

# Leopoldo Pla Sempere

SOFTWARE ARCHITECT · PART-TIME LECTURER

✉ b64encoded: bGVvcG9sZG9wbGFzZW1wZXJlQGdtYWlsLmNvbQ== | 🏠 lpla.github.io | 📧 lpla | 📺 leopoldoplasempere

## Experience

### Part-time lecturer

UNIVERSITY OF ALACANT

San Vicent 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.

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

UNIVERSITY OF ALACANT

San Vicent del Raspeig

Nov. 2017 - present

- Lead developer in EU funded projects in coordination with
  - University of Edinburgh, Prompsit, TAUS, Omniscien and John 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)
  - University of Groningen, Jožef Stefan Institute, Prompsit:
    - ★ MaCoCu: Massive collection and curation of monolingual and bilingual data: focus on under-resourced languages
- The work focus on creating corpora for all EU official languages (full list at Paracrawl site and MaCoCu site) by a broad web crawling effort.
- Designed state-of-the-art methods for the entire processing chain, from identifying web sites with translated text all the way to cleaned parallel corpora that are ready as training data for CEF Automated Translation.
- Coordinated the development of the Open Source suite Bitextor [Github] (bash, python, C++, snakemake, awk...), the production-ready tool that merges all this necessary technology chain for these CEF projects, run under HPCC-Cloud scale (Cambridge Service for Data-Driven Discovery, CSD3).
- 6 people team under my supervision.
- Cloud resources provided by Amazon (AWS S3), Microsoft (Azure) and Docker.
- OS 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 300 in Global Alexa Ranking)
  - several million page-views per day, big data Cloud scale computing
  - grown in an agile development team
  - using technologies as SVN, pandas, jupyter, bash (UNIX tools), word2vec, keras, scipy and Apache Solr
  - Language resources I was developing were in all available languages in the app (Arabic, German, English, French, Hebrew, Italian, Japanese, Dutch, Polish, Portuguese, Romanian and Russian)
  - High responsibilities in product releases and Quality Assessment.

### Full-stack developer intern

PROMPSIT LANGUAGE ENGINEERING

Elche

Feb. 2013 - May 2013

- Developing news sentiment analysis
- 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

### Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

València

Sep. 2014 - Jun. 2015

- GPA: 8.81/10
- 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

## Computer Science Degree

UNIVERSITY OF ALACANT

San Vicent del Raspeig

Sep. 2010 - Jun. 2014

- GPA: 8.53/10
- B.Sc. specialized 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

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

ACL 2020

Online

6 July, 2020

- Marta Bañón, Pinzhen Chen, Barry Haddow, Kenneth Heafield, Hieu Hoang, Miquel Esplà-Gomis, Mikel L. Forcada, Amir Kamran, Faheem Kirefu, Philipp Koehn, Sergio Ortiz Rojas, Leopoldo Pla Sempere, Gema Ramírez-Sánchez, Elsa Sarrías, Marek Strelec, Brian Thompson, William Waites, Dion Wiggins, and Jaume Zaragoza.

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

BEST ACADEMIC RECORD

University of Alicante, Generalitat Valenciana

San Vicent del Raspeig

2015

### B.Sc thesis defense

HIGHEST MARK (OUTSTANDING WITH HONORS)

University of Alicante

San Vicent del Raspeig

2014

## 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, pandas, sklearn, numpy, scipy, matplotlib

### Programming

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

### Languages

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

## Extracurricular Activity

### Clarinet Professional degree

ELCHE SCHOOL OF MUSIC

Elche

Sep. 2004 - Jun. 2010

- 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 from the University of North Carolina at Chapel Hill
- After studies, gained experience in several youth orchestras, symphonic wind orchestras and rock, swing and jazz bands, writing arrangements, improvising, while playing clarinet and piano as soloist. Non-professional experience with saxophone, guitar, ukulele, bass and drums.

### Hardware-related projects

SELF-TAUGHT

- Electronic repairs, modding and programming on old video-game platforms. Winner of the "best sound" award at "Retroconsolas Alicante 2013" for my Amstrad CPC 464 entry.
- Open source 3D design (Autodesk), modeling and printing. Volunteer with "3D Makers Elche" during COVID-19 lockdown by printing more than two hundred face-masks for near hospitals.
- Open source PCB layout design using EAGLE and KiCad, with real production in factories as OSHPark and JLCPCB, for computers and consoles from the 80's (Commodore and Game Boy).