

Deep learning for Information retrieval / Campagne d'évaluation TREC

Encadrement équipe MLIA/LIP6 : Laure Soulier (laure.soulier@lip6.fr)

1 Contexte

La campagne d'évaluation TREC¹ propose tous les ans de nombreuses tâches en recherche d'information ou traitement automatique du langage naturel.

Dans le cadre du PLDAC, l'idée est de participer à la campagne d'évaluation TREC - Deep Learning Track - avec pour objectifs :

- Réalisation d'un état de l'art
- Implémentation/Proposition d'un modèle et comparaison avec des modèles de référence
- Suivant l'implication des étudiants/résultats, participation à la campagne avec un "nouveau" modèle (envoi des résultats pour obtenir le leaderboard)
- Suivant l'implication des étudiants/résultats, réalisation d'un article pour expliquer le modèle proposé pour la conférence.

Pour plus d'information, voir les références suivantes :

- <https://microsoft.github.io/msmarco/TREC-Deep-Learning>
- [RDV⁺19, MNP⁺20, HZM⁺20, KBG⁺18]

Références

- [HZM⁺20] Sebastian Hofstätter, Hamed Zamani, Bhaskar Mitra, Nick Craswell, and Allan Hanbury. Local self-attention over long text for efficient document retrieval. In Jimmy Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu, editors, *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, pages 2021–2024. ACM, 2020.
- [KBG⁺18] Tom Kenter, Alexey Borisov, Christophe Van Gysel, Mostafa Dehghani, Maarten de Rijke, and Bhaskar Mitra. Neural networks for information retrieval. In Yi Chang, Chengxiang Zhai, Yan Liu, and Yoelle Maarek, editors, *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018*, pages 779–780. ACM, 2018.
- [MNP⁺20] Sean MacAvaney, Franco Maria Nardini, Raffaele Perego, Nicola Tonellotto, Nazli Goharian, and Ophir Frieder. Efficient document re-ranking for transformers by precomputing term representations. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20*, page 49–58, New York, NY, USA, 2020. Association for Computing Machinery.

1. <https://trec.nist.gov/pubs/call2021.html>

- [RDV⁺19] Kirk Roberts, Dina Demner-Fushman, Ellen M. Voorhees, William R. Hersh, Steven Bedrick, Alexander J. Lazar, Shubham Pant, and Funda Meric-Bernstam. Overview of the TREC 2019 precision medicine track. In Ellen M. Voorhees and Angela Ellis, editors, *Proceedings of the Twenty-Eighth Text REtrieval Conference, TREC 2019, Gaithersburg, Maryland, USA, November 13-15, 2019*, volume 1250 of *NIST Special Publication*. National Institute of Standards and Technology (NIST), 2019.