Waren Long

26, calle San Mateo, Madrid, Spain

📱 +33 6 77 80 44 89 • ✉ waren@sourced.tech • 🌐 warenlg.github.io
 warenlg • 🐦 warenlg • French & American citizenship

Experience

source{d} — **Madrid, Spain**
Machine Learning Software Engineer — *2017 to present – 2 years 7 months*

- Mine and build massive datasets from software repositories.
- Design and train ML models applied to source code using a multi GPU cluster.
- Develop solutions for code completion, program repair, clone detection, identity matching, developer similarity.
- Publish research papers and talk in top-tier Data Mining and ML conferences.
- Conduct product demonstrations and integrate source{d} solutions into customers' environment.
- Love and contribute to Open Source.

CEDRIC Lab — **Paris, France**
Research Software Engineer Intern — *2017 – 6 months*

- Implement a GRASP heuristic in Python to solve the Dial-A-Ride taxi-sharing Problem with Money as an incentive.
- Evaluation on public datasets of GPS taxi traces.

UCLA Department of Mathematics — **Los Angeles, USA**
Visiting Researcher — *2016 – 4 months*

- Design and implement in Matlab a macroscopic model to analyze traffic congestion using fluid mechanics equations.

Education

Conservatoire National des Arts et Métiers — **Paris, France**
MPRO Research Master's Degree in Computing Science, Operational Research — *2016 – 2017*

ENSTA ParisTech — **Paris, France**
Master of Science in Engineering, Optimization and Statistics — *2014 – 2017*

Pasteur College — **Neuilly-sur-Seine, France**
Preparatory classes PCSI/PC, Physics and Chemistry* — *2011 – 2014*

Technical skills

- **Languages:** English, French, Spanish.

- **Programming:** Python3, Go, C++, CUDA, SQL, Matlab, Julia, AMPL, LaTeX.

- **Front & BI tools:** HTML5, CSS3, JavaScript, React, Superset, Metabase, Tableau.

- **Models & Algorithms:** bag-of-words, n-grams, GBDT, bayesian, clustering, embeddings, neural nets, topic modeling, linear optimization, shortest path, PageRank, community detection.

- **ML frameworks:** TensorFlow, Keras, PyTorch, scikit-learn, DGL, NLTK, DyNet.

- **Technologies:** Linux, Git, Docker, Kubernetes, Apache Spark^{TM}.

- **Database:** MySQL, PostgreSQL, Cassandra.

Publications

Vadim Markovtsev and **Long Waren**. **Public Git Archive: A Big Code Dataset for All**. In *Proceedings of the 15th International Conference on Mining Software Repositories*, MSR, New York, NY, USA, 2018. ACM.

Vadim Markovtsev, **Long Waren**, Egor Bulychev, Romain Keramitas, Konstantin Slavnov, and Gabor Markowski. **Splitting source code identifiers using Bidirectional LSTM Recurrent Neural Network**. In *Proceedings of the 1st International FLOC Workshop on Machine Learning for Programming*, MLP, 2018.

Vadim Markovtsev, **Long Waren**, Hugo Mougard, Konstantin Slavnov, and Egor Bulychev. **STYLE-ANALYZER: Fixing Code Style Inconsistencies with Interpretable Unsupervised Algorithms**. In *Proceedings of the 16th International Conference on Mining Software Repositories*, MSR, Piscataway, NJ, USA, 2019. IEEE Press.

Long Waren, Vadim Markovtsev, Hugo Mougard, Egor Bulychev, and Jan Hula. **Identifying collaborators in large codebases**. In *Proceedings of the 1st International ICSE Workshop on Machine Learning for Software Engineering*, ML4SE, 2019.

Talks

- Machine Learning on Source Code : from n-grams to GGNNs [slides]
  ML Meetup 2019 - Nantes, France

- style-analyzer: Fixing Code Style Inconsistencies with Interpretable Unsupervised Algorithms [slides]
  MSR 2019 - Montreal, Canada

- Empower developers with ML assisted code review [slides]
  AI Meetup 2019 - Paris, France

- How AI can help programmers write better code? [slides]
  AI & Society Meetup 2019 - Paris, France

- Splitting source code identifiers using Bidirectional LSTM [slides]
  MLP 2018 - Oxford, UK

- Public Git Archive: A Big Code Dataset for All [slides]
  MSR 2018 - Gothenburg, Sweden

- Analyzing GitHub, how developers change programming languages over time [slides]
  PyDays 2018 - Vienna, Austria

Interests and Hobbies

- **Sport:**
  - Rugby player at Club de Rugby Liceo Francés
  - Golf (hcp. 7.3)
  - Scuba diving (N2 FFESSM, Advanced Open Water Padi)

- **Adventure:** Hiking and kayaking (Teide, Salcantay, Licancabur, Colorado river, Everglades).