



Bilan 2024 de l'Action DOING

ANIMATION SCIENTIFIQUE (merci de remplir le tableau détaillé en annexe)

Réunions de groupes de travail (nombres de réunions, nombre moyen de participant·e·s)

Possiblement réunions internes de l'Action/Atelier

- Environ une réunion tous les deux mois des coordinatrices de l'action (3 personnes).
- Réunions hebdomadaires autour des tâches liées aux projet APR-IA DOING (LIFO, LIFAT, LIRIS)

Ecoles d'été/automne..., semaines de formation, hackathons (nombre de jours, nombre de participant·e·s)

Journées, évènements workshops (nombre de journées, nombre de participant·e·s moyen)

Ateliers dans des conférences ... Evènement très ouvert...

- Workshop international - satellite à ADBIS'2024 (5eme édition)
 - DOING 2024 : 5 articles complets acceptés. Chaque article a été évalué par 3 membres du comité de programme. Le programme final présentait des travaux dans les domaines de l'extraction d'informations à partir de données textuelles, de la classification de textes, de l'analyse de données, ainsi qu'une revue sur l'intégration de règles logiques dans des plongements de graphes de connaissances.
 - Environ 15 participants
- Atelier DOING au symposium MADICS 30 mai
<https://www.madics.fr/event/symposium-madics-6/#doing>

Autres

Nous avons promu des actions de collaboration internationale avec le Brésil. Dans ce cadre, nous avons soumis un projet CAPES-COFECUB en partenariat avec Paris Dauphine, l'Université d'Orléans, le BRGM et le LIRIS, ainsi qu'avec l'Université de São Paulo, l'Université de Campinas, l'Université Fédérale de Rio Grande do Norte et l'Université Fédérale du Paraná. Ce projet porte sur les thématiques développées par l'équipe DOING.

FAITS MARQUANTS

Faits marquants scientifiques de l'année 2024 (articles et résultats majeurs publiés par la communauté MaDICS, distinctions pour des chercheurs ou chercheuses, prix pour des articles de la communauté MaDICS, ...)

Faits marquants sociétaux de l'année 2024 (communication grand public, autre, sur des travaux ou des résultats de membres du GDR)

Nouvelles collaborations mise en place durant l'année 2024

- BRGM, Geosciences, textual and graph-driven approaches
- Université Fédérale do Paraná, Biodiversité, Gestion de l'eau (projet PRINT RESMA)

ACTIONS CIBLEES



Activités en faveur des doctorantes et doctorants (*rencontres dédiées, mise en avant de résultats, écoles, émergence/prix, sensibilisation au métier de la recherche, à la reproductibilité, à la médiation scientifique, à l'éthique en science ...*)

Activités en faveur des permanentes et permanents nouvellement recrutés (*rencontres, valorisation des travaux, recensement, soutien à des projets scientifiques, intégration dans des groupes de travail...*)

Actions en faveur de l'interdisciplinarité au-delà de MaDICS (*GT communs à plusieurs GDR, journées co-organisées...*)

- Rencontres avec des chercheurs du BRGM, liens avec l'ARD JUNON
- Rencontres avec des chercheurs Brésiliens dans le domaine de la biologie

Activités en lien avec les partenaires issus du monde socio-économique (*journées, recensement, partenariat, note de prospective ...*)

Actions collectives (*ressources mutualisées utiles à la communauté : logiciels, plateformes, jeux de données, annuaire, cartographie, recensement production scientifique, recensement de thèses soutenues, production de documents de bonnes pratiques...*)

VEILLE SCIENTIFIQUE, PROSPECTIVE

Au travers des activités de l'atelier et de l'action DOING, nous avons pu identifier les problématiques et défis majeurs liés à l'extraction d'informations et à leur transformation en connaissances à partir de contenu textuel. Nous avons mis en évidence les défis de représentation et de modélisation, notamment avec les graphes, qui apparaissent comme des structures puissantes pour représenter les connaissances extraites des textes et pour modéliser des informations, tout en explorant les textes pour y découvrir des contenus correspondant à ces représentations abstraites. Nous avons également observé le rôle crucial des modèles d'IA dans des pipelines visant à découvrir des informations potentiellement cachées, que ce soit directement dans le contenu textuel ou via les graphes qui modélisent ce contenu.

Cette exploration nous a permis de questionner l'importance de la persistance pour enrichir et compléter les processus d'analyse, qui sont souvent réalisés en mémoire vive dans les solutions existantes. À travers ces réflexions scientifiques, nous avons identifié une communauté nationale et internationale qui contribue activement à ces problématiques d'extraction et de transformation des connaissances textuelles. De plus, grâce à des exemples concrets dans les domaines de la santé et de la géologie, nous avons pu identifier des cas d'usage pertinents pour les défis et approches étudiés dans le cadre de DOING.

Sur une période de cinq ans, la recherche et la technologie ont également évolué, et nous observons aujourd'hui l'émergence de nouveaux systèmes de gestion de bases de données basés sur des représentations vectorielles issues de l'analyse du contenu textuel. Ces avancées renouvellent les questions d'interrogation et de gestion efficace, ouvrant de nouvelles perspectives pour les communautés de la linguistique, des bases de données et de l'intelligence artificielle.

Journées ou écoles thématiques ciblées (*préciser le programme, le nombre de participant-e-s...*)

Participations variées au débat public



Documents de prospective produits (verrous y compris dans le domaine d'application, enjeux du domaine, enjeux sur l'interdisciplinarité, ... mettre les liens si les documents sont en ligne)

1. Mirian Halfeld Ferrari Alves, Anne-Lyse Minard-Frost and Genoveva Vargas-Solar, Transforming Text into Knowledge with Graphs: Report of the GDR MADICS DOING Action. In the Proceedings of the 3rd International Workshop on Knowledge Graphs Analysis on a Large Scale (K-GALS@ADBS), 28th August 2024, Bayonne, France.
2. Genoveva Vargas-Solar, Processing the Narrative: Innovative Graph Models and Queries for Textual Content Knowledge Extraction, Electronics 2024, 13, 3688.
<https://doi.org/10.3390/electronics13183688>, 2024 September 17 (Q2 SJR 2024)
3. Bouvresse, V., Chabin, J., Crvenka, V., Alves, M. H. F., Vargas-Solar, G., & Wang, L. (2024, January). Vers des requêtes déclaratives en science des données: EASI-GDS pour Neo4J. In EGC 2024 (pp. 433-440).
4. Construction automatique de bases de données pour le domaine médical : Intégration de texte et maintien de la cohérence. Nicolas Hiot. PhD thesis. Université d'Orléans, LIFO;
<https://theses.fr/s265149>
5. From Text to Databases: attribute grammar as database meta-model. Jacques Chabin, Mirian Halfeld-Ferrari, Nicolas Hiot. Preprint: <https://arxiv.org/abs/2410.09441>

ANNEXE : Listes et intitulés des réunions organisées ou co-organisées par le GDR

Intitulé de la réunion	Objectif/URL du programme	Organisateur ou organisatrice	Visio / présentiel	# de participant·e·s	Date
DOING Symposium MADICS	Programme	Coordinatrices DOING	Présentiel	~40	30 mai 2024 - 9:00 - 11:00
DOING@ADBS	https://www.univ-orleans.fr/lifo/evenements/doin g/?page_id=1259	Mirian Halfeld Ferrari Alves, U. Orléans, Fr Cristina Aguiar, USP, Br Carmen S. Hara, UFPR, Br	Présentiel	~15	28 août 2024