

<https://obtic.sorbonne-universite.fr/actualite/club-de-lecture-ml4dh-appel-a-participation/>



ceres



Club de lecture ML4DH

Sorbonne Université

Présentation

Le club de lecture ML4DH rassemble des chercheurs, des enseignants, des doctorants et postdoctorants de Sorbonne Université et d'autres établissements et organismes qui s'intéressent à l'intersection de l'apprentissage automatique et des humanités numériques. Le club se réunit régulièrement pour discuter d'articles scientifiques récents en rapport avec ces sujets. Il est organisé par le [SCAI](#), le [CERES](#) et l'[ObTIC](#).

L'objectif du club de lecture ML4DH est de se tenir au courant des derniers développements dans ce domaine interdisciplinaire et d'évaluer de manière critique les recherches publiées. Au cours des réunions, les participants résumeront un article et discuteront ensuite de ses forces, de ses faiblesses et de ses implications pour les recherches à venir.

En plus des discussions, le club organisera d'autres activités pour promouvoir échanges et collaborations. Ces activités pourront inclure par exemple des présentations par des membres du club ou des invités sur leurs projets de recherche actuels, des ateliers ou des séances de formation pour acquérir de nouvelles compétences ou techniques liées à l'apprentissage automatique et aux humanités numériques, des projets numériques ou des hackathons pour explorer de nouvelles applications de l'apprentissage automatique dans les humanités, des conférences par des experts invités.

Format

Chaque participant proposera à l'avance un article scientifique à présenter. Il ou elle pourra ensuite détailler les points suivants:

- Situation du travail présenté par rapport à l'état de l'art.

- Méthodologie, outils et ressources employés.
- Résultats obtenus.

Les échanges peuvent naturellement apporter un regard critique vis-à-vis de la recherche présentée.

Programmation - 2024-2025

Date	Horaire	Lieu	Intervenant(e) + Sujet / Article à présenter
17/10/24	13h-14h	Serpente (D117)	Marceau Hernandez: "A Fast Alignment Scheme for Automatic OCR Evaluation of Books" (Ismet Zeki Yalniz, R. Manmatha ; ICDAR 2011) Thématiques : OCR, évaluation de qualité des textes, alignement
14/11/24	13h-14h	SCAI (salle Conseil)	Mikhail Biriuchinskii: correction automatique de textes océrés: "Leveraging Large Language Models for Post-OCR Correction of Nineteenth-Century British Newspapers", Alan Thomas, 2024.
12/12/24	13h-14h	Serpente	Gaël Lejeune: "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers" (Marco Siino , Ilenia Tinnirello , Marco La Cascia; Information Systems N° 121, 2024) Thématiques :Pré-traitement de textes, Correction, Classification
09/01/25	13h-14h	SCAI salle Conseil)	Floriane Chiffolleau: reconnaissance automatique de textes (état de l'art)
06/02/25	13h-14h	Serpente	
06/03/24	13h-14h	SCAI (salle Conseil)	Motasem Alrahabi: "Difficulty Estimation and Simplification of French Text Using LLMs" (Henri Jamet, Yash Raj Shrestha, Michalis Vlachos, 2024).
	13h-14h	Serpente	
	13h-14h	SCAI	

Participation: les personnes intéressées pour présenter un article peuvent nous écrire ou bien s'inscrire directement dans ce [document](#) en précisant leur adresse mail, le thème traité et le lien de l'article.

Assister sur place: l'entrée est libre, mais en raison du nombre limité de places, merci de bien vouloir compléter ce formulaire :

<https://framaforms.org/inscription-seminaire-ml4dh-du-1710-1727099273>

Plan et accès:

- SCAI (voir [PJ](#))
- Serpente : Maison de la Recherche 28 rue [Serpente](#).

Suivre en ligne: voici le lien permanent de la visio:

https://rendez-vous.renater.fr/muted_user100/ML4DH_Sorbonne-Universite_80c11f-83b4d4-8c8032

Contactez les organisateurs:

- motasem.alrahabi@sorbonne-universite.fr (OBTIC)
- gael.lejeune@sorbonne-universite.fr (CERES/STIH)
- xavier.fresquet@sorbonne-universite.fr (SCAI)

Programmation - printemps 2023

Date	Horaires	Intervenant(e) + Sujet/Article	Adresse mail
9 février	14h-15h30	1. Ljudmila Petkovic -> Détection de phénomènes intertextuels	ljudmila.petkovic@gmail.com
		2. Xavier Fresquet -> Linked Open Images: Visual similarity for the Semantic Web	xavier.fresquet@sorbonne-universite.fr
23 février	14h-15h30	1. Johanna Cordova -> mBERT	johanna.cordova@sorbonne-universite.fr
		2. Hamed Rahimi: <i>Document Embedding-Based Neural Topic Models and Topic Evolution</i> -> Top2Vec , BERTopic , ANTM  Document Embedding-Based Neural...	hamed.rahimi@sorbonne-universite.fr https://github.com/hamedR96/ANTM
9 mars	14h-15h30	1. Riccardo Barontini The Rise and Fall of Biodiversity in Literature, And Biodiversity Is Not Declining in Fiction	barontini.riccardo@gmail.com
6 avril	14h-15h30	1. Gaël Lejeune -> Emerging trends: Subwords, seriously? (Kenneth Church)	gael.lejeune@sorbonne-universite.fr
13 avril	14h-15h30	1. Marguerite Bordry → Franco Moretti, <i>False movimento</i> (Nottetempo, 2022) / sur le Distant Reading	marguerite.bordry@gmail.com
		2. Caroline Parfait -> <i>Cybernetics and Society - The Human Use of Human Beings</i> . N. Wiener, 1950.	caroline.parfait@ensapc.fr
27 avril	14h-15h30	Xavier Fresquet	xavier.fresquet@sorbonne-universite.fr
11 mai	14h-15h30	1. David Reymond: SoVisu, un système d'identification d'experts (expert finding system) qui aide à identifier les chercheurs à l'aide de mots-clés.	david.reymond@univ-tln.fr
25 mai	14h-15h30	1. Valentina Fedchenko, Dario Nicolosi Présentation du projet ERC ModERN	valentina.fedchenko@sorbonne-universite.fr