Dr. Maxime Masson

@ maxime-masson@outlook.com | @ maxime.masson@univ-pau.fr | maxime-masson.github.io

LinkedIn | Google Scholar | DBLP | ORCID | HAL

Nationality: French • Location: France

I am a doctor in computer science from the University of Pau and Pays de l'Adour (UPPA) and in language analysis and processing from the University of the Basque Country (UPV/EHU). With a strong foundation in both disciplines, I am now seeking postdoctoral fellowship and research and development opportunities, aiming to contribute to innovative projects and advance my expertise in the field.

Education

Ph.D. in Computer Science — October 2021 – September 2024
University of Pau and Pays de l'Adour — *Pau, France*
Supervisor: Dr. Christian Sallaberry
Research Unit: LIUPPA Laboratory , T2I Research Team

Ph.D. in Language Analysis and Processing — October 2021 – September 2024
University of the Basque Country — *San Sebastián, Spain*
Supervisor: Dr. Rodrigo Agerri
Research Unit: HiTZ Center , Ixa Research Group

Both Ph.D. degrees were obtained through an **International Joint Ph.D. course (Cotutelle).**
Title: Generic Framework for the Multidimensional Processing and Analysis of Social Media: A Proxemic Approach
Keywords: Business Intelligence, Natural Language Processing, Deep Learning, Information Extraction, Visualization
Jury: Prof. Josiane Mothe , Prof. Elena Cabrio , Dir. Ana-Maria Raimond , Dir. Maguelonne Teisseire
Mention: International Ph.D.

M.Sc. in Computer Science — September 2019 - August 2021
University of Pau and Pays de l'Adour — *Pau, France*
Specialization: Internet Technologies
Achievements: Graduated with Honors, Top 5 of the Class

B.Sc. in Computer Science — September 2018 - August 2019
University of Pau and Pays de l'Adour — *Bayonne, France*
Specialization: Software Engineering
Achievements: Graduated with Honors, Valedictorian

University Degree in Technology (DUT) in Computer Science — September 2016 - August 2018
University Institute of Technology (IUT) of Bayonne — *Bayonne, France*

High School Diploma (Baccalaureate) — July 2016
Largenté High School — *Bayonne, France*
Type: Scientific Baccalaureate
Achievements: Graduated with Honors

Work Experience

LIUPPA Laboratory and HiTZ Center - Ixa Group – APs Project — October 2021 - September 2024, 3 years
Academic Researcher (Ph.D. Student) — *Pau, France and San Sebastián, Spain*

- **Objectives**: the APs (Augmented Proxemic services) project aims to propose a multilingual framework for processing and analyzing social media data within semantically defined application domains (e.g., through ontologies or thesauri). The main objective of this project is to assist decision-makers and local stakeholders across various application domains in getting insights and indicators based on social media to address domain-specific requirements.
- **Selective Funding**: E2S UPPA and Urban Community of Pau Béarn Pyrénées .
- **Keywords**: Business Intelligence, Natural Language Processing, Deep Learning, Information Extraction, Data Visualization.
- Proposition of a formal methodology to build multidimensional datasets from social media. Building accurate and exhaustive datasets is a recurrent challenge in the Web and Social Media Search research field. However, most approaches currently used are ad hoc and, therefore, difficult to reuse. This methodology addresses this issue by proposing an iterative and incremental pipeline applied to several data feeds (e.g., posts, metadata, media, etc.), incorporating both human feedback and automatic mechanisms to improve quality.

- Set up of a comparative study of rules-based, fine-tuning, and few-shot learning techniques alongside various new large language models (LLMs) for extracting knowledge from unstructured, multilingual, and noisy social media texts in the tourism domain. Social media posts are challenging in Natural Language Processing (NLP) due to their multilingualism, brevity, the presence of informal language, and frequent grammatical errors, among other factors. Additionally, I investigated a recurrent challenge faced by researchers: determining the minimum number of training annotations required to achieve competitive results in a specific domain. Manual annotations are both time-consuming and costly; thus, researchers aim to annotate as few samples as possible while still maintaining high-quality results.
- Design of modular, domain-adaptive indicators by reinterpreting the theory of proxemics for social media. The challenge is that most indicators for social media are domain-specific, meaning they are effective within a specific domain of application but difficult to adapt to others. My indicators, expressed as similarity measures, stand out due to their modularity based on proxemic dimensions, and their applicability across heterogeneous entities, such as users, demographics, themes, time periods or places.
- **Developed Software Platform: TextBI, An Interactive Dashboard for Visualizing Multidimensional Analyses in Social Media Data (Demonstration Video)**
  TextBI is an interactive dashboard designed to visualize multidimensional indicators on social media across various dimensions (e.g., spatial, temporal, thematic, personal, sentimental). It addresses the challenge of presenting complex information in a way that is adaptable to various domains and easily interpretable by non-computer scientists, such as local stakeholders (e.g., tourism offices, municipal councils). Unlike existing Business Intelligence (BI) tools, TextBI offers interactive visuals specifically designed for social media, featuring sentiment and engagement overlays, multilevel timelines, thematic maps, proxemic crosshairs and interaction graphs.
- **Main Technologies and Languages:** Python for the back-end (Data Collection and Pre-processing), Deep Learning with LLMs (Mistral, LLaMA 2, FlanT5) and MLMs (BERT, XLM-RoBERTa) using Huggingface Transformers (Text Mining, Information Extraction), HTML / CSS / JavaScript with Visualization Libraries for the front-end (Plotly, Cytoscape, Leaflet, etc.).

LIUPPA Laboratory and L3i Laboratory – DA3T Project February 2021 - September 2021, 8 months
Research Engineer (Internship) *Pau, France and La Rochelle, France*

- **Objectives**: the DA3T project (Digital Trace Analysis Device for the Valorization of Touristic Territories) aims to develop a system for analyzing multidimensional mobility tracks both outdoor in cities and indoor, for example, in museums, to assist local planners and decision-makers in managing and promoting touristic areas. It focuses on creating tools and methods for extracting, processing and analyzing mobility tracks. A mobile application named Geoluciole was developed to collect tracks from tourists in various touristic cities of the region.
- **Selective Funding**: French Nouvelle-Aquitaine Region and Charente Tourism .
- **Keywords**: Geomatics, Geospatial Analysis, ETL, GPS Tracks, Mobility, Semantic Trajectories, Data Integration, Similarity Measures.
- Work on a multi-level, multi-aspect model for analyzing semantic trajectories, addressing several challenges in the geomatics field. Specifically, the model focuses on: modeling semantic trajectories with data enrichment associated with positions or segments; defining enrichment generically to integrate various dimensions; and structuring this enrichment according to a hierarchical organization.
- Design and implementation of 3D visualization modules, including a customizable space-time cube, and semantic trajectory enrichment modules by leveraging open data sources, including OpenStreetMap, Google Maps Places, and the DataTourisme ontology, as well as through semi-structured interviews.
- Participation in the design of composite and interpretable semantic trajectory similarity measures that assist geographers in assessing the similarity of touristic trajectories at various granularities (e.g., macro, meso and micro).
- **Developed Software Platform: MobilityETL, A Novel ETL (Extract, Transform, Load) Tool Dedicated to Processing Mobility Tracks (Demonstration Video)**
  Design and development of a novel ETL platform (Extract, Transform, Load) dedicated to processing mobility tracks. It represents the first mobility-specific ETL system and addresses the challenge of seamlessly analyzing heterogenous mobility tracks coming from various sources. More precisely, it allows geographers to automatically integrate (e.g., process, enrich, visualize) many mobility tracks through modular and reactive pipelines accessible to users who are not necessarily computer scientists.
- **Main Technologies and Languages:** C#, .NET Framework, and WPF (Windows Presentation Foundation) to build the GUI of the desktop front-end application. Python, along with various spatial processing libraries (e.g., Geopandas), for the processing modules in the back-end. The 3D space-time cube visualization was developed in JavaScript with the Three.JS WebGL library.

MaxSea TimeZero – MapMedia Project March 2018 - August 2019, 1 year and 4 months
Software Engineer (Work-Study Program) *Bidart, France*

- **Background**: MaxSea TimeZero is one of the leading companies in the field of maritime navigation applications and the production of cartographic data.

- Work within an Agile team on the company's vector and raster cartographic production cloud processes (hosted on Microsoft Azure), integration of bathymetric data, and implementation of a GUI application for managing production flows (correction of polygons, monitoring and orchestration of heavy cloud processes, etc.).
- Redesign of the graphical user interfaces (GUIs) of the company's internal tools to improve user experience and migrate them from WinForms and ASP.NET to Windows Presentation Foundation (WPF). This initiative aimed to enhance maintainability, scalability, and usability for the company's geographers, who are non-computer scientists. Responsibilities included analyzing existing systems, gathering requirements, designing, and developing solutions using C# and .NET. Additionally, set up automated unit and integration tests to ensure quality and reliability.
- **Main Technologies and Languages:** ASP.NET, Azure Cloud Suite (Azure Batch, Azure Blob Storage, Azure SQL), SQL Server, WinForms, Windows Presentation Foundation (WPF), C#, .NET Framework and Core, OSGeo's OGR and GDAL components for .NET.

Core Skills

Languages and Technologies:

- Python (Pandas, spaCy, NLTK, Folium, Matplotlib, BeautifulSoup libraries)
- C#, .NET Framework (WPF, WinForms, XAML, ASP .NET, Azure Batch, Azure Blob Storage)
- Java, JavaFX, Android Programming with Java and Kotlin
- Web Technologies (PHP, HTML, CSS, JavaScript), Data Visualization libraries (WebGL with Three.JS, Plotly, Cytoscape, Leaflet, etc.)
- Databases (MySQL, SQLite, SQL Server, MongoDB)
- Machine Learning (Huggingface Transformers with Large Language Models: Mistral, LLaMA 2, FlanT5; Masked Language Models: BERT, XLM-RoBERTa)
- Linux Bash, Windows PowerShell, LaTeX, UML / OCL modelling

Methodologies and Tools:

- Agile, Scrum, Kanban, Object-oriented Programming (OOP), Test-driven development (TDD)
- Visual Studio 2022, Visual Studio Code, PyCharm, IntelliJ IDEA, Android Studio, Overleaf, QGIS, PowerBI

Soft Skills:

- Time Management, Teamwork, Problem-solving, Documentation, Scientific Paper Writing, Engaging Presentation and Reports

Languages

English: Full Professional Proficiency – CEFRL C2 Level

French: Native Proficiency

Japanese: Currently Learning (Beginner Level)

Awards

Geodata Challenge Award
Winner (ranked 1st) of the Geodata Challenge of the **French National Geonumeric Days** (**GeoDataDays 2023**) held at the Reims Convention Center on September 12-13 2023 with the proposal "*Visualization of data from social media: a Business Intelligence type platform*". This event was organized by the French Association for Geographic Information (Afigéo), DécryptaGéo and sponsored by ESRI (ArcGIS).

Publications

Ph.D. Thesis (defended September 23rd 2024)

M. Masson (2024). Generic Framework for the Multidimensional Processing and Analysis of Social Media Content: A Proxemic Approach.

International Conferences and Journals (Peer-reviewed)

M. Masson, P. Roose., C. Sallaberry, M. N. Bessagnet, A. Le Parc Lacayrelle, R. Agerri (2023). ProxMetrics: Modular Proxemic Similarity Toolkit to Generate Domain-Adaptable Indicators from Social Media. *Social Network Analysis And Mining*, 14, 124. Springer International Publishing. (Impact Factor: 2.8).

M. Masson, C. Sallaberry, M. N. Bessagnet, A. Le Parc Lacayrelle, P. Roose, R. Agerri. (2024). TextBI: An Interactive Dashboard for Visualizing Multidimensional NLP Annotations in Social Media Data. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations (EACL 2024)* (pp. 1-9) (St. Julians, Malta). Association for Computational Linguistics. (CORE Rank: A, ERA Rank: A).

M. Masson, P. Roose, C. Sallaberry, R. Agerri, M. N. Bessagnet, A. Le Parc Lacayrelle. (2023). APs: A Proxemic Framework for Social Media Interactions Modeling and Analysis. In *International Symposium on Intelligent Data Analysis (IDA 2023)* (pp. 287-299) (Louvain-La-Neuve, Belgium). Cham: Springer Nature Switzerland (CORE Rank: B, ERA Rank: A).

M. Masson, C. Sallaberry, R. Agerri, M. N. Bessagnet, P. Roose, A. Le Parc Lacayrelle. (2022). A Domain-independent Method for Thematic Dataset Building from Social Media: The Case of Tourism on Twitter. In *International Conference on Web Information Systems Engineering (WISE 2022)* (pp. 11-20) (Biarritz, France). Cham: Springer International Publishing (CORE Rank: B, ERA Rank: A).

M. Masson, C. Cayèré, M. N. Bessagnet, C. Sallaberry, P. Roose, C. Faucher. (2022). An ETL-like Platform for the Processing of Mobility Data. In *Proceedings of the 37th ACM Symposium on Applied Computing (ACM SAC 2022)* (pp. 547-555) (Brno, Czech Republic) (CORE Rank: B, ERA Rank: B).

C. Cayèré, C. Sallaberry, C. Faucher, M. N. Bessagnet, P. Roose, M. Masson, J. Richard. (2021). Multi-level and Multiple Aspect Semantic Trajectory Model: Application to the Tourism Domain. *ISPRS - International Journal of Geo-Information*, 10(9), 592. (Impact Factor: 3.4)

National Conferences and Journals (Peer-reviewed)

M. Masson, C. Sallaberry, M. N. Bessagnet, A. Le Parc Lacayrelle, P. Roose, R. Agerri. (2024). Visualization of data from social media: a Business Intelligence type platform. In *Mappemonde. Quarterly Journal on the Geographic Image and the Forms of the Territory*. OpenEdition Journals. (to be published).

C. Cayèré, C. Sallaberry, C. Faucher, M. N. Bessagnet, P. Roose, M. Masson. (2024). Multidimensional Similarity Measures for Semantic Trajectories. In *Open Journal in Information Systems Engineering*. 4, 2. ISTE Open Science.

M. Masson, R. Agerri, C. Sallaberry, M. N. Bessagnet, P. Roose, A. Le Parc Lacayrelle. (2024). Optimal Strategies for the Multidimensional Analysis of Multilingual Content from Social Media. In *Proceedings of the 42nd Conference on Computer Science for Organizations and Information and Decision Systems* (INFORSID 2024) (Nancy, France).

M. Masson, S. Abdelhedi, C. Sallaberry, R. Agerri, M. N. Bessagnet, A. Le Parc-Lacayrelle, P. Roose. (2023). Interactive Visualization of Tourist Activity Trajectories: Application to Data Extracted from Twitter. In *Workshop "Exploring traces in an all-digital world: challenges and perspectives"* at INFORSID 2023 (La Rochelle, France).

M. Masson. (2022). Augmented Proxemic Services for Cultural Heritage and Tourism Practices. In *Young Researchers' Forum* at INFORSID 2022 (Dijon, France).

C. Cayèré, C. Sallaberry, C. Faucher, M. N. Bessagnet, P. Roose, M. Masson. (2022). Similarity Measurement for Semantic Trajectories: Taking into Account Three Levels of Granularity. In *Proceedings of the 40th Conference on Computer Science for Organizations and Information and Decision Systems* (INFORSID 2022) (Dijon, France). Recipient of an Outstanding Paper Award.

M. Masson, C. Cayèré, M. N. Bessagnet, C. Sallaberry, P. Roose, C. Faucher. (2022). Spatio-temporal Visualization of Outdoor Tourist Mobility Data . In *Workshop on Spatial and Temporal Data Management and Analysis (GAST), Francophone Conference on Knowledge Extraction and Management* (EGC 2022) (Blois, France)

Talks

Workshops

M. Masson. (November, 2023). TextBI: A Generic Dashboard for Interactive Visualization of Multidimensional Data from Social Media. *Workshop on Spatialized Digital Humanities, Annual Meeting of the GdR CNRS MAGIS (CNRS Research Network on Methods and Applications for Geomatics and Spatial Information)*. Maison des Suds (Bordeaux, France).

M. Masson, S. Laborie. (June, 2023). A Generic Framework for the Extraction, Processing, Analysis and, Valorization of Social Media Content. *Symposium "Constitution of corpus for the needs of digital marketing in the domain of fashion" (European Cassini Program)*, Parthenope University of Naples (Naples, Italy).

M. Masson. (November, 2022). APs: A Proxemic Approach for Data Analysis on Social Media. *Workshop Smart city, smart destination: from management to territorial experience*, IRGO - Research Institute in Organizational Management, University of Bordeaux (Bordeaux, France).

M. Masson. (September, 2022). APs: A Proxemic Approach for Data Analysis on Social Media. *Inter-association Day EGC/INFORSID*, IRIT, University of Toulouse III (Toulouse, France).

Webinars and Seminars

M. Masson. (January, 2024). TextBI: An Interactive Platform for Visualizing Multidimensional Data from Social Media. Invited Speaker: *Webinar on Cartography and Geovisualization of the GdR CNRS MAGIS (CNRS Research Network on Methods and Applications for Geomatics and Spatial Information)* (Online).

M. Masson, P. Roose. (July, 2023). Analyzing Touristic Data in the Basque Country. *Urban community of the Basque Country* (Bayonne, France).

M. Masson. (June, 2023). A Generic Framework for the Extraction, Processing, Analysis, and Valuation of Social Media content: Application to the Domain of Tourism and the Social Media Twitter. *Ixa Seminar*, University of the Basque Country (EHU/UPV) (San Sebastián, Spain).

Teaching Activities

Operating Systems Management. October 2023. 1st year of Bachelor of Technology in Computer Science (BUT), University Institute of Technology (IUT) of Bayonne and the Basque Country (9h tutorials).
Subjects: Windows, Linux, PowerShell and Bash Scripting.

Server-side Programming. September to December 2023. 2nd year of Bachelor of Technology in Computer Science (BUT), University Institute of Technology (IUT) of Bayonne and the Basque Country (27h tutorials).
Subjects: PHP and Asynchronous JavaScript (Ajax).

Development of Web Applications. September to December 2023 and 2024. 2nd year of Bachelor in Computer Science (L2), Faculty of Science (STEE), University of Pau and Pays de l'Adour (100h lectures and tutorials). Course coordinator (administrative responsibility).
Subjects: HTML, CSS, JavaScript, and the SAM (State-Action-Model) pattern.

Office Automation. September to December 2023. Bachelor in Insurance (LP), European Research Center for Family, Insurance, Personal, and Health Law - CERFAPS, University of Bordeaux (23h lectures). Course designer and coordinator (administrative responsibility).
Subjects: Advanced Excel (Data Integration, Charting, Formula, Macro, etc.).

Office Automation. September to December 2022. Bachelor in Insurance (LP), European Research Center for Family, Insurance, Personal, and Health Law - CERFAPS, University of Bordeaux (23h lectures). Course designer and coordinator (administrative responsibility).
Subjects: Advanced Excel (Data Integration, Charting, Formula, Macro, etc.).

Computer Science I. September to December 2022. 1st year of Bachelor in Economy and Management (L1), Faculty of Social Sciences and Humanities (SSH), University of Pau and Pays de l'Adour (32h tutorials).
Subjects: Computer Science for Economy and Management (Excel, PowerBI, Tableau Public).

Computer Science II. January to June 2022. 2nd year of Bachelor in Economy and Management (L2), Faculty of Social Sciences and Humanities (SSH), University of Pau and Pays de l'Adour (32h tutorials).
Subjects: Computer Science for Economy and Management (Excel, PowerBI, Tableau Public).

Supervision and Miscellaneous Activities

Student Volunteer (2024) at the 22nd International Conference on Pervasive Computing and Communications (PerCom 2024) (CORE Rank: A*, ERA Rank: A). https://percom.org/2024

Co-supervision of Master's Students (January to June 2024): Alexy Del Amo Alonso and Maxime Silla from the *University of Pau and Pays de l'Adour, France*.
Master's Project Title: "Extraction of Web Documents About Marine Biology".

Co-supervision of an Intern (February to August 2023): Siwar Abdelhedi from the *Higher Institute of Computer Science, Tunisia, ISI*.
Internship's Project Title: "Generic, Multi-dimensional, and Multi-level Visualization of Data from Social Media: Application to the Tourism Domain".

Co-supervision of Master's Students (January to June 2023): Victor Laffarguette and Thomas Procureur from the *University of Pau and Pays de l'Adour, France*.
Master's Project Title: "Annotation of Tourism-Related Phrases in Tweets".

Co-supervision of Master's Students (January to June 2022): Aitor Cachenaut and Benjamin Laby from the *University of Pau and Pays de l'Adour, France*.
Master's Project Title: "Extraction of Spatial and Thematic Named Entities from Multilingual Tweets".

Interests

Traveling: I enjoy traveling and discovering new places and cultures. So far, I have visited: Morocco, Tunisia, Spain, Italy, Greece, Malta, Belgium, China and Japan.

Japan: I am highly interested in Japan and are currently learning Japanese following several trips to the country.

References

References available upon request.