

Wenqi Jiang

Data Engineer / Backend Engineer

About Me

Hello, my name is Wenqi Jiang.

I'm now studying a master's degree in **Data Science** at the Polytechnic University of Madrid (UPM) while working as an **ontology engineer** at Ontology Engineering Group (OEG).

Before that, I worked in mainland China for over three years as a **big data engineer** and **backend engineer**, so I have experience in handling Tbs data, and I am most skilled in: **Spark / Flink / Elasticsearch / Spring** and **Java / Scala / Python**.

Besides daily work, I am constantly looking for new challenges, I learned to **React / Flutter**. Recently, I have been working on **Web 3.0** in my spare time.

Category	Techniques
Big Data	Kafka, Hadoop, CDH, Sqoop, HBase, Hive, Impala, Spark, Flink, Elasticsearch
Backend	Maven, Gradle, Spring Boot, Dubbo, MyBatis, Hibernate, FreeMarker, Mysql, Redis, Nginx, Tomcat
Data Science	NLP, Ontology, Statistical Analysis, Machine Learning, Deep Learning, Time Series Mining
Frontend & Mobile	React.js, VUE.js, Scrappy, Flutter
Blockchain	Dapp, Solidity, Web3.js, Hardhat, Alchemy, Infura
DevOps	Docker, Github Action
Programming Language	Java, Scala, JavaScript, Python, Solidity, C++, R

Contact

- Mobile: KzM0IDY5MS03NTItMTU3
- Email: jiangwenqi1995@gmail.com
- Online Resume: jiangwenqi.info
- GitHub: github.com/jiangwenqi
- Blog: dev.to/jiangwenqi

Experience and Projects

Ontology Engineer (Scholarship)

OEG-UPM | Dec. 2021 - Present

Ontology Engineering Group (OEG) is located at the ETS de Ingenieros Informáticos and is formed by more than 30 people from worldwide. The group has more than 25 years of experience in the creation and use of semantic technologies, linguistic engineering and data integration.

LOOM-LD

Feb. 2022 - Present

A student **scholarship** for the development of linking algorithms and graph linkers.

- Analyzed the **ontologies** in different data sources;
- Specification and development of **linking discovery** algorithms, including text similarity and geographic relationship;
- Development of a **SPARQL-based** knowledge graph linker using customized **Apache Jena** functions;
- Evaluation of the algorithms with **OAEI-2021** datasets;
- Complement of the master's thesis.

Keywords: Ontology Graph Linker RDF Text Similarity DE-9IM WKT SPARQL Apache Jena OAEI

Big Data Engineer

TOGIC | *May 2019 - Aug. 2021 · 2 yrs 4 mos*

Shenzhen Togic Software Technology Co. was established in 2010. The company focuses on home Internet TV clients, with its Togic (similar to Netflix) and WEBBOX (similar to Apple TV) delivering a simple, intuitive, and lightweight experience for family entertainment, quickly becoming the most popular app on TV systems. There are currently over 700,000 daily users on this platform, and near 9 million customers worldwide used these products.

Offline Data Warehouse

Dec. 2019 - Jun. 2020

Togic's offline data warehouse serves as the company's data center, providing data support for daily operation, procurement, and decision-making.

- The data warehouse contains over **200 TB** of data and more than 150 tables, including user activity logs and system logs from the Togic App and Webbox System, which generate over **300 GB** of data every day.
- The data warehouse structure is **STAR** modelling and is organized into four layers: ODS, DWD, DWS, and ADS, which saved over **80%** storage space and eased the small files' problem on **HDFS**.

Keywords: CDH Kafka HDFS HIVE Spark

Data Statistics System

May 2019 - Mar. 2021

Offline Statistics System includes over 90 multidimensional statistics items, including user behavior, program playing, home page clicks and advertising etc.

- The statistics system was **restructured** based on the new Offline Data Warehouse, and almost all big data components **upgraded**, including **HDFS**, **Spark**, **Kafka** etc.;
- The code was rewritten by using **DataFrames**, and was more readable and maintainable, which **reduced the process time** to 1/4 of the previous one.
- Implemented more than **30 new statistical items** and generated reports;

- Fixed the bugs that the former system caused.

Keywords: Sqoop Hive Spark SQL Scala Elasticsearch MySQL MongoDB Python

Log Collection Service

Sept. 2019 - Sept. 2019

The log collection service is one of the most important service of the whole company, all data uploaded will be backed up into Data Warehouse for statistics.

- **Restructured** the log collection service to improve performance and stability;
- Used **Nginx** and **Netty** to receive and forward logs, and handle error data;
- Produced all logs into the message queue of **Kafka**, which has five brokers, three data backups, and seven days of log redundancy to **improve the robustness** of this service.

Keywords: Nginx Netty Java Kafka HDFS

Backend Engineer

WEBSTUDIO | Jul. 2018 - Apr. 2019 · 10 mos

Beijing Webstudio Information Technology Co. was founded in 2007 and focuses on quality-based and innovative development, providing data analysis and visualization, as well as official website marketing solutions which span a wide range of business disciplines and application situations.

- Implemented backend APIs of big data visualization dashboards and general reports for the **BI** applications;
- Developed more than 30 **data operators**, such as summation, date conversion, string processing etc.;
- Implemented the one-click function of syncing all user information;
- Expanded 16 data sources, including **MySQL**, **SQL Server**, **MongoDB** etc.;
- Built and maintained the big data platform;
- Fixed bugs and developed **new features** for the previous company's projects.

Keywords: Dubbo Spring Boot MyBatis Hibernate Java Nginx MySQL Redis Docker FastDFS CDH Hadoop HBase Impala Spark ML

Software Engineer (Internship)

KEG-THU | Feb. 2018 - Jul. 2018 · 6 mos

The Knowledge Engineering Group (KEG) of Tsinghua University established in 1996, and devotes to research on social network analysis, news mining, semantic web, knowledge graph construction, etc.

- Organized and annotated critical illness insurance documents;
- Trained models to extract the logical structure of insurance documents by using **GROBID**;
- Created a desktop application using above models;
- Maintained datasets in **NoSQL database**, like MongoDB, Neo4j;
- Completed the thesis.

Keywords: NLP CRF GROBID Neo4j MongoDB Java FX

Open Source

Dapp-Learning

Dapp Learning DAO | Apr. 2022 - Present

Definitive Guide for Decentralized-app(Dapp) Development on Blockchain.
Step-by-step Dapp practice through actual projects.

- Contributions of various tutorial tasks;
- Translations of the basic tasks;
- Shares of the latest blockchain news and cutting-edge technics in community.

Keywords: Blockchain web3.js Solidity Dapp Ethereum Hardhat Alchemy Infura DAO.

Education

MS Data Science

UPM | 2021 - Present

The Polytechnic University of Madrid (UPM) is a public university was founded in 1971 as the result of merging different Technical Schools of Engineering and Architecture, originating mainly in the 18th century. The Engineering and Architecture Schools of UPM have contributed significantly to the history of Spanish technology.

The major subjects include:

- Big Data / Cloud Computing and Big Data Ecosystems Design;
- Data Processes / Information Retrieval, Extraction and Integration / Time Series Data Mining;
- Statistical Data Analysis / Machine Learning;
- Open Data and Knowledge Graphs / Intelligent Systems / Deep Learning;
- Introduction to Research Methodology / Ethical, Legal and Social Aspects etc.

BE Computer Science & Technology

BISTU | 2014 - 2018

The Beijing Information Science and Technology University (BISTU) was founded in 2008 by merging the Beijing Machinery Industry Institute and the Beijing Information Engineering College, which are both created in the 19th century.

It offers an excellent academic discipline system in information science and technology and servers around 15,000 students from China and abroad.

The major subjects include:

- Principles of Computer Compositions / Computer Architectures;
- Compilation Principles / Practice of Embedded Application System;
- Operating System / Principles and Applications of Database;
- Advanced Math / Linear Algebra / Discrete Mathematics / Probability and Statistics;
- Programming Languages: C, C++, Java;
- Data Structure, Algorithm, etc.