

# Leopoldo Pla Sempere

SOFTWARE ARCHITECT · PART-TIME LECTURER

✉ b64encoded: bGVvcG9sZG9wbGFzZW1wZXJlQGdtYWlsLmNvbQ== | 🌐 lpla.github.io | 📷 lpla

## Experience

### Part-time lecturer

UNIVERSITY OF ALICANTE

San Vicente del Raspeig

Oct. 2019 - present

- Teaching in **Mathematics and Computer Science** degrees
- Programming algorithmic fundamentals and advanced techniques for software specification
- Use real-world examples and challenges from Google, Amazon, Microsoft, and Spotify.

### Innovation projects engineer

GRUPO ENERCOOP

Crevillent

Nov. 2023 - present

- Lead engineer in **Energy Production and Transition** area, managing EU funded innovation projects.
  - **InEEsS**: Innovative Energy (Efficiency) Service Models for Sector Integration via Blockchain.
  - **DIGIRED-IA**: digital twin and intelligent management tools for electrical distribution networks based on artificial intelligence.
  - **MODERATE**: Marketable Open Data Solutions for Optimized Building-Related Energy Services.
  - **STREAM**: Creating an innovative and robust flexibility ecosystem on the low voltage (LV) grid side of existing power markets.

### Senior Data Scientist, Music Information Retrieval and Machine Learning engineer

UNIVERSITY OF ALICANTE

San Vicente del Raspeig

Jun. 2023 - Nov. 2023

- Lead developer in **MultiScore** project, training neural models to learn both OMR and AMT holistically (end-to-end) for transcribing music

### Senior Data Scientist, Natural Language Processing and Machine Learning engineer

UNIVERSITY OF ALICANTE

San Vicente del Raspeig

Nov. 2017 - Jun. 2023

- Lead developer and coordinator of **software Work Packages in EU funded projects (CEF)**, in coordination with
  - University of Edinburgh, Prompsit, TAUS, Omniscien and Johns Hopkins University:
    - \* **ParaCrawl** (Provision of Web-Scale Parallel Corpora for Official European Languages)
    - \* **Paracrawl 2** (Broader Web-Scale Provision of Parallel Corpora for European Languages)
    - \* **Paracrawl 3** (Continued Provision of Web-Scale Parallel Corpora for European Languages) (paracrawl.eu)
  - University of Groningen, Jožef Stefan Institute, Prompsit:
    - \* **MaCoCu**: Massive collection and curation of monolingual and bilingual data: focus on under-resourced languages (macocu.eu)
- My work focused on **creating parallel corpora** for EU official languages **by a broad web crawling effort**.
- Coordinated the development of the **Open Source Machine Learning based suite Bitextor** (bash, python, C++, snakemake, awk...), the production-ready tool that merges the technology pipeline for these CEF projects, run under **HPCC and Cloud scale**.
- 6 people team under my supervision, in **Agile** environments.
- Cloud resources provided by **Amazon (AWS S3), Microsoft (Azure) and Docker**.
- OS and system administration of various Linux-based systems (Scientific Linux, Ubuntu, Fedora, Debian)

### Data Scientist, Natural Language Processing and Machine Learning engineer

PROMPSIT LANGUAGE ENGINEERING

Elche

May 2015 - Oct. 2017

- Developing natural language related projects as Reverso Context, processing full range of corpora and language resources
  - developed for Reverso (Softissimo, approx. rank 250 in Global Similarweb Ranking)
  - **several million page-views per day**, big data Cloud scale computing, from an Agile development team
  - using technologies as SVN, pandas, jupyter, bash (UNIX tools), word2vec, keras, scipy and Apache Solr

### Full-stack developer intern

PROMPSIT LANGUAGE ENGINEERING

Elche

Feb. 2013 - May 2013

- Developing an internal news sentiment analysis suite
- Python and bash scripting for news crawling and plain text extraction from all downloaded HTML data
- AJAX, jQuery, CSSv3 and PHP for real-time interactive interface for dataset generation, MySQL for dataset storage

## Education

### M.Sc. Artificial Intelligence, Pattern Recognition and Digital Imaging, GPA: 8.81/10

POLYTECHNIC UNIVERSITY OF VALENCIA

Valencia

Sep. 2014 - Jun. 2015

- M.Sc. specialised in **natural language processing and deep learning**
- Coursed machine learning MOOCs of Andrew Ng (Stanford University) and Yaser Abu-Mostafa (Caltech)
- Master Thesis: Audio classical composer identification in MIREX 2015: submission based on Structural Analysis of Music

### B.Sc. Computer Science, GPA: 8.53/10

UNIVERSITY OF ALICANTE

San Vicente del Raspeig

Sep. 2010 - Jun. 2014

- B.Sc. specialised in Informatics (robotics, machine learning, artificial vision, compilers)
- **Enrolled in the group of high profile academics programme**
- Diploma Thesis: Dodecaphonic music composer assistant with OpenMusic

## Publications & Awards

### Building Domain-specific Corpora from the Web: the Case of European Digital Service Infrastructures

LREC 2022

van Noord et al.

Marseille, France

June, 2022

### MaCoCu: Massive collection and curation of monolingual and bilingual data: focus on under-resourced languages

EAMT 2022

Bañón et al.

Ghent, Belgium

June, 2022

### ParaCrawl: Web-Scale Acquisition of Parallel Corpora

ACL 2020

Bañón et al.

Online

6 July, 2020

### Computer Science Degree graduation and M.Sc scholarship

BEST ACADEMIC RECORD

University of Alicante, Generalitat Valenciana

San Vicente del Raspeig

2015

### B.Sc thesis defense

HIGHEST MARK (OUTSTANDING WITH HONORS)

University of Alicante

San Vicente del Raspeig

2014

## Presentation & Associations

### EAMT

WEBMASTER FOR THE EUROPEAN ASSOCIATION OF MACHINE TRANSLATION

- Technical role managing services as eamt.org, lists.eamt.org, mt-archive.net and several EAMT congress websites

Remote

July 2021 - Present

### DockerCon 2022

PRESENTER FOR <THE BITEXTOR PIPELINE>

- Introduced the Bitextor pipeline and explained its uses with Docker

Remote

May, 2022

### Docker Community All-Hands #4 2021

PRESENTER FOR <THE BITEXTOR PIPELINE>

- Introduced the Bitextor pipeline and explained its uses with Docker, including a live demo

Remote

Dec. 2021

## Skills

### Natural Language Processing tools

Moses, Marian, Bleualign, Bitextor, Bicleaner

### Scripting and tools

git, CMake, tmux, Python, Lisp, Bash, AWK, Perl, systemd

### Machine Learning Toolkits

TensorFlow, Keras, PyTorch, HuggingFace, pandas, sklearn, numpy, scipy, matplotlib

### Programming

C++, C, Java, Golang, JAVA, Rust, LaTeX, CMake, Snakemake

### Languages

English (proficient), Spanish (native), Catalan (native)

## Extracurricular Activity

### Clarinet Professional degree

PROFESSIONAL CONSERVATORY OF MUSIC OF ELCHE

- Specialized in jazz improvisation with Miguel García Ferrer
- MOOCs on jazz improvisation with Gary Burton (Berklee School of Music)
- MOOC on fundamentals of rehearsing music ensembles with Dr. Evan Feldman (University of North Carolina at Chapel Hill)
- After studies, performed in several youth orchestras, symphonic wind orchestras and rock, pop, swing and jazz bands, writing arrangements, improvising, both playing clarinet and piano as soloist. Non-professional experience with saxophone, guitar, ukulele, bass, and drums.

Elche

Sep. 2004 - Jun. 2010

### Hardware-related projects

SELF-TAUGHT

- Open source PCB layout design using EAGLE and KiCad, with real production in factories as OSHPark and JLCPCB, for 80s computers (GB, C64).
- Electronic repairs and coding on vintage video-game platforms. "Best sound" award at "Retroconsolas Alicante 2013" for the Amstrad CPC 464.
- Open source 3D design (Autodesk), modelling and printing. Volunteer with "3D Makers Elche" group during COVID-19 lockdown by 3D-printing more than two hundred facemasks for near hospitals.