

Réu 16 mars 2026

- Tableau des tâches
- Des réflexions au fil de l'eau
- Exploration biblio Continuous Learning et IR interprétable
- Administratif

Tableau des tâches sur le git

La tâche qui nous concerne

- caractériser la difficulté
 - lexical overlap
 - performance des modèles sur d'autres bench
 - diversité des domaines (overlap inter datasets)
 - présence de documents faux positifs

Problématiques scientifiques

Quantité de données pour l'entraînement : BDD IJ est petite

- Few-shot learning [[Izacard et al. 2023](#)] Atlas: Few-shot Learning with Retrieval Augmented Language Models

Autour de l'apprentissage continu et IR

- [[Belkacem & Teissède 2021](#)] Outil Interactif et Évolutif pour l'Extraction d'Information dans des Documents Techniques **A LIRE**
- [[Hou et al., 2025](#)] Advancing continual lifelong learning in neural information retrieval: Definition, dataset, framework, and empirical evaluation
- [[Saez 2023](#)] Continuous Evaluation Framework for Information Retrieval Systems
 - [[Saez 2022](#)] Continuous Result Delta Evaluation of IR Systems **A LIRE**

⇒ Pas de survey qui couple Information Retrieval et Continuous Learning

[[Hou et al., 2025](#)] Advancing continual lifelong learning in neural information retrieval: Definition, dataset, framework, and empirical evaluation

- Formalisation des tâches de CL
- Framework d'évaluation/ comparaison
 - 5 modèles (3 embedding-based + 2 pre-training based)
 - 6 méthodes de CL (4 Regularization-based, 2 Replay-based)
- Restructuration de MS-MARCO par topic
 - Évaluation de la gestion du *topic-shift*

Autour de la représentation de documents

- [[Clinchant, 2023](#)] Aggregating Continuous Word Embeddings for Information Retrieval **A LIRE**

Autour de l'IR interprétable

- [[Parry et al., 2025](#)] MechIR: A Mechanistic Interpretability Framework for Information Retrieval **A LIRE**

⇒ Recherches à compléter

Environnement pour manipuler les modèles

- Google Colab
- torch, transformers (HuggingFace)
- Téléchargement sur drive des git (zip)

Administratif et organisationnel

- GDR TAL-I
- Réu Place du Numérique 11h30