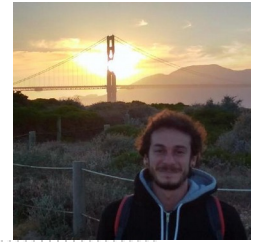


Antonio Ercole De Luca

GITHUB <https://github.com/eracle>

SKYPE [ercole.deluca](#) •

E-MAIL eracle@posteo.eu



Big Data Scientist and Senior Software Developer

Main professional experience with Python from 2016,
previous working experience with Java from 2013.

Focused on all aspects of data management such as
Retrieval, Storage, Scaling, Analysis and Mining.
Experience on many projects involving:
Web Scraping, Relational Databases, Cloud infrastructure,
REST API and Jupiter notebook, then Regression and Continuous Model
Deployment.

Always fascinated by the potential given intersecting data coming from the Web
and Data Mining techniques such as Data Clustering, Graph Analysis, Information
Retrieval and Machine Learning.

Github Contributions:

Dask: Random Sampling on Dask's Bags.

Github personal project:

linkedin: Scraping software aimed to collect info from linkedin users. It uses
scrapy, docker, Selenium Webdriver and Python3.

Some Skills:

Python3, R, Java, Groovy, C, Bash;
Git, PyCharm, Gradle;
docker, docker-compose, Amazon AWS, heroku, Ubuntu Server;
Django, Django REST framework, Scrapy, Selenium Webdriver;
Pandas, dask, Jupyter;
postgresql, elasticsearch, celery, redis;
Scrum, Agile, TDD;
CrowdFlower, BPMN, Petri Nets, OpenStreetMap;

Antonio Ercole De Luca

GITHUB <https://github.com/eracle> SKYPE [eracle.deluca](#) E-MAIL eracle@posteo.eu

Working Experience:

- 2020** **Push**
Data Scientist
- Data cleaning - GeoSpatial Data;
- python3, pandas, dask, Jupyter-lab, OpenStreetMap.
- 2019** **ErnieApp**
Dublin, Ireland
Remote
Backend Test Developer
- Code Refactoring, Increasing Testing Coverage metrics;
- python3, docker, django, celery, postgres, on Heroku.
- 2018** **XPLOD Social**
Los Angeles, California
Remote
Software Architect
- Design an Analytics Platform to show insights on data gathered from Instagram;
- python3, docker, django, celery, redis, Selenium, Instapy, Laravel, Mysql, on DigitalOcean.
- 2018** **Trademark Pro**
Canada
Remote
Data Scientist
- Image similarity search engine (MVP);
- python3, docker, flask, keras, tensorflow, scipy, numpy, annoy, postgres, on EC2 and S3.
- 2018** **Kiwicampus.com**
Berkeley, California
On Site
Backend Software Developer
- In charge to build the Integration platform with [Doordash Inc.](#)
- django, postgres, celery, redis, postgres, on Heroku.

- 2018** **Decyde.io**
San Francisco, California
Remote
Python Software Developer
- python3, scrapy, scrapinghub.com, google-custom-search, selenium.
- 2017** **Einhorn Manufaktur**
Berlin, Germany
Remote
Backend Software Developer
- REST API Developer.
- Python3, Django, Postgres, TDD, Github.
- 2016** **Legalkite Ltd**
Zurich, Switzerland
Remote
Backend Software Developer
- REST API, Data storage and Cloud infrastructure.
- Python3, Django, Amazon AWS, ElasticSearch, Agile, TDD, Jira.
- 2015** **SpazioDati s.r.l.**
Pisa, Italy
[Internship/Thesis](#)
- Supervisors: Prof. [Paolo Ferragina](#) and [Ugo Scaiella](#)
- Development of an ETL workflow which involved CrowdFlower (rebranded now as [Figure Eight](#)).
- Particular focus was given to cost reduction by leveraging an [Active Learning](#) technique.
- The output dataset was used to train a logistic regression which replaced one of the steps of SpazioDati's proprietary Entity Linking system (SaaS): namely [Dandelion](#)
- Tools used: Groovy, IntelliJ IDEA, WEKA, Gradle, Git and Ubuntu server.

2013 **Synthema S.R.L.**
Pisa, Italy
Software Developer

Design and implementation of Java modules for Crawling and Linguistic Analysis (Romanian, Japanese, Chinese and German),
part of the [CAPER](#) european project.

Education:

2006 **Bachelor's Degree**
Laurea triennale
Computer Science
Informatica
[University of Pisa](#), Italy
Outstanding final marks 110/110
Thesis: Document classification using lexical and semantic resources.
(Catalogazione di libri tramite risorse di semantica lessicale)

2009 **Erasmus LLP Program**
Facultad de Informática
[Universidad Politécnica de Valencia](#), Spain

2011 **Postgraduate Master's Degree**
Laurea Magistrale
Data Science and Business Informatics
[University of Pisa](#), Italy
Outstanding final marks 110/110
Thesis: Applying Active Learning on Entity Linking.
([Algoritmi di Active Learning applicati al problema dell'Entity Linking Pruning](#)).

2019 **Bachelor's Degree**
Laurea triennale
Statistics for Data Analysis
Statistica per l'analisi dei dati
Università di Palermo, Italy
Classes: R Laboratory, Probability Calculus, Exploratory Statistics,
Statistical Inference, Multivariate Statistical Analysis.

Publications:

Semantic Crawling: an Approach based on Named Entity Recognition
Proceedings of FOSINT-SI 2014
Antonio Ercole De Luca, Carlo Aliprandi, Giulia Di Pietro, Matteo Raffaelli, Tiziana Soru.

CAPER: Crawling and Analysing Facebook for Intelligence Purposes

Proceedings of ASONAM2014

Carlo Aliprandi, Antonio E. De Luca, Davide Gazzè, Mariantonietta N. La Polla, Giulia Di Pietro, Matteo Raffaelli, Andrea Marchetti, Maurizio Tesconi.

Languages:

English: Fluent

Spanish: Fluent

Italian: Native

Activities:

Co-Founder [Python Meetup Palermo](#).