RODRIGO MARTÍNEZ CASTAÑO

E-mail	rodrigo@martinez.gal
LinkedIn	https://www.linkedin.com/in/brunneis/
GitHub	https://github.com/brunneis, https://github.com/labteral
Google Scholar	https://scholar.google.com/citations?user=LDGO-2EAAAAJ

Professional Experience

June 2018 – now

Lead Backend / Blockchain Engineer
Councilbox Technology

- *Redesigned the massive mailing system in Node.js with RabbitMQ to properly scale up and recover from failures, guaranteeing that every notification is sent only once.*
- *Designed and developed Graphn: a horizontally-scalable microservice-based Python blockchain for evidence generation that pegs to trustworthy blockchains such as Bitcoin or Ethereum. [explorer]*
- *Developed Stopover: a simple and robust message broker built on top of RocksDB for communications between critical microservices. [server] [sdk] [microservices]*
- *Developed DigSig: a digital signature library for Python with RSA and ECDSA support. [code]*
- *Developed Binnakle: a JSON-based logger for Python and Javascript to facilitate log analysis on tools such as Datadog.*
- *Developed Liebre: a RabbitMQ Python library to simplify communications between microservices.*
- *Developed Wudder and Withose: the first and second generation of APIs with authentication to create blockchain evidence at scale for external clients.*
- *Developed PGKV: a Python library to use PostgreSQL as a horizontally scalable key-value store with the Citus extension. [code]*
- *Developed Synced: a Python library to create disk-synced variables. [code]*
- *Developed EasyWeb3: a Python library to deploy and interact with Ethereum smart contracts. [code]*
- *Developed Txlog: a Python library to define crash-resistant Python code through transactions. [code]*
- *Developed Quies: a modular Python monitoring system to notify interruptions of service through arbitrary scripts.*
- *Developed Quartz: a scalable timer for HTTP requests.*
- *Developed a backend for Quorum / Ethereum blockchains to build block explorers.*

September 2019 – January 2021

Computer Security Teaching Assistant
University of Santiago de Compostela (USC)

- *Served multiple classes on the subject and reviewed students' activities.*
- *Created lessons and code to explain Proof-of-Work consensus, the blockchain data structure, and smart contracts. [code]*

September 2016 – January 2021

Big Data and AI R&D Engineer
Singular Research Centre on Intelligent Technologies (CiTIUS, USC)

- *Developed Catenae: a Python microservices framework on top of Apache Kafka. [code] [paper]*

- *Developed Ancoris: a simple docker container orchestrator for clusters in Python. [paper]*
- *Developed Alien: an anonymous, horizontally-scalable Reddit scraper for real-time data collection. [code]*
- *Designed and developed a distributed, horizontally-scalable real-time system for early detection of signs of depression on Reddit with Catenae microservices. [video] [code] [paper]*
- Developed an HAProxy-based proxy balancer to feed web scrapers.
- Designed and developed a distributed system to perform real-time extraction of depressed users from Reddit with minimal context with several strategies.
- Developed Ernie, a Python library for building BERT-based Sentence Classification AI models. [code]
- Developed a system with Ernie to detect early signs of committing self-harm on Reddit in real time and won a scientific competition. [paper]

March 2016 – June 2016

System Administrator Intern
CESGA (Galicia Supercomputing Center)

- *Developed dockerfiles for several Big Data processing tools packaged in the software distributions from Hortonworks (HDP) and Cloudera (CDH).*
- *Automated the deployment of Docker-based clusters of those distributions in the supercomputing infrastructure with Python Fabric.*

November 2015 – June 2016

Big Data R&D Intern
Singular Research Centre on Intelligent Technologies (CiTIUS, USC)
Department Collaborator Scholarship (Ministry of Education)

- *Designed and developed a distributed, horizontally-scalable real-time system for sentiment analysis on Twitter with Big Data technologies. [video] [code] [paper]*
  - *Improved and integrated Polypus: the massive Twitter scraper.*
  - *Used Aerospike (a memory-based key-value store similar to Redis) to store tweet IDs to avoid duplicates, and as a fast temporary buffer for real-time processing.*
  - *Coded a scalable real-time processing architecture with Java for Apache Storm.*
  - *Designed a key-value store on Apache HBase to store all the downloaded tweets with efficient key-based scanning.*
  - *Developed a module with Scala for Apache Spark to automatically pre-process the sentiment based on a set of configurable keywords.*
  - *Developed an HTTP API to retrieve statistics and analyze the sentiment associated with arbitrary keywords in real-time within a time frame.*
  - *Developed a web UI with AngularJS to interact with the API.*
  - *Packaged all the modules in Docker containers and deployed the architecture on AWS.*

June 2014 – December 2014

Big Data R&D Intern
Singular Research Centre on Intelligent Technologies (CiTIUS, USC)
Talemtum Long Track Scholarship (Telefónica)

- *Developed Polypus: a massive anonymous Twitter scraper that could retrieve up to 30% of the published tweets in real time. Coded in Java with multi-thread support. While only 1% of the tweets were made available in real time, exploiting the anonymous Search API used by the front-end with a massive keyword dictionary through a battery of proxies.*
- *Developed a Hadoop MapReduce application in Java for batch processing the sentiment analysis on a massive amount of tweets stored on Apache HBase.*
- *Developed a web app to interact with the system.*
- *Configured, deployed, and benchmarked the architecture on AWS.*

Education

2012 - 2016 Santiago de Compostela, Spain	**University of Santiago de Compostela** B.Sc. in Computer Science Best Final Project
2016 - 2018 Santiago de Compostela, Spain	**University of Santiago de Compostela y Universidad de Murcia** M.Sc. in Big Data & Data Science Best Final Project, Best Academic Record
2018 - Santiago de Compostela, Spain	**University of Santiago de Compostela** Ph.D. in Big Data & Information Retrieval (unfinished)

Scientific Publications

2022

Real-Time Focused Extraction of Social Media Users
R Martínez-Castaáo, DE Losada, JC Pichel
IEEE Access 10, 42607-42622

Early Risk Detection of Self-Harm Using BERT-Based Transformers
R Martínez-Castaño, A Htait, L Azzopardi, Y Moshfeghi
Early Detection of Mental Health Disorders by Social Media Monitoring, 183-206

2021

Hybrid intelligence strategies for identifying, classifying and analyzing political bots
B García-Orosa, P Gamallo, P Martín-Rodilla, R Martínez-Castaño
Social sciences 10 (10), 357

BERT-based transformers for early detection of mental health illnesses
R Martínez-Castaño, A Htait, L Azzopardi, Y Moshfeghi
International Conference of the Cross-Language Evaluation Forum for European Languages

2020

Early Risk Detection of Self-Harm and Depression Severity using BERT-based Transformers: iLab at CLEF eRisk 2020
R Martínez-Castaño, A Htait, L Azzopardi, Y Moshfeghi
CLEF 2020 Working Notes 2696 (50)

A Big Data Platform for Real Time Analysis of Signs of Depression in Social Media
R Martínez-Castaño, JC Pichel, DE Losada
International Journal of Environmental Research and Public Health 13 (17), 23

Ignis: An efficient and scalable multi-language Big Data framework
C Piñeiro, R Martínez-Castaño, JC Pichel
Future Generation Computer Systems, 105, 705-716.

2019

Blockchain: A Technical Introduction
R Martínez-Castaño
4ª Revolución Industrial: Impacto de la Automatización y la Inteligencia Artificial en la Sociedad y la Economía Digital, 19

2018

Building Python-Based Topologies for Massive Processing of Social Media Data in Real Time
R Martínez-Castaño, JC Pichel, DE Losada
Proceedings of the 5th Spanish Conference on Information Retrieval, 18

A Micromodule Approach for Building Real-Time Systems with Python-Based Models: Application to Early Risk Detection of Depression on Social Media
R Martínez-Castaño, JC Pichel, DE Losada, F Crestani
European Conference on Information Retrieval, 801-805

Polypus: a Big Data Self-Deployable Architecture for Microblogging Text Extraction and Real-Time Sentiment Analysis
R Martínez-Castaño, JC Pichel, P Gamallo
arXiv [cs.DC] - Distributed, Parallel, and Cluster Computing

Awards

2019

Predoctoral Scholarship
Xunta de Galicia (Government)

M.Sc. Best Final Project
Unviersity of Santiago de Compostela

M.Sc. Best Academic Record
Unviersity of Santiago de Compostela

M.Sc. Extraordinary Award
Unviersity of Santiago de Compostela

M.Sc. Best Final Project
DXC Technology

III Hackathon Big Data & Analytics CEIN (1st Prize)
DXC Technology

2018

Alastria's Open Call 2018 of Castilla y León (1st Prize)
Alastria (alastria.io)

2017

B.Sc. Best Final Project
University of Santiago de Compostela

2016

I Hackathon Big Data & Analytics CEIN (1st Prize)
Hewlett Packard Enterprise

2014

Talentum Startups Long Track Scholarship
Telefónica