

Lionel Tadjou



As a young researcher in Natural Language Processing, I am currently in my 4th year of PhD with a defense date set for September 2023. My thesis work has focused on named entity recognition, corpus building and pseudonymization, and conversation disentanglement. Before starting my PhD, I worked as a full-stack developer for about 5 years on projects, some of which were European (PARTHENOS H2020). Throughout my professional career, I have supervised an intern and an apprentice.

Currently looking for new challenges where I could apply my knowledge and skills acquired during my PhD on data analysis problems in inter-disciplinary domains such as those of NLP which today includes LLM with libraries such as LangChain which I have already had the opportunity to use. My full-stack developer skills will also help me for rapid application prototyping for example.

PROFIL

+ 33 6 56 84 98 97

lioneltadjou@gmail.com

Paris, France

Cameroon

linkedin.com/in/tatali

EDUCATION

PhD | 2020 – Sept 2023

Field: Natural Language Processing
Sorbonne University Paris - France

Master | 2008 - 2014

Field: Computer Science
University of Dschang, Cameroon

LANGUAGES

French : Native Language

English : B2

Certification

Linux Professional Institute
LPIC Level1: 04/2015 – 04/2020

PROFESSIONAL EXPERIENCES

PhD Student

March 2020 - September 2023

Orange/ Inria | Caen/Paris, France

➤ Topic : Disentanglement of email conversations

- Perform a state of the art of the problem
- Develop an email extraction tool on employees' Outlook mailboxes in compliance with the GDPR
- Implementation of a pseudo-anonymization chain of the collected data based on of Named Entities Recognition (NER)
- Analysis of conversational corpora and production of statistical graphs
- Trained models for dialogue acts recognition and text segments pairing
- Write and present articles

R&D Engineer

May 2017 - February 2020

French National Institute for Computer Science and Applied Mathematic (Inria Paris), France

➤ Project 1 : Standardization Survival Kit (SSK - [Github](#)) within PARTHENOS H2020

SSK is Platform dedicated to support the modeling and management of research data for researchers in Digital Humanities. It has been used as core for the actual [SSH Open Marketplace](#). In this project I was the only developer and below are the tasks that I performed:

- Design and implementation and deployment of platform (Back-end, Front-End and database)
- Integration of a UX designer's sketches into interactive web pages
- Set up of Elasticsearch as a database to facilitate textual and faceted search in the web application
- Build API to interconnect components of the platform and XML files stored on github

➤ Project 2 : [LECTAUREP](#), [LECTure Automatique de REPertoires](#)

This project aims to revolutionize the use that is currently being made of directories of notary deeds by researchers and users of notarial archival funds by developing a platform equivalent to a real tool for transcription, reading and computer analysis of writings organized in the notary's directory.

- Define specifications for a transcription crowdsourcing platform
- Training of segmentation and transcription models for the recognition of notary deeds with the [Kraken tool](#)

HOBBIES

Reading books on personal growth, running, visit natural parks, Chess, basketball, baby foot, football

Network administrator & Helpdesk Support

Jan 2017 - April 2017

Ovamba Cameroon Solutions LTD, Douala, Cameroon

- Management of the computer park for about thirty users
- Configuration and maintenance of Nano wifi repeaters, pfSense firewall and XenServer hypervisor.

Full-Stack Developer

January 2016 - December 2017

Ovamba Cameroon Solutions LTD, Douala, Cameroon

- Track and fix bugs
- Analysis and integration of check management, signatory management, and accounting modules

Full-Stack Developer/Part time

August 2013 – September 2015

Moneytel S.A, Douala Cameroun

In a team of 4 developers including a tech lead, we were working on the design and implementation of a payment system on mobile and web.

Skills

➤ Applications Development

- Programming Languages: Java, Python, JavaScript/TypeScript, C#, SQL
- Frameworks: Spring Boot, Django, Bootstrap, Angular
- Data formats & Database: JSON, XML; MySQL, PostgreSQL, NoSQL, Elasticsearch, MongoDB
- CI/CD : Jenkins, Travis CI, GitLab CI/CD
- Others: Latex, Docker, Junit, Mockito

➤ Natural Language Processing & Machine learning

- Programming Language: Python, Shell scripts
- Libraries: Transformers, Pytorch, TensorFlow, scikit-learn, Spacy, Gensim, NLTK, Pandas, Seaborn, LangChain, OpenAI,

Publications

➤ [Constitution de sous-fils de conversations d'emails](#)

L. Tadonfouet tadjou [et al.] (Submitted and accepted to coria-taln-2023, published by the end of May)

➤ [Building A Corporate Corpus For Threads Constitution](#)

Lionel Tadonfouet Tadjou | Fabrice Bourge | Tiphaine Marie | Laurent Romary | Éric de la Clergerie; RANLP 2021

➤ [A Tree Pattern Matching Algorithm for XML Queries with Structural Preferences. Available](#)

Tchendji, M.T., Tadonfouet, L. and Tchendji, T.T. (2019) A Tree Pattern Matching Algorithm for XML Queries with Structural Preferences.

➤ [SSK by example. Make your Arts and Humanities research go standard](#)

Marie Puren, Laurent Romary, Lionel Tadjou, Charles Riondet, Dorian Seillier

REFERENCES

- Laurent Romary, Director of scientific culture and information
Inria, 3 rue Simone Iff, 75012 Paris
laurent.romary@inria.fr (+49 176 30620964)
- Eric de la Clergerie, Inria NLP Researcher (ALMAAnaCH Team)
Inria, 3 rue Simone Iff, 75012 Paris
Eric.De_La_Clergerie@inria.fr (+33 783 62 68 42)
- Fabrice Bourge, Research Engineer at Orange
42 rue des coutures, 14000, Caen
fabrice.bourge@orange.com (+33 679 93 81 99)