Leopoldo **Pla Sempere**

Software Architect · Part-time lecturer

b64encoded: bGVvcG9sZG9wbGFzZW1wZXJlQGdtYWlsLmNvbQ== | lpla.github.io | lpla | leopoldoplasempere

Summary

Senior Software Engineer at Transducens group in University of Alicante. 6+ years experience specializing in the NLP-focused tools development, HPCC-Cloud scale, and machine learning. Super nerd who loves Vim, Linux and OS X and enjoys to customize all of the development environment. Interested in devising a better problem-solving method for challenging tasks, and learning new technologies and tools if the need arises.

Experience

Part-time lecturer

San Vicent del Raspeig

University of Alacant

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

San Vicent del Raspeig

University of Alacant

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

Elche

Prompsit Language Engineering

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

Elche

Prompsit Language Engineering

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

València

Universitat Politècnica de 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

San Vicent del Raspeig

University of Alacant

Sep. 2010 - Jun. 2014

- GPA: 8.53/10
- B.Sc. specialized in Informatics (robotics, machine learning, artificial vision, compilers)
- Enroled in the group of high profile academics programme
- Diploma Thesis: Dodecaphonic music composer assistant with OpenMusic