démontré qu'il est possible de découper les entités médicales complexes en sous-entités plus simples grâce à des grammaires, des lexiques ou des modèles préalablement entraînés. Cela permet d'extraire des unités telles que des dates, des nombres, des noms de personnes, ainsi que du vocabulaire contrôlé spécifique au domaine médical. Ces entités simples servent ensuite à identifier des entités plus complexes à l'aide de cascades de CRFs. La première partie du travail met l'accent sur la structuration des informations extraites pour constituer une base de données. La seconde partie présente la mise à jour d'une base de données incomplète, en se concentrant sur les graphes. Bien que les travaux se basent sur des bases de données graphes, ils sont également applicables à un modèle relationnel. L'objectif est de parvenir à une base de données qui respecte les contraintes tout en étant minimale, c'est-à-dire sans redondances, même si certaines valeurs sont inconnues. Avec cette seconde partie on s'assure que les informations présentes dans la base de données sont fiables et pertinentes.	PAMDA
Sébastien Rivault	Vers une jointure par similarité d'ensemble extensible en utilisant MapReduce et LSH	La jointure par similarité consiste à retrouver l'ensemble des couples d'objets ayant une distance inférieure à un seuil donné. Nous nous concentrons sur des ensembles en utilisant la distance de Jaccard qui est notamment utilisée pour le nettoyage de données, la détection de textes similaires et le filtrage collaboratif. Elle permet également de réduire l'espace de recherche de similarité entre séquences.	PAMDA

AUTEUR	Titre	Résumé	Équipe
Usman Isah	Towards Intrusion Detection Systems Dedicated to Agriculture Based on Federated Learning	<p>The recent advancements in technologies such as Artificial Intelligence (AI), Internet of Things (IoT), drones, and embedded systems have led to significant changes in industrial systems architectures. This shift from simple systems to complex systems with hundreds of devices has made securing these systems difficult. Intrusion Detection Systems (IDS) are essential tools for securing computer networks and systems against cyber attacks. IDS is an area where AI is increasingly used to detect and respond to malicious traffic on heterogeneous systems that may include IoT, embedded systems, and classic LAN. However, IDS also involves other techniques such as signature-based detection and anomaly-based detection, which are non-AI techniques for detecting intrusions. Machine Learning (ML) and Deep Learning (DL) are powerful techniques that have been explored as a means to automate IDS, giving them an edge to detect new forms of attacks. In this research, we will focus on IDS for networks dedicated to agriculture.</p>	SDS
Clément Lecomte	Environment optimization of Data Centres with Deep RL	<p>Energy considerations become a critical issue for modern man-made systems and the need for efficient eco-responsible solutions ensuring energy savings is crucial. This paper develops an energy aware method allowing to optimize the energy consumption of data centers systems thanks to the introduction of Inhibitor P-Time Petri nets (IP-TPN) and Deep Reinforcement Learning techniques. Indeed, thanks to a schedulability analysis method and being given an energy cost function, the global energy consumed for a particular behavior of the system considered can be computed.</p>	SDS

AUTEUR	Titre	Résumé	Équipe
Shahrzad Heydarshahi	DNA Tile Self-assembly for 3D Surfaces : Genus Identification	We introduce a new DNA tile self-assembly model: the Surface Flexible Tile Assembly Model (SFTAM), where 2D tiles are placed on host 3D surfaces made of axis-parallel unit cubes glued together by their faces, called polycubes. The bonds are flexible, so that the assembly can bind on the edges of the polycube. We are interested in the study of SFTAM self-assemblies on 3D surfaces which are not always embeddable in the Euclidean plane, in order to compare their different behaviors and to compute the topological properties of the host surfaces. We focus on a family of polycubes called cuboids. Order-0 cuboids are polycubes that have six rectangular faces, and order-1 cuboids are made from two order-0 cuboids by subtracting one from the other. Thus, order-1 cuboids can be of genus 0 or of genus 1 (then they contain a tunnel). We are interested in the genus of these structures, and we present a SFTAM tile assembly system that determines the genus of a given order-1 cuboid. The SFTAM tile assembly system which we design, contains a specific set Y of tile types with the following properties. If the assembly is made on a host order-1 cuboid C of genus 0, no tile of Y appears in any producible assembly, but if C has genus 1, every terminal assembly contains at least one tile of Y. Thus, we are able to distinguish the host surfaces according to their genus, by the tiles used in the assembly. This system is specific to order-1 cuboids but the techniques we use should be generalizable to other families of shapes.	GAMOC
Darine Rammal	Synchronous cooperative threading model in MSSL	We propose a synchronous cooperative threading model for a programming language called MSSL. In MSSL, threads are executed cooperatively by a round-robin scheduler. Meanwhile, shared data between threads does not require the locking primitives for protection. This result is achieved by introducing a new type of smart pointers, named Trc, that combines sharing and mutability. Furthermore, using the notion of logical time, called instants, threads can progress synchronously with a consistent view of the state of signals. Signals are powerful, flexible, and have compelling means of communication. In this paper, we present a subset of the formal semantics of MSSL in the format of structural operational semantics.	LMV
Nicolas Hiot	Database with incomplete information: construction from unstructured text and incremental updates	La pharmacovigilance vise à identifier les effets secondaires indésirables des médicaments. Les entreprises pharmaceutiques souhaitent automatiser la transcription des déclarations de cas dans une base de données. Nous considérons les étapes pour passer des textes aux graphes, en nous basant sur des travaux d'extraction d'entités et de relations. Il est démontré qu'il est possible de découper les entités médicales complexes en sous-entités plus simples grâce à des grammaires, des lexiques ou des modèles préalablement entraînés. Cela permet d'extraire des unités telles que des dates, des nombres, des noms de personnes, ainsi que du vocabulaire contrôlé spécifique au domaine médical. Ces entités simples servent ensuite à identifier des entités plus complexes à l'aide de cascades de CRFs. La première partie du travail met l'accent sur la structuration des informations extraites pour constituer une base de données. La seconde partie présente la mise à jour d'une base de données incomplète, en se concentrant sur les graphes. Bien que les travaux se basent sur des bases de données graphes, ils sont également applicables à un modèle relationnel. L'objectif est de parvenir à une base de données qui respecte les contraintes tout en étant minimale, c'est-à-dire sans redondances, même si certaines valeurs sont inconnues. Avec cette seconde partie on s'assure que les informations présentes dans la base de données sont fiables et pertinentes.	PAMDA
Sébastien Rivault	Vers une jointure par similarité d'ensemble extensible en utilisant MapReduce et LSH	La jointure par similarité consiste à retrouver l'ensemble des couples d'objets ayant une distance inférieure à un seuil donné. Nous nous concentrons sur des ensembles en utilisant la distance de Jaccard qui est notamment utilisée pour le nettoyage de données, la détection de textes similaires et le filtrage collaboratif. Elle permet également de réduire l'espace de recherche de similarité entre séquences.	PAMDA

AUTEURS	Titre	Résumé	Équipe	Salle
Valentin Bouvresse et Virgile Crvenka	<b>EASI-GDS</b>	EASI-GDS, a data science interface that puts in practice a declarative machine learning framework over graph databases. <b>WE INVITE YOU TO TRY IT AND GIVE YOUR FEEDBACK ON OUR TOOL</b>	PAMDA	