Pedro Paulo BALAGE
Curriculum vitae

✍ Lisbon, Portugal
☎ +351 910 326 543
✉ pedrobalage@gmail.com
⚡ www.pedrobalage.com

WORK EXPERIENCE

APRIL 2016 – PRESENT DAY

Priberam Labs
Research Scientist

I research and develop new technologies to transform, extract, classify and cluster textual data. I have worked on tasks such as named-entity recognition; entity linking; text summarization; and sentiment analysis in the context of the H2020 SUMMA project.

MAY 2015 – DEC 2015

Boo Labs
Data Scientist

I built tools to extract information and to classify e-commerce products for product matching and catalog management. I worked in an agile environment frequently pushing changes into production. Boo Labs startup was later acquired by the Brazilian e-commerce giant B2W.

SEPTEMBER 2014 – DECEMBER 2015

BuscaOpinioes
Founder

Busca Opinioes was a startup focused on helping consumers to achieve the best e-commerce shopping experience by the use of automatically curated user reviews.

JANUARY 2013 – SEPTEMBER 2013

Dicionário Criativo
NLP Researcher

Dicionário Criativo is a Brazilian startup which was conceived to offer linguistic tools to help users in creative writing. I was responsible to organize the linguistic resources and to develop the first onomasiological thesaurus for Portuguese available online.

MAY 2011 – SEPTEMBER 2011

Xerox Research Centre Europe (XRCE) - France
NLP Intern

In this internship I developed an interface in Java to support the Xerox opinion mining system. My work was awarded as the 2nd best internship project in that summer and I got a paper published in NAACL-Demo.

EDUCATION

2012 – 2017	**PhD in Natural Language Processing** *University of São Paulo - Brazil*
2010 – 2012	**Masters in Natural Language Processing** WITH DISTINCTION *University of Wolverhampton, UK* *University of Algarve, Portugal*
2005 – 2010	**Bachelor in Computer Science** *University of São Paulo - Brazil*

COMPLEMENTARY EDUCATION

2017	7th Lisbon Machine Learning School (LxMLS). Member of the organization - monitor in the lab sessions.
2016	12th International Conference on the Computational Processing of Portuguese - PROPOR. Chair of the 1st Student Workshop

COMMUNICATION SKILLS

PORTUGUESE	Native speaker
ENGLISH	Oral: good – Written: good
SPANISH	Oral: good

SOFTWARE SKILLS

NLP	NLTK, SpaCy, Stanford CoreNLP
ML	SVM, CRF, MLP, ConvNets, LSTMs, Seq2seq, Sklearn, Tensorflow, PyTorch
DEV OPS	AWS, Vagrant, Docker, Git, Apache, Shell Scripting, Vim
DBs/INDEX	PostgreSQL, MongoDB, Apache Cassandra, Elasticsearch

SELECTED PUBLICATIONS

Liepins, R. et al. (2017), The SUMMA Platform Prototype, *in* 'Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics', Association for Computational Linguistics, Valencia, Spain, pp. 116–119.

Mendes, A., Nogueira, D., Broscheit, S., Aleixo, F., Balage, P., Martins, R., Miranda, S. & Almeida, M. S. (2017), 'SUMMA at TAC Knowledge Base Population Task 2017', pp. 1–10.