

Roque Enrique López Condori

Curriculum Vitae

Personal Information

Name Roque Enrique López Condori
Nationality Peruvian
Website <https://roquelopez.com>
Github <https://github.com/roquelopez>
LinkedIn <https://linkedin.com/in/roquelopez>
G. Scholar <https://scholar.google.com/citations?user=eCBv1NMAAAAJ>

Education

2013–2015 **MSc in Computer Science**, *São Paulo University*, Brazil.
2006–2010 **BSc in Computer Science**, *San Agustin National University*, Peru.
2001–2005 **High School Diploma**, *Colegio Técnico Parroquial San Martín de Porres*, Peru.

Awards

- 2023, Distinguished participation in the External Operations Platform Data Challenge (JIATF), USA.
- 2023, Best Paper Honorable Mention Award at IEEE Visualization Conference (VIS), Australia.
- 2023, Best Paper Award at Data Management for End-to-End Machine Learning (DEEM) Workshop, co-located with SIGMOD, USA.
- 2020, Finalist team for the Wells Fargo Campus Analytics Challenge 2020, USA.
- 2016, Finalist for the Best MSc Dissertation in IEEE Latin American Conference on Computational Intelligence, Colombia.
- 2014, Young Scientific Best Paper in VII International Conference on Intelligent Information and Engineering Systems, Poland.
- 2014, Master Scholarship granted by Samsung-Brazil (2014-2015), Brazil.
- 2013, Top 5th Bachelor thesis in Simposio Peruano de Inteligencia Artificial, Peru.
- 2010, Placed in top 5th in the undergraduate, Peru.
- 2005, First place and scholarship High School Diploma, Peru.

Experience

Dec 2018 **Research Associate**, *New York University*, USA.

present I am working on multiple DARPA and ARPA-H research projects at NYU's Visualization, Imaging, and Data Analytics lab (VIDA).

- In the BDF project, I'm working on building a Python toolkit for biomedical data integration using Generative Artificial Intelligence (GenAI) and Natural Language Processing (NLP) techniques.
- In the D3M project, using Reinforcement Learning (RL) methods, I have implemented an automated Machine Learning (ML) system that automatically searches for models and derives end-to-end pipelines that read, pre-process the data, and train the model.
- In the PTG project, I explored and implemented different techniques based on Large Language Models (LLMs) for reasoning tasks, e.g. to identify errors in recipe steps and to extract preconditions and effects of actions in recipes.

Feb 2017 **Research Engineer**, *Inria*, France.

Aug 2018 I worked on extracting symbolic knowledge about objects and their characteristics from unstructured text (relying on natural language processing, machine learning and machine reading techniques), as well as available ontologies and knowledge on the Semantic Web.

Oct 2016 **CTO**, *DoctorCV*.

Sept 2020 Peruvian startup (<https://doctor-cv.com/>) which evaluates and proposes improvements in your CV using NLP techniques. It is like having a virtual recruiter expert who will tell what to change in the resume to increase the chances of being called for the job interview.

July 2015 **Researcher, Software Engineer**, *Elabora Consultoria*, Brazil.

Jan 2017 I researched some information retrieval techniques with machine learning methods and word embedding representations to create hierarchies of terms in collections of patent about different topics. In this project, I used Spark, HBase and Python.

June 2014 **Associate Editor**, *CompuScientia*.

Dec 2014 CompuScientia is a Peruvian electronic journal of the Computer Science Student Society. I had to review and write articles about the ecosystem of Peruvian startups.

Feb 2014 **Teaching Assistant**, *São Paulo University*, Brazil.

July 2014 I served as TA to professor Dr. Thiago Pardo for the course *Professional Information in Computer Science*. I was responsible for evaluating and grading examinations or assignments. In collaboration with another TA, we supported 50 students during discussion sections and office hours.

Oct 2013 **Software Engineer**, *Dicionário Criativo*, Brazil.

Dec 2013 I implemented a crawler for Twitter in Python. This crawler collected public comments about music, books and films. I also developed a module to normalize these comments (spell checker, etc.). With this data and other resources the database was increased.

Jan 2012 **Researcher, Software Engineer**, *Lindexa - Wayra Telefónica*, Peru.

Dec 2012 I researched and developed some automatic techniques of opinion mining and sentiment analysis for Spanish language. In addition, I implemented a crawler for Facebook and Twitter in Python. This crawler collected all public comments written on these social networks. Lindexa was a winner startup of the first edition of Wayra Peru.

- Jan 2011 **Technical Researcher, Developer**, *Cátedra CONCYTEC*, Peru.
- Dec 2011 In general, we researched techniques to generate extractive summaries of many texts. These summaries selected the most important data in terms of coverage and representativeness. I also researched and implemented text classification and clustering algorithms.
- Sept 2010 **Interface Designer**, *Software Solutions S.A.C.*, Peru.
- Dec 2010 We developed the Graphical User Interfaces of the Credits and Collections module, which included the elaboration of the navigability documents and prototypes. I also developed the executable prototype.
- Sept 2009 **Analyst, Developer**, *EPISUNSA Development Centre*, Peru.
- Dec 2009 Together with a group of colleagues, we worked on the design and development of the website of the Professional School of Systems Engineering.

Research Projects

- Mar 2024 ***BDF: Biomedical Data Fabric Toolbox***, *New York University*, USA.
 present The BDF project, supported by the Advanced Research Projects Agency for Health (ARPA-H), aims to streamline the integration of biomedical research data from thousands of sources, addressing challenges posed by incompatible data formats.
- Dec 2021 ***PTG: Perceptually-enabled Task Guidance***, *New York University*, USA.
 June 2024 The PTG project, funded by the Defense Advanced Research Projects Agency (DARPA), seeks to develop methods, techniques, and technology for artificially intelligent assistants that provide just-in-time visual and audio feedback to help with task execution.
- Dec 2018 ***D3M: Data-Driven Discovery of Models***, *New York University*, USA.
 Mar 2024 The D3M project, a Defense Advanced Research Projects Agency (DARPA) funded effort, aims to develop automated model discovery systems that enable users with subject matter expertise but no data science background to create empirical models of real and complex processes.
- Feb 2017 ***ALOOF: Autonomous Learning of the Meaning of Objects***, *Inria*, France.
 Aug 2018 The goal of ALOOF is to significantly advance the ability of today's robots and autonomous systems to adapt to ever changing, dynamic real world environments by enabling them to learn about the meaning of objects from resources accessible through the Web.
- July 2015 ***Patent Faceted Searches Using Topic Modeling***, *Elabora Consultoria*, Brazil.
 Feb 2017 This project aimed to integrate advanced techniques of information retrieval for making faceted searches over big collections of patents about different topics.
- Jan 2014 ***Semantic Text Processing in Brazilian Portuguese***, *Samsung – ICMC/USP*,
 June 2015 Brazil.
 The overall objective of this project was to advance the state of the art in semantic processing of texts written in Brazilian Portuguese. With these resources, we built mining and summarization texts applications, focused on Internet opinions.
- Jan 2011 ***Automatic Summarization of Medical Records using Natural Language***
 Dec 2011 ***Processing techniques***, *Cátedra CONCYTEC*, Peru.
 The goal of this project was to investigate and apply techniques to generate automatic summaries of many documents. We developed different methods to extract, select and organize the most important information.

Publications

- 2024 **Sonia Castelo, Joao Rulff, Parikshit Solunke, Erin McGowan, Guande Wu, Iran Roman, Roque Lopez, Bea Steers, Qi Sun, Juan Bello, Bradley Feest, Michael Middleton, Ryan Mckendrick, Claudio Silva**, *HuBar: A Visual Analytics Tool to Explore Human Behavior based on fNIRS in AR Guidance Systems*, IEEE Visualization Conference (VIS), St. Pete Beach, USA.
(to appear)
- 2023 **Roque Lopez, Raoni Lourenco, Remi Rampin, Sonia Castelo, Aecio Santos, Jorge Ono, Claudio Silva, Juliana Freire**, *AlphaD3M: An Open-Source AutoML Library for Multiple ML Tasks*, AutoML Conference (AutoML-Conf), Potsdam/Berlin, Germany.
- 2023 **Sonia Castelo, Joao Rulff, Erin McGowan, Bea Steers, Guan-de Wu, Shaoyu Chen, Iran Roman, Roque Lopez, Ethan Brewer, Chen Zhao, Jing Qian, Kyunghyun Cho, He He, Qi Sun, Huy Vo, Juan Bello, Michael Krone, Claudio Silva**, *ARGUS: Assistive Visualization of Human-AI Collaboration for Task Guidance in Augmented Reality*, IEEE Visualization Conference (VIS), Melbourne, Australia.
- 2023 **Haoxiang Zhang, Roque Lopez, Aecio Santos, Jorge Ono, Aline Bessa, Juliana Freire**, *Using Pipeline Performance Prediction to Accelerate AutoML Systems*, Workshop on Data Management for End-to-End Machine Learning (DEEM), co-located with SIGMOD, Seattle, USA.
- 2020 **Jorge Ono, Sonia Castelo, Roque López, Enrico Bertini, Juliana Freire and Claudio Silva**, *PipelineProfiler: A Visual Analytics Tool for the Exploration of AutoML Pipelines*, IEEE Visualization Conference (VIS).
- 2018 **Valerio Basile, Roque López and Elena Cabrio**, *Measuring Frame Instance Relatedness*, Joint Conference on Lexical and Computational Semantics (*SEM), co-located with NAACL, New Orleans, USA.
- 2017 **Roque López and Thiago Pardo**, *Opinion Summarization Methods: Comparing and Extending Extractive and Abstractive Approaches*, Expert Systems with Applications (ESWA), volume 78.
- 2016 **Lucia Castro, Roque López, Gabriel Cavalcante and Luiz Lapolla**, *Towards Automatic Building of Term Hierarchies from Large Patent Datasets*, Artificial Intelligence and Natural Language Conference (AINL), Saint-Petersburg, Russia.
- 2015 **Roque López and Thiago Pardo**, *Experiments on Sentence Boundary Detection in User-Generated Web Content*, International Conference on Intelligent Text Processing and Computational Linguistics (CICLING), Cairo, Egypt.
- 2015 **Roque López, Lucas Avanço, Pedro Balage, Alessandro Bokan, Paula Cardoso, Márcio Dias, Fernando Nóbrega, Marco Sobrevilla, Jackson Souza, Andressa Zacarias, Ariani Di Felippo, Eloize Seno and Thiago Pardo**, *A Qualitative Analysis of a Corpus of Opinion Summaries based on Aspects*, Linguistic Annotation Workshop (LAW), co-located with NAACL, Colorado, USA.
- 2014 **Verônica Agostini, Roque López and Thiago Pardo**, *Automatic Alignment of News Texts and their Multi-document Summaries: Comparison among Methods*, International Conference on Computational Processing of Portuguese (PROPOR), São Carlos, Brazil.

- 2014 **Roque López, Javier Tejada and Mikhail Alexandrov**, *Medical Texts Classification based on Keywords using Semantic Information*, International Conference on Intelligent Information and Engineering Systems (INFOS), Rzeszow-Krynica, Poland.
- 2014 **Márcio Dias, Alessandro Bokan, Carla Chuman, Cláudia Barros, Erick Maziero, Fernando Nobrega, Jackson Souza, Marco Sobrevilla, Marina Delege, Lucía Castro, Naira Silva, Paula Cardoso, Pedro Balage, Roque López, Vanessa Marcasso, Ariani Felippo, Maria Graças, Thiago Pardo**, *Enriquecendo o Corpus CSTNews - a Criação de Novos Sumários Multidocumento*, Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish (ToRPorEsp), co-located with PROPOR, São Carlos, Brazil.
- 2012 **Roque López, Javier Tejada and Mike Thelwall**, *Spanish Sentistrength as a Tool for Opinion Mining Peruvian Facebook and Twitter*, International Conference on Intelligent Information and Engineering Systems (INFOS), Rzeszow-Krynica, Poland.
- 2012 **Angels Catena, Mikhail Alexandrov and Roque López**, *Parameterization of comments from Peruvian Facebook and Twitter: Lexical Resources and Algorithm*, International Conference on Intelligent Information and Engineering Systems (INFOS), Rzeszow-Krynica, Poland.
- 2012 **Alessandro Bokan, Roque López**, *Método No Supervisado para la sugerencia de Tags utilizando información semántica basada en conocimiento*, Congreso Internacional de Computación y Telecomunicaciones (COMTEL), Lima, Peru.
- 2011 **Roque López, Dennis Barreda and Javier Tejada**, *MFSRank: An Unsupervised Method to Extract Keyphrases Using Semantic Information*, Mexican International Conference on Artificial Intelligence (MICAI). Lecture Notes in Artificial Intelligence. Springer, Puebla, Mexico.
- 2011 **Roque López, Dennis Barreda, Javier Tejada and Luis Alfaro**, *Método Supervisado orientado a la clasificación automática de documentos. Caso Historias Clínicas*, Encuentro Chileno de Computación (ECC), Curicó, Chile.
- 2011 **Roque López, Mikhail Alexandrov and Javier Tejada**, *LexisTerm - The program for term selection by the criterion of specificity*, International Conference on Intelligent Information and Engineering Systems (INFOS), Polanczyk, Poland.
- 2011 **Ales Bourek, Mikhail Alexandrov and Roque López**, *Folksonomy - supplementing RICHE expert based taxonomy by terms from online documents*, International Conference on Intelligent Information and Engineering Systems (INFOS), Polanczyk, Poland.
- 2011 **Roque López, Dennis Barreda, Javier Tejada and Luis Alfaro**, *Clasificación automática de Historias Clínicas basada en Prototipos utilizando técnicas de Procesamiento de Lenguaje Natural*, Jornadas Peruanas de Computación (JPC), Pucallpa, Peru.
- 2011 **Dennis Barreda, Roque López, Javier Tejada and Luis Alfaro**, *Un algoritmo genético para la agrupación de documentos aplicado en corpus con características diferentes*, Jornadas Peruanas de Computación (JPC), Pucallpa, Peru.

Monographs

- MSc Thesis *Title:* Automatic Aspect-Based Opinion Summarization Methods
Description: In this master's project are presented investigations to generate extractive and abstractive summaries of opinions using an aspect-based approach. Besides using known methods in the area, it also was proposed two new methods for Portuguese language which got the best performance in the experiments.
- BSc Thesis *Title:* Medical Documents Classification based on Keywords using Semantic Information
Description: In this monograph, it is presented a method to classify medical documents which improves the results of Naive Bayes and Rocchio algorithm. This method, in addition to considering statistical information, taking into account the semantic relatedness between the keywords of medical documents.

Congress and Scientific Events

- 2023 **Assistant**, *AutoML Conference*, Potsdam/Berlin, Germany.
- 2022 **Assistant**, *Tandon Research Excellence Exhibit*, New York City - USA.
- 2021 **Assistant**, *D3M Winter Workshop*, Online.
- 2020 **Assistant**, *International Conference on Very Large Data Bases*, Online.
- 2020 **Panelist**, *Simposio Peruano de Deep Learning*, Arequipa - Peru.
- 2019 **Assistant**, *IBM AI Systems Day*, Cambridge - USA.
- 2018 **Reviewer**, *SIMBig - Track on Applied Natural Language Processing*, Lima - Peru.
- 2018 **Assistant**, *North American Chapter of the Association for Computational Linguistics*, New Orleans - USA.
- 2018 **Assistant**, *CHIST-ERA Projects Seminar*, Paris - France.
- 2017 **Reviewer**, *SIMBig - Track on Applied Natural Language Processing*, Lima - Peru.
- 2016 **Assistant**, *Latin American Conference on Computational Intelligence*, Cartagena - Colombia.
- 2015 **Assistant**, *International Conference on Intelligent Text Processing and Computational Linguistics*, Cairo - Egypt.
- 2014 **Assistant**, *International Conference on Computational Processing of Portuguese*, São Carlos - Brazil.
- 2014 **Assistant**, *International Conference on Intelligent Information and Engineering Systems*, Krynica - Poland.
- 2013 **Assistant**, *Simposio Peruano de Inteligencia Artificial*, Arequipa - Peru.
- 2011 **Assistant**, *Mexican International Conference on Artificial Intelligence*, Puebla - Mexico.
- 2011 **Assistant**, *Jornadas Peruanas de Computación*, Pucallpa - Peru.
- 2010 **Assistant**, *Simposio Peruano de Inteligencia Artificial*, Arequipa - Peru.
- 2010 **Assistant**, *Congreso Internacional de Dirección de Proyectos*, Arequipa - Peru.
- 2010 **Assistant**, *Congreso Internacional INFOTEK y FETEC*, Arequipa - Peru.

Languages

Spanish Fluent
English Advanced Level
Portuguese Advanced Level

native language
completed studies
completed studies

Programming Knowledge

ML Scikit-learn, Numpy, Pandas
RL Gymnasium, Ray
NLP NLTK, SpaCy, TextBlob
GenAI PyTorch, Transformers, LangChain
Languages Python, Java, C++
Others Git, Docker, Linux

Main Interests

- Automated Machine Learning
- Applied Machine Learning
- Reinforcement Learning
- Natural Language Processing
- Generative Artificial Intelligence