

DOING@MADICS

Bilan 2021 (version détaillée)

Mirian Halfeld Ferrari Alves (LIFO)
Anne-Lyse Minard Forst (LLL)
Genoveva Vargas-Solar (LIRIS)

Octobre 2021

1 Introduction

DOING aborde l'exploitation intelligente, efficace et sûre des documents par une recherche interdisciplinaire surpassant une simple mise à disposition de données. Deux enjeux de la thématique guident les réflexions proposées par notre action :

- La transformation des données en information.
- La transformation de l'information en connaissances.

La réflexion et le travail sur ces enjeux sont faits dans une perspective multi-disciplinaire avec notamment les domaines du *traitement automatique des langues, des bases de données et de l'intelligence artificielle*. Notre réflexion est, jusqu'à maintenant, associée au domaine d'application de la santé.

Nous rappelons que DOING@MADICS est né d'une extension nationale du groupe de travail régional DOING@DIAMS qui motive des collaborations au sein de la Région Centre Val du Loire.

2 Activités réalisées

L'année 2021 a, comme l'année 2020, été marqué par la crise sanitaire. Notre planning d'activités a ainsi été impacté, comme beaucoup d'autres. Nous l'avons adapté aux circonstances. Dans cette section, nous proposons un bilan organisé par type d'activité ; les activités sont décrites dans un ordre chronologique.

Nous avons adoptée une méthodologie de travail basée sur l'organisation des Webinars avec des présentations invitées et sur la création de deux groupes de travail. Lors des Webinars nous avons consacré du temps à la discussion sur les thèmes des groupes.

2.1 *Webinars*

Cette année nous avons organisé 3 webinars, dont un qui a eu lieu pendant le symposium du GDR MADICS. Le programme des deux premiers webinars a prévu un keynote puis la discussion sur le thème d'un des groupes de travail thématique. Les discussions étaient alimentées par deux présentations courtes. Les discussions dans les groupes de travail étaient riches et nous pensons reproduire ce format lors des prochaines rencontres DOING. Pendant le troisième webinar nous avons eu deux présentations invitées et un tutoriel.

2.1.1 Webinar 1 : 10 mars 2021

Le programme du premier *webinar* DOING 2021 était le suivant ¹:

- 9:00 - 9:30 - Présentation de l'action DOING et du programme de travail de la réunion
- 9:30 – 10:30 - **Keynote : Language-aware indexing for conjunctive path queries**, George Fletcher, Eindhoven University of Technology, Netherlands
- 11:00 – 12:30 - **Groupe de travail sur les requêtes : premières étapes de construction de requêtes déclaratives data science**. Deux courtes présentations permettront d'alimenter les discussions :
 - **Analyse du texte à partir de graphes de connaissances**, Salima Benbernou and Mourad Ouziri, LIPADE, Université Paris Descartes
 - **From source data to data narratives: accompanying users in the way to interactive data analysis**, Veronika Peralta, LIFAT, Université de Tours

Cette première rencontre s'est focalisée sur les interrogations analytiques déclaratives, mais les liens avec l'extraction d'information à partir des données textuelles ont été aussi abordés, notamment au vu des travaux présentés par Mourad Ouziri. Cette rencontre a compté 31 participants (détails dans le tableau 1).

2.1.2 Webinar 2 : 17 juin 2021

Le 17 juin le *webinar* DOING a suivi le programme ci-dessous ² et a compté 39 participants (tableau 1).

- 9:05 – 10:05 - **KEYNOTE: Building Scientific Knowledge Graphs from Scholarly Data**, Davide Buscaldi, Associate Professor, LIPN, Université Sorbonne Paris Nord

¹https://www.univ-orleans.fr/lifo/evenements/doing/?page_id=604

²https://www.univ-orleans.fr/lifo/evenements/doing/?page_id=663

	Institutions des intervenants	Participants
Webinar 1 (10 mars 2021)	Eindhoven University of Technology, Netherlands U. Paris Descartes, LIPADE U. de Tours, LIFAT	31 participants Labs/Instituts (8+) : LIFAT, ESILV, AgroParisTech, LISITE, LIPADE, LIFO, IGN, DISP Lab, etc. UMR (3+) : LIRIS, LISN, LLL, etc. International : ?
Webinar 2 (17 juin 2021)	LIPN, U. Sorbonne Paris Nord ESILV LLL, U. d'Orléans	39 participants Labs/Instituts (9+) : BRGM, LIFAT, ESILV, LIFO, DISP Lab Lyon, LIPADE, DAVID, IRISA, MAS Laboratory, etc. UMR (5+) : LISN, LIPN, LLL, LIP, ATILF, etc. International : ?
Webinar 3 (5 juillet 2021)	LISN CNRS Pontificia Universidad Católica de Chile, Chili Universidade Federal do Amazonas, Brésil	26 participants Labs/Instituts (4+) : BRGM, LIFAT, ESILV, LIFO, etc. UMR (4+) : LISN, LLL, LIP6, GREYC, etc. International : Univ. Manaus Brésil, UFRN Brésil

Table 1: DOING@MADICS' Webinars

- 10:20 – 11:20 - **Groupe de travail sur le TAL en domaine médical.**
Deux courtes présentations permettront d'alimenter les discussions :
 - **Information retrieval on medical data: techniques based on the hybridization of symbolic and statistical methods on text**, Christophe Rodrigues, IT, Big Data and Connected Objects Department, ESILV
 - **Extraction of information in the medical field: extraction of unsupervised relationships**, Anne-Lyse Minard Forst, Université d'Orleans, LLL

Le deuxième webinar s'est focalisé sur les thématiques de la sous-partie TAL-BD. Différents aspects de l'extraction d'information ont été discutés. Les travaux de Davide Buscaldi abordent la construction d'un graphe de connaissance à partir d'articles scientifiques. La méthode utilisée cible le domaine d'application et ne semble pas être directement exploitable pour d'autres domaines. Christophe Rodrigues a présenté des travaux sur la recherche d'information dans le domaine médical. La réflexion a porté sur l'hybridation de méthodes symboliques et statistiques, en particulier des méthodes d'apprentissage profond. Dans DOING nous nous intéressons pas directement à la recherche d'information dans des textes, mais les méthodes présentées peuvent nous aider (combinaison à la fois des vecteurs de mots et des vecteurs de concepts). Pour compléter la discussion autour du TAL en domaine médical, Anne-Lyse Minard a présenté des travaux sur l'extraction de relations, et s'est penchée sur la question de l'extraction de relations non supervisée.

2.1.3 Third Symposium GDR CNRS MADICS: Webinar DOING

Le 5 juillet 2021, dans le cadre du Symposium MADICS, DOING a organisé une demi-journée ³ avec deux *Keynotes* et un tutoriel. 26 participants ont participé

³https://www.univ-orleans.fr/lifo/evenements/doing/?page_id=678

au webinar.

- 14:30 – 13:45 - Introduction et présentation de la demi-journée
- 13:45 – 14:45 - **Keynote : Natural language processing for epidemiology & public health**, Aurélie Névéol, CNRS, LISN, France
- 14:45 – 15:45 - **Keynote : Evaluating navigational queries over graphs**, Domagoj Vrgoc, Pontificia Universidad Católica de Chile, Chili
- 15:45 - 16:00 - Coffee break
- 16:00 – 17:30 - **Tutorial : Say the word, and you'll be free: Methods and Techniques for Natural Language Database Interfaces**, Altigran Soares da Silva, Universidade Federal do Amazonas, Brésil

La première présentation par Aurélie Névéol a consisté en une présentation des problématiques liées au traitement automatique de corpus médicaux, puis en présentant différentes tâches (identification d'entités cliniques, analyse temporelle, etc.) elle a soulevé les problèmes de l'adaptation au domaine et de la reproductibilité des systèmes. L'adaptation au domaine est un point qui nous intéresse particulièrement dans DOING, et nous nous interrogeons sur la généralité des systèmes permettant une adaptation (au domaine et au type de textes) à moindre coût.

La deuxième présentation, par Domagoj Vrgoc de l'Université Catholique du Chili, a été intitulée: *Navigational Queries over Graphs*. La présentation a porté sur les requêtes de graphes qui permettent de traverser des chemins de longueur arbitraire, par opposition aux patrons de taille fixe. Certains résultats théoriques concernant cette classe de requêtes ont été discutés. Ensuite il a montré comment ces résultats ont été pris en charge par les moteurs de bases de données de graphes existants. Enfin, il a discuté sur les algorithmes qu'on pourrait utiliser pour exécuter efficacement de telles requêtes.

Le tutoriel présenté par Altigran da Silva (Universidade Federal do Amazonas) a porté sur le traitement de requêtes en langue naturelle. La maturité technologique dans des domaines tels que l'apprentissage automatique, le traitement du langage naturel et la recherche d'informations, permet d'extraire la sémantique des textes avec une grande précision et efficacité. Dans ce tutoriel, A. da Silva a présenté une vue d'ensemble des méthodes et techniques actuelles qui ont permis d'améliorer les algorithmes et les modèles utilisés pour construire les bases de données en langue naturelle.

2.2 Un *Workshop* associé

En août 2021, le *workshop* DOING@ADBIS⁴, en connexion avec notre atelier DOING@MADICS, a eu lieu comme événement satellite de la conférence ADBIS⁵.

⁴https://www.univ-orleans.fr/lifo/evenements/doing/?page_id=551

⁵<https://adbis2021.cs.ut.ee/accepted-workshops/>

Proposé et organisé par Carmem S. Hara de l'*Universidade Federal do Paraná* (UFPR), Brésil et Mirian Halfeld Ferrari de l'Université d'Orléans, le workshop a reçu 9 soumissions, 3 articles acceptés comme *full papers* et 3 comme *short paper*, ayant ainsi un taux d'acceptation de 50%. Chaque article a été relu par 3 membres du comité de programme. Le comité de programme était composé de spécialistes de différents pays et continents, travaillant dans un des trois domaines phares de DOING : traitement automatique des langues, bases de données et intelligence artificielle.

Le *workshop* a eu lieu par visio-conférence le 24 août 2021, avec environ 15 participants.

Les chairs de DOING@ADBIS comptent re-postuler pour un *workshop* en 2022 au sein de la même conférence.

2.3 Réunions prévues

Nous sommes en train d'organiser des réunions de travail d'ici la fin de l'année.

- Séminaire de Nicolas Travers (ESILV - École d'Ingénieurs Généraliste Leonard da Vinci) au LIFO, suivi d'une discussion pour la mise en place de collaboration avec DOING.
- Journée de travail sur la conception de bases de données graphes pour représenter le contenu des cas cliniques avec Agata Savary (Université Paris Saclay - LISN). Nous allons échanger sur les problèmes associés à la conception des schémas de graphes de données adaptée à la représentation des textes, particulièrement, des cas cliniques.
- Invitation des responsables de l'atelier ROCED de MADICS pour échanger autour de la représentation des connaissances et le lien au texte.
- Visite de Martin Musicante de l'*Universidade Federal do Rio Grande do Norte*, Brésil, comme chercheur invité du LIFO. Dans ce cadre, DOING organise des réunions de travail dans le but d'entamer de nouvelles étapes dans les recherches en partenariat avec Martin Musicante, qui est un collaborateur de longue date de Genoveva Vargas et Mirian Halfeld Ferrari.

3 Stages et Groupes de travail

Dans le cadre des appels régionaux, nous avons eu l'occasion de postuler pour le financement des stages. Les propositions faites avaient comme but de continuer des collaborations inter-thématiques débutées l'année dernière. Les stages ont ainsi permis de coopérer à travers des sujets pratiques entre nos laboratoires et d'avoir des expériences concrètes sur les objectifs et les problèmes abordés dans le cadre de DOING, en général. Avec les stages, nous pensons aussi contribuer à la formation d'étudiants par le recherche, en espérant, pour les plus jeunes, surtout, de les éveiller à nos thématiques.

3.1 Stage : Extraction de relations

Nous avons obtenu un financement par la fédération ICVL⁶ (Informatique Centre Val de Loire) pour un stage de 6 mois sur la thématique de l'extraction de relations entre entités dans le domaine médical. Le stage a été encadré par Nicolas Hiot (LIFO, Orléans), Mirian Halfeld Ferrari (LIFO, Orléans), Anne-Lyse Minard (LLL, Orléans) et Agata Savary (LIFAT, Tours), tous les quatre membres de DOING (@MADICS et @DIAMS). Nous avons recruté une étudiante de M2 (master Sciences du langage, parcours Industrie de la langue de l'université de Grenoble), Alena Silvanovich.

Nous avons travaillé à partir du corpus DEFT 2020 (que nous avons déjà utilisé dans le cadre d'un stage en 2020) qui est composé de cas cliniques rédigés en français et annotés avec des entités médicales (examens, traitements, signes ou symptômes, doses, etc.). Il a d'abord fallu définir les relations d'intérêts, puis annoter une sélection de ces relations (5 relations : moyen, mesure de, accompagne, révèle/recherche/teste, localisation de) dans un sous corpus (14 textes) et enfin développer un système d'extraction automatique. Le système repose sur un ensemble de règles qui s'appliquent principalement à l'analyse en dépendances des phrases d'un texte. Le système a été évalué sur 5 documents du sous-corpus annoté et obtient des résultats très bons : une précision de 0,94, un rappel de 0,83 et une F-mesure de 0,89.

Le stage a conduit à la réalisation (i) d'un état de l'art sur l'extraction de relations dans le domaine médical ; (ii) l'annotation d'un sous-corpus de 14 documents avec 5 relations ; (iii) au développement d'un système à base de règles pour l'extraction de relations entre entités cliniques.

Ce stage nous a également amené à travailler sur la définition d'un schéma de base de données graphe à partir des informations que nous étions capable d'extraire automatiquement dans les textes.

Nous prévoyons de présenter ce travail dans un article pour le workshop DOING'2022.

3.2 Stage : Karim Dao

Encadrement d'un stage de 6 mois de l'étudiant Karim Dao, du master Data Science de l'Université Paris Daupine à Tunis. Ce stage a été dirigé par Genevieve Vargas-Solar (LIRIS, Lyon) et Mirian Halfeld Ferrari (LIFO, Orléans). Il a été financé par une bourse de la région Afrique.

Ce projet de recherche, qui sera soutenu en Novembre, a eu comme objectif la proposition d'une méthode semi-automatique pour traiter les questions des experts métier, exprimées en langue naturelle pour dériver des patrons ("templates") de requêtes science des données qui fournissent des réponses à ces questions. Nous proposons NLDS-QL un évaluateur de requêtes "classique" et data science qui portent sur les bases de données graphe. L'évaluateur NLDS-QL a été validé dans un cas d'usage associé au domaine médical.

⁶<http://www.info.univ-tours.fr/ICVL/>

3.3 Stage : Cypher et l’analyse de données

Encadrement du stage de 10 semaines de l’étudiant Mathis Dupuis, de l’IUT d’Orléans. Ce stage a été dirigé par Mirian Halfeld Ferrari et Jacques Chabin (LIFO, Orléans) membres de DOING (@MADICS et @DIAMS). Il s’agissait ici d’un stage d’éveil à la recherche comprenant une introduction aux bases de données graphe, au langage cypher et aux algorithmes d’analyse comme *betweenness* et *Page rank*. Le stage nous a permis d’étudier en détail certains algorithmes et de comprendre les implémentations choisies par Neo4J.

3.4 Groupe de travail : Classification de documents

Dans le cadre d’un groupe de travail sur le Traitement Automatique des Langues dans le domaine médical rassemblant des personnes du LLL et du LIFO, nous avons participé à l’édition 2021 de DEFT (Défi Fouille de Textes)⁷. La tâche qui nous a intéressée portait sur l’identification du profil clinique du patient à partir de cas cliniques. Le système que nous avons développé est arrivé premier (sur 7 participants). Une communication orale a été faite à l’atelier DEFT 2021 (organisé en ligne) le 28 juin 2021 et un article a été publié dans les actes de l’atelier (Hiot et al., 2021).

3.5 Groupe de travail : Recherche d’information sur des données médicales

Ce groupe de travail rassemble des collègues qui travaillent sur le croisement entre la recherche d’information, le traitement de la langue naturelle et les bases de données (LIPN, LIRIS, ERIC, ESILV). Les présentations de Davide Buscaldi du LIPN et de Christophe Rodrigues de l’ESILV ont introduit des approches data science utilisant par exemple des méthodes symboliques et statistiques pour traiter le texte. Nous nous intéressons particulièrement à ces techniques dans le cadre de la conception des bases de données graphes et le traitement de requêtes data science appliquées aux données médicales.

4 Les contacts

Comme déjà mentionné lors de nos commentaires évaluant les *webinars* DOING, nous avons pu, pendant cette année, identifier des collègues intéressés et intéressants qui peuvent amener des points de vues riches et divers à notre groupe thématique.

Un travail en amont avec des spécialistes dans le domaine de la santé, rendant plus clair les besoins dans nos domaines différents, a été entamé. Via un questionnaire, nous souhaitons caractériser le profil des médecins de notre réseau, leurs contextes et leurs attentes. Nous avons actuellement des échanges avec quelques médecins pour affiner notre questionnaire, questionnaire que nous

⁷<https://deft.lisn.upsaclay.fr/2021/>

souhaitons diffuser plus largement à partir de janvier 2022. Nous avons des contacts avec des médecins au Brésil, Mexique, USA et Espagne, à qui nous demanderons de faire circuler le questionnaire à leurs collègues. Nous voyons ici non seulement une opportunité d'échange avec les spécialistes de la santé mais aussi une ouverture internationale très intéressante qui vient renforcer les liens que nous avons en particulier avec le Brésil et le Mexique.

5 Projets

5.1 Projet région: APR-IA

Avec les participants à l'action DOING de la région Centre Val de Loire (laboratoire LIFAT, LIFO et LLL), nous avons répondu à un appel à projet APR-IA de la région Centre Val de Loire.

Contexte et positionnement : Le projet DOING (Données Intelligentes) a comme ambition la concrétisation des propositions scientifiques qui émanent du groupe de travail DOING du RTR DIAMS. DOING se construit autour d'une recherche multidisciplinaire impliquant trois laboratoires de nos deux universités (LIFO, LIFAT et LLL), sur trois sites. Le groupe DOING réunit des chercheurs intéressés par la transformation des données en information et puis en connaissance. [...] Le projet DOING est structurant sur trois aspects chers à l'appel APR-IA : (A) la consolidation d'un travail de recherche multidisciplinaire, multi-établissement et multi-sites autour de la manipulation intelligente données; (B) l'intégration de ce travail à un réseau national (action dans le GDR MADICS, collaborations en cours avec le LIRIS-Lyon, le LIPADE-Paris 5, le LATTICE-ENS-Paris), ainsi qu'à un réseau international (workshop à AD-BIS, collaborations en cours avec le Brésil); (C) le tremplin pour concrétiser certaines propositions scientifiques du groupe et se préparer pour des soumissions nationales ou internationales dans la suite.

Le projet est découpé en trois tâches :

- Tâche 1 - Extraction des informations dans les données textuelles (TAL/BD)
- Tâche 2 - Requêtes data science : langage et algorithmes (BD/IA)
- Tâche 3 - Analyse et prédiction de besoins des médecins (BD/IA/TAL)

Le projet a été retenu, néanmoins, la région CVL ré-évalue son enveloppe budgétaire et des coupures financières plus importantes que celles déjà demandées dans la première phase de sélection sont à craindre. Nous attendons des informations supplémentaires.

Le montage de ce projet a été l'occasion d'étendre nos collaborations en région. Ainsi, le groupe de travail TAL se voit renforcé par des collaborations avec des collègues linguistes du laboratoire LLL et les échanges BD-IA se voient enrichis par l'implication de nouveaux collaborateurs, chercheurs dans le domaine de l'IA au LIFAT. La tâche 2 du projet est entièrement conçue sur cette nouvelle collaboration.

5.2 Participation au projet région JUNON

Dans le contexte du dispositif *Ambition Recherche Développement Centre-Val de Loire*, l'ARD-CVL JUNON est un ambitieux programme de développement d'un pôle de recherche numérique sur l'environnement continental (agricole, urbain, forestier et fluvial), en région. Ce programme a été retenu, mais reste en attente des crédits de la région. Piloté par le BRGM (Bureau de recherches géologiques et minières), ce pôle vise à concevoir des services numériques pour améliorer le suivi et la compréhension de l'environnement, pour une meilleure gestion des ressources naturelles. Le LIFO y participe en particulier avec l'engagement de certains membres de DOING. Cette participation concerne le volet *interrogation intelligente* et JUNON représente une opportunité pour le développement des recherches sur les requêtes *data science*. Nous espérons obtenir le financement d'une thèse via ce programme.

6 Bilan scientifique (2021) et éléments de perspectives

Le progrès de DOING en 2021 se fait sentir dans les directions de travail et des résultats que nous décrivons dans la suite. Ces résultats sont la conséquence de notre politique de groupes de travail et co-encadrements de stages, permettant une plus grande interaction entre les membres DOING.

Extraction d'information à partir des textes. De nouveaux algorithmes pour l'extraction d'information à partir des textes du corpus DEFT ont été proposés, notamment pour l'extraction de cinq types de relations.

Interrogation et analyse de graphes à partir de requêtes en langue naturelle. Dans la recherche d'information, les bases de données et dernièrement la *data science*, le traitement des requêtes en langue naturelle est un problème d'actualité qui vise à rendre accessible l'interrogation, l'exploration et l'analyse de données à des personnes avec peu de connaissances en langages d'interrogation formels. Afin de rendre l'interrogation et l'analyse de données et en particulier de graphes aux expert.e.s métier, nous avons travaillé sur la proposition d'une méthode semi-automatique pour traiter les questions de recherche et les transformer vers des requêtes classiques ou *data science* exécutables sur un gestionnaire de bases de données graphes.

Perspectives Nos avancées montrent que plusieurs défis, intrinsèques au traitement de la langue naturelle, restent d'actualité. En outre, un défi majeur concerne le mapping entre les informations extraites et un schéma de base de données utile à l'utilisateur final. La construction d'une instance de base de données à partir d'un texte exige des décisions complexes méritant une analyse assez globale, tant les liens sémantiques sont disséminés dans le texte. La

définition de la bonne granularité que la base doit représenter est intimement liée à l’usage envisagé en avant.

La proposition du projet APR-IA a permis l’élargissement de nos collaborations avec le laboratoire LIFAT à Tours. Dans ce cadre, nous pensons focaliser nos premiers efforts sur l’analyse prédictive - terme qui combine la gestion des données, l’apprentissage automatique et l’optimisation, et qui témoigne de la demande croissante de tels outils pour le traitement des problèmes de sciences des données.

Enfin, DOING a entamé la construction d’un réseau de collaboration avec des spécialistes dans le domaine de la santé. C’est via ce réseau, fondamentalement international, que nous comptons caractériser les attentes de l’utilisateur final. Les méthodes pour atteindre notre but ont été décrites auparavant et nous espérons pouvoir cueillir des informations qui guident la construction de notre base et nos requêtes d’analyse.

Activités à venir

1. Pour la partie extraction d’information, nous faisons face aux défis intrinsèques au traitement de la langue naturelle : traitement de la temporalité, coréférence, distinction entre des actions prescrites et celles effectivement exécutées, etc.
2. Dans la construction d’une instance d’une base de données à partir des textes, un défi majeur concerne le mapping entre les informations extraites et un schéma de base de données utile à l’utilisateur final. La construction d’une instance de base de données à partir d’un texte exige des décisions complexes méritant une analyse assez globale, tant les liens sémantiques sont disséminés dans le texte. La définition de la bonne granularité que la base doit représentée est intimement liée à l’usage envisagé en aval.
3. La construction d’un réseau de collaboration avec des spécialistes dans le domaine de la santé (étape que nous avons entamée comme expliqué auparavant). Via ce réseau, fondamentalement international pour l’instant, nous comptons caractériser les attentes de l’utilisateur final. Cette enquête auprès des utilisateurs cibles est importante pour, d’un côté définir l’importance de chaque type d’information à stocker et, de l’autre côté, comprendre les analyses de données à privilégier.
4. Dans le contexte des requêtes *data science*, nous envisageons, en 2022, de nous focaliser sur l’analyse prédictive. Nous espérons combiner la gestion des données, l’apprentissage automatique et l’optimisation. Cela est la base de notre nouvelle collaboration avec le LIFAT.
5. À plus long terme, il serait très important, d’établir une typologie de cas d’usage, pour aller vers la construction de solutions plus génériques. Ici, la question consisterait à proposer un noyau commun à un type de cas d’usage, pour proposer des solutions à deux couches : un noyau et une couche de personnalisation.

7 Communication DOING

- Site DOING@MADICS : <https://www.madics.fr/ateliers/doing/>
- Site DOING@DIAMS : <https://www.univ-orleans.fr/lifo/evenements/doing/>
- Twitter : <https://twitter.com/NetworkDoing>
- Canal Slack : https://join.slack.com/t/doing-madics/shared_invite/zt-eg5yg240-pqu-Q2igaD0Xs10pmz00Hg

8 Bilan financier

A cause des mesures sanitaires, toutes nos réunions ont été réalisées de manière virtuelle et nous avons utilisé une partie du budget accordé à DOING pour l'année 2021 (1500 euros). Les montants donnés ici sont approximatifs car certaines dépenses sont en cours.

- Missions (120 euros): prise en charge d'une mission pour assister à la journée MADICS à Paris (Mirian Halfeld Ferrari-Alves)
- Achat de livres - 550 euros
- Visites courtes inter-laboratoires (500 euros encore à dépenser): nous avons fait un appel pour le financement de visites courtes pour promouvoir l'interaction entre les participants de DOING.
- Journées de travail (ca. 240 euros): nous avons organisé deux journées hybrides à Orléans. Pour chacune des journées nous avons financé l'invitation d'un.e collègue pour travailler sur les aspects TAL et les aspects interrogation de graphes à Orléans. *Les repas des collègues d'Orléans n'a pas pu être pris en charge par DOING.*

9 Publications associées à DOING

1. Nicolas Hiot, Anne-Lyse Minard, Flora Badin. DOING@DEFT : utilisation de lexiques pour une classification efficace de cas cliniques. Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Atelier Défi Fouille de Textes, DEFT@TALN 2021, Lille, France, June 28 - July 2, 2021.
2. Jacques Chabin, Cristina Dutra de Aguiar Ciferri, Mirian Halfeld Ferrari Alves, Carmem S. Hara, Raqueline R. M. Penteado: *Role-Based Access Control on Graph Databases*. SOFSEM 2021: 519-534

3. Genoveva Vargas-Solar, Pierre Marrec, and Mirian Halfeld Ferrari Alves: *Comparing graph data science libraries for querying and analysing datasets: towards data science queries on graphs*. 3rd Workshop STRAPS, in conjunction with ICSOC 2021 (submitted)
4. Karim Dao, *Requêtes sciences de données pour des bases de données à graphes : des besoins applicatifs à la construction de requêtes*, Master Thesis, Université Paris Dauphine Tunis, Novembre, 2021