



GABRIEL DE JESUS

Researcher (Ph.D.) | Information Retrieval | Natural Language Processing

PROFILE

I hold a Ph.D. in Informatics Engineering and I am an Affiliate Researcher at INESC TEC, Porto, Portugal, based in Timor-Leste. My research focuses on Information Retrieval, Natural Language Processing, and AI for the Tetun language, including RAG and Agentic AI. I develop language resources and computational methods to advance Tetun language technology, contributing to digital inclusion for under-resourced languages.

PROJECTS:

- **Labadain Chat:** An Agentic AI conversational assistant designed specifically for Tetun speakers (www.labadain.com).
- **Labadain Chat Old:** An LLM-based conversational assistant for Tetun speakers (old.labadain.com).
- **Labadain Search:** A monolingual search engine developed specifically for Tetun speakers (search.labadain.com).
- **Timor News:** An online news portal exclusively publishes news in Tetun (www.timornews.tl).
- **Timor News (old site):** The legacy news portal that was used during the initial years of the news agency's operations (<https://old.timornews.tl>).
- **Labadain Crawler:** A data collection pipeline for web crawling (<https://github.com/gabriel-de-jesus/labadain-crawler>).

EDUCATION

Faculty of Engineering, University of Porto (FEUP), Portugal

September 2021 – September 2025

Ph.D. in Informatics Engineering

Thesis: <https://hdl.handle.net/10216/169208>

Faculty of Science, University of Porto (FCUP), Portugal

September 2011 – July 2013

Master's in Computer Science

National University of Timor-Leste (FUP / UNTL), Timor-Leste

September 2005 – June 2008

Bachelor's in informatics engineering

WORK EXPERIENCE (SELECTED)

Timor News | Founder

May 2019–Present (5+ years)

I am responsible for developing strategic plans and implementing standard business procedures to achieve long-term objectives, including overseeing and guiding the team to ensure that the company's goals are met.

Chemonics International (HRH2030 Program, USAID-funded project)

| Human Resource Information System Manager

July 2019–February 2021 (1 year 8 months)

My roles included conducting a feasibility study, developing functional requirements for the Training Management Information System (TMIS), and customizing the iHRIS Train toolkit to meet those needs. I also prepared comprehensive user and system administrator manuals for the TMIS, including conducting formal training sessions.

Autoridade Nacional de Comunicações (ANC) | Non-Executive Board Member

December 2018–August 2021 (2 years 9 months)

My responsibilities were defined under Article 8 of the Telecommunications Decree-Law no. 15/2012, dated March 28, 2012. In this capacity, I contributed to governance and strategic oversight of the telecommunications sector, participated in regulatory decision-making, and supported the development and enforcement of policies that promoted compliance and fair competition.

Food and Agriculture Organization (FAO) | Database Design and Implementation Expert Consultant

November 2018–March 2019 (5 months)

My roles included ensuring quality and compliance of the Suco Level Food Security Monitoring System (SLMS) development agreement, while advising the Ministry of Agriculture and Fisheries on the SLMS workflow. Additionally, I provided training and support to end-users to ensure the successful implementation of the system.

TEACHING EXPERIENCE

Feb 2015-Jun 2016: Invited lecturer at the Instituto Profissional de Canossa (IPDC). *Taught courses in data warehouse, data mining, and object-oriented programming.*

Aug 2013-Dec 2013 and **Feb 2009-Jun 2009:** Invited lecturer at the Universidade Nacional Timor Lorosa'e (UNTL). *Taught ICT and database courses.*

SUPERVISION EXPERIENCE

September 2024-July 2025:

- Co-supervising a master's student in Informatics at FEUP, Portugal, working on *integrating semantic search in a sports search Engine.*

- Supervising one undergraduate student in Information Systems at the Escola Superior de Tecnologia e Gestão (ESTG), Portugal, working on *cross-lingual information retrieval in Tetun.*

TRAINING

Aug 2025, Retrieval Augmented Generation (RAG), Coursera: <https://www.coursera.org/verify/NJJ7LIZHYHWQ>

Feb 2023, Machine Learning Specialization, Coursera: <https://coursera.org/verify/specialization/PX7PT8Q4JYJ5>

June 12-16, 2017, Situational Leadership Training, Bangkok, Thailand.

CONTACT

Mobile/WA: +670 7733 2224

Email: gabrieldejesusphd@gmail.com

Personal page: <https://gabrieldejesus.labadain.com/>

LinkedIn: <https://www.linkedin.com/in/gabrieljesus>

Major Project Secretariat (MPS)/Infrastructure Fund (IF) | Coordinator of the Administration and Information Technology (IT) Directorate
January 2014–September 2017 (3 years 9 months)

I served as the head of the IT and Administration Directorate, overseeing the Department of Administration & HR and the Department of Procurement, Logistics, IT, and Databases. In this role, I planned, managed, and supervised the IT infrastructure, information systems, administration, HR, procurement, budget planning, and logistics to ensure efficient and effective operations.

Timor Telecom | Senior Database Administrator
September 2009–August 2011 (2 years)

I managed and administered the mediation and billing databases, which included the integrated customer management system, NGIN and the post-paid mobile and IS provider. In this role, I collaborated with the marketing team to address fraud issues and develop new applications for the company.

Quidgest Lda | Software Developer
July 2008–August 2009 (1 year 2 months)

My responsibilities included software and database development, implementation, and maintenance. I also provided support and training to end-users to ensure efficient use of the software.

RESEARCH AND PUBLICATIONS (SELECTED)

de Jesus, G., & Nunes, S. (2025). Insights into LLM-based conversational search: A study of Tetun-speaking users' search behavior. In *Proceedings of the 2025 International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval (ICTIR)* (pp. 297–306). Association for Computing Machinery. <https://doi.org/10.1145/3731120.3744596>

de Jesus, G., Singh, S. A. K., Nunes, S., & Yates, A. (2025). Zero-shot and hybrid strategies for Tetun ad-hoc text retrieval. In *Proceedings of the 2025 International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval (ICTIR)* (pp. 264–274). Association for Computing Machinery. <https://doi.org/10.1145/3731120.3744593>

de Jesus, G., & Nunes, S. (2024). Exploring large language models for relevance judgments in Tetun. In C. Siro, M. Aliannejadi, H. A. Rahmani, N. Craswell, C. L. A. Clarke, G. Faggioli, B. Mitra, P. Thomas, & E. Yilmaz (Eds.), *Proceedings of the First Workshop on Large Language Models for Evaluation in Information Retrieval (LLM4Eval 2024)* (Vol. 3752, pp. 19–30). CEUR-WS. <https://ceur-ws.org/Vol-3752/>

de Jesus, G., & Nunes, S. (2024). Data collection pipeline for low-resource languages: A case study on constructing a Tetun text corpus. In N. Calzolari et al. (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC/COLING 2024)* (pp. 4368–4380). ELRA and ICCL. <https://aclanthology.org/2024.lrec-main.390>

de Jesus, G., & Nunes, S. (2024). Labadain-30k+: A monolingual Tetun document-level audited dataset. In M. Melero et al. (Eds.), *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024* (pp. 177–188). ELRA and ICCL. <https://aclanthology.org/2024.sigul-1.22>

The full list of my research publications is available on Google Scholar:
<https://scholar.google.com/citations?user=puPeqEUAAA&hl>