

Lila BOUALILI

PH.D. IN NATURAL LANGUAGE PROCESSING AND INFORMATION RETRIEVAL

Ile-de-France, France

✉ boulili18lila@gmail.com | 🌐 BOUALILILila.github.io | 📄 BOUALILILila | 📺 lilaboulili

Education

IRIT Laboratory, University of Paul Sabatier

Toulouse, France

PH.D. IN COMPUTER SCIENCE

2019 - 2022

- Thesis Topic: Studying Relevant Signals for Document Retrieval using Transformer Models
- Highlights: Enhancing, fine-tuning, and evaluating encoder models for ad hoc retrieval based on cross-encoder and dual-encoder architectures, using single-vector and multi-vector representations, using direct supervision or distillation from teacher models.
- Advisors: Jose G. Moreno, Mohand Boughanem
- Area of study: Deep Learning, Information Retrieval, and Natural Language Processing

Higher National School of Computer Science (ESI)

Algiers, Algeria

MASTER'S DEGREE IN COMPUTER SCIENCE AND ENGINEERING

2018 - 2019

- Thesis Topic: Real Time Tweet Summarization
- Advisors: Lynda Said Lhadj, Mohand Boughanem
- Area of study: Deep Learning, Information Retrieval, and Natural Language Processing

Higher National School of Computer Science (ESI)

Algiers, Algeria

ENGINEERING DEGREE IN COMPUTER SCIENCE AND ENGINEERING (VALEDICTORIAN)

2014 - 2019

- Majored in Information Systems & Software

Experience

Mindflow

Paris, France

NLP RESEARCHER

oct. 2023 - present

- Designing no-code solutions powered by Generative AI agents to streamline task orchestration in cybersecurity operations, reducing complexity and enhancing efficiency.
- Designed and deployed stateful multi-agent systems leveraging Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG), enabling seamless natural language interactions with over 3,000 API integrations. [Patent Pending]
- Evaluated different retrieval approaches for RAG experimenting with sparse, dense, and hybrid models to improve retrieval accuracy and reduce latency.
- Studied various multi-agent workflows, including collaboration through shared memory and hierarchical agent interactions to enhance scalability and reduce cost.
- Designed an evaluation framework for assessing the planning and tool-use capabilities of LLMs, ensuring high-quality performance and decision-making in complex scenarios.

LIG, University of Grenoble Alps

Grenoble, France

POSTDOC RESEARCH FELLOWSHIP

dec. 2022 - sept. 2023

- Enhancing the systematic compositional generalization capabilities of sequence-to-sequence (seq2seq) models by integrating structural (syntactic) information into the decoding process through hyperbolic representations of dependency trees.
- Designed, trained, and evaluated a hybrid Euclidean-Hyperbolic transformer-based seq2seq model from scratch.

Research Internships

Max Planck Institute for Informatics (MPI)

Saarbrücken, Germany

A STUDY OF TERM-TOPIC EMBEDDINGS

Jun. 2021 - Feb. 2022

- Advisor: Andrew Yates
- Studied advancements with the ColBERT architecture, which relies on token-level representations with late interactions for document ranking.
- Proposed a novel approach for distilling ColBERT's contextualized token embeddings into a more explicit and principled process by aggregating a finite set of frozen, pre-trained term-topic embeddings, with each term-topic capturing a contextual topic of a token.

Institut de Recherche en Informatique de Toulouse (IRIT)

Toulouse, France

RESEARCH ON MICROBLOG RETRIEVAL AND SUMMARIZATION

Dec. 2018 - Jul. 2019

- Advisor: Mohand Boughanem
- Development of a tweet summarization approach based on Deep Learning Models. Manipulation of various state-of-the-art Models for both tweet representation and relevance estimation of tweets with respect to users' interests.

Publications

- Lila Boualili
Deep learning for information retrieval: studying relevant signals for ad hoc search based on transformer models
Thesis, Université Paul Sabatier-Toulouse III
- Lila Boualili and Andrew Yates
A Study of Term-Topic Embeddings for Ranking
Advances in Information Retrieval: 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2–6, 2023, Proceedings, Part II
- Lila Boualili, Jose G. Moreno, and Mohand Boughanem
Highlighting Exact Matching via Marking Strategies for Ad hoc Document Ranking with Pretrained Contextualized Language Models
Information Retrieval Journal 25 (4), 414-460
- Lila Boualili, Jose G. Moreno, and Mohand Boughanem
MarkedBERT: Integrating Traditional IR Cues in Pre-trained Language Models for Passage Retrieval
Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20), X'ian, China, July 25-30, 2020.
- Lila Boualili, Lynda Said Lhadj, and Mohand Boughanem
Retrospective Tweet Summarization: Investigating Neural Approaches for Tweet Retrieval
Proceedings of the Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020), Samatan, Gers, France, July 6-9, 2020.

Projects

SciWATCH

- Co-author of SciWatch, a Python package designed to facilitate scientific literature scanning by sending newsletters, configured to specific time intervals, with relevant papers to user-defined queries, featuring comprehensive generated summaries of each paper. The relevance is estimated with a boolean retrieval model combining diverse terms with a set of logical operators.

HYBRIDSEQ2SEQ

- Implemented a Euclidean-hyperbolic hybrid seq2seq model based on the Transformer architecture to address compositional generalization in COGS semantic parsing tasks. The model integrates hyperbolic embeddings to encode complex hierarchical structures, leveraging hyperbolic space's continuous tree-like properties.

Teaching Activities

University of Paul Sabatier

Toulouse, France

IN FRENCH.

2019 - 2021

- Data Structures and Fundamental Algorithms (Semester I)
- Database Programming and administration (Semester II)

Scholarships

2019 **CIMI Excellency Doctoral scholarship**, three years PhD full scholarship

Toulouse, France

2019 **IRIT Internship Support Scholarship**, Six months Research Internship

Toulouse, France

Skills

Programming Python, TypeScript, rust

Libraries Pytorch, TensorFlow, Sickit-learn, Transformers, Sentence-Transformers, Langchain, Fairseq

Operating Systems Microsoft Windows, Linux and other UNIX variants

Agile Methodologies Scrum and Kanban

Languages French (native), English (professional), Arabic (professional), Kabyle (native)

Hobbies

- Avid reader with a focus on Lovecraftian worlds, including manga like Berserk.
- I enjoy drawing and I am exploring photography as a new creative outlet.