Urchade Zaratiana

Member of technical staff at Fastino

✉ urchade.zaratiana@gmail.com | ⓞ urchade.github.io

Professional Experience

Fastino   *Reunion Island*
MEMBER OF TECHNICAL STAFF   *Jul 2024 - present*
- Conducting research on Small Language Models (SLMs).

FI GROUP   *Paris, France*
RESEARCH SCIENTIST   *Dec 2020 - Jul 2024*
- I was a research scientist at FI Group as part of my PhD study "CIFRE" contract.
- Selected projects: Knowledge Graph Construction, Cross-lingual search and QA, keyphrase and entity extraction on scientific texts, scientific taxonomy construction. (python, pytorch, transformers, Large Language Models (LLMs))

ASTEK   *Paris, France*
RESEARCH INTERN   *Feb 2020 - Jul 2020*
- My research focused on the adapation of pretrained language models such as BERT for web article recommendation. Advisor: Alexandre Pascault

Education

Université Paris 13   *Paris, France*
PHD IN COMPUTER SCIENCE   *Jul 2021 - present*
- In-going PhD thesis at Laboratoire Informatique de Paris Nord (LIPN) under the RCLN team. My thesis subject is entity and relation extraction on scientific texts.
- Topics: Natural language processing, Structured Prediction, Combinatorial optimization, Probabilistic graphical models, Graph generation
- Advisors: Dr. Nadi Tomeh and Prof. Thierry Charnois

Johannes Kepler University   *Linz, Austria*
MASTER PROGRAM IN COMPUTER SCIENCE   *2019 - 2020*
- Courses: Deep Learning and Neural Nets, NLP with Deep Learning, Probabilistic models, Multimedia Information Retrieval, Deep Learning in life sciences. (python, pytorch)
- Selected project: Abstractive text summarization for French. Advisor: Dr. Navid Rekab-saz

Institut Supérieur d'Electronique de Paris   *Paris, France*
SOFTWARE ENGINEERING DEGREE   *2017 - 2020*
- Courses: Advanced algorithmics and programming, Artificial Intelligence, Web programming. (python, keras, java, javascript)
- Selected project: Home automation website. Advisor: Dr. Rafael Angarita

Selected publications

Urchade Zaratiana, Nadi Tomeh, Pierre Holat, Thierry Charnois. 2023. **GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer**. Annual Conference of the North American Chapter of the Association for Computational Linguistics, (NAACL 2024).

Urchade Zaratiana, Nadi Tomeh, Pierre Holat, Thierry Charnois. 2024. **An Autoregressive Text-to-Graph Framework for Joint Entity and Relation Extraction**. Proceedings of the 38th AAAI Conference on Artificial Intelligence, (AAAI 2024).

Urchade Zaratiana, Nadi Tomeh, Niama El khabir, Pierre Holat, Thierry Charnois. 2023. **Filtered Semi-Markov CRF**. In Findings of the Association for Computational Linguistics: (EMNLP 2023).

Student supervision

2023	**Mehdi Saissi Hassani**, Research intern on "Benchmarking open-source large language models (LLMs) for retrieval augmented question answering in French".	*Paris, France*
2022	**Dennis Núñez Hernandez**, Research intern on "Data-efficient architecture for extractive question answering".	*Paris, France*

Awards

2023	**1st position at Wojood Named Entity Recognition Shared task at ArabNLP@EMNLP 2023**, Our system is build on models (`github.com/urchade/Filtered-Semi-Markov-CRF`) we developed in prior research on span-based Named Entity Recognition.	*Singapore*
2024	**2nd position at Multilingual AI-generated Text Detection Shared task at SemEval 2024**, Our system is entitled "A Syntactically Motivated Architecture for Multilingual Machine-Generated Text Detection".	*Maxico*