

# 程世文

硕士 ◊ 计算机科学与技术 ◊ 1991年1月

(+86) 15201523580 ◊ chengshiwen0103@gmail.com

北京市海淀区清华大学东主楼9-325室, 中国, 100084

<http://chengshiwen.com/> ◊ <https://github.com/chengshiwen>

## 兴趣领域

---

**大数据:** 大数据分析 & 处理, 如大数据平台架构设计、大数据算法研究、大数据挖掘及分析  
**智能运维:** 智能化运维研发, 如性能测量、异常检测、异常定位、性能瓶颈分析、日志聚类  
**算法研发:** 算法研究与设计, 如基于特定场景的算法研究, C/C++算法、大数据算法研究  
**后端开发:** 后端设计与开发, 如服务器、缓存、数据库、负载均衡、分布式系统设计与开发

## 教育背景

---

**清华大学, 硕士** 2014年8月- 2017年7月  
计算机科学与技术系  
研究方向: 大数据分析 & 处理, 异常检测 & 定位, 互联网服务性能瓶颈分析

**清华大学, 学士** 2010年8月- 2014年7月  
计算机科学与技术系  
高二提前保送  
排名: 前20%

## 主要经历

---

**质量部架构组, 滴滴出行** 2016年10月- 2017年1月  
实习 北京

- ◊ 基于机器学习的海量非规范错误日志的清洗与聚类.
- ◊ 研究、实现并改进日志清洗算法和基于K-Means的聚类算法.
- ◊ 设计系统架构, 基于JStorm、Kafka、Django、Gunicorn、Redis、InfluxDB等开源方案.
- ◊ 数据请求峰值达到8万QPS, 每条日志4KB.

**大数据分析 & 处理系统, 清华大学计算机系网络所** 2016年3月- 2016年8月  
系统架构师 & 核心开发者 北京

- ◊ 负责设计大数据分析 & 处理系统的架构, 该系统可以对实时流数据和离线数据进行处理, 其使用了当前流行的开源框架, 包括Kafka, Logstash, Spark, Hadoop, OpenTSDB, HBase, Redis, MangoDB等.
- ◊ 该系统由数据采集、数据分发、数据处理、数据存储、数据查询、数据可视化组成.
- ◊ 数据处理、数据存储、数据查询、数据可视化模块的核心开发者.

**技术部研发, 北京水木快车科技有限公司** 2014年9月- 2016年1月  
联合创始人 & CTO 北京

- ◊ 北京水木快车科技有限公司的联合创始人和CTO, 主要产品是基于匿名社交的移动互联网应用: 枝丫.
- ◊ 领导管理技术部, 并对所有的技术方案进行决策.
- ◊ 后端的设计与开发者, 使用Python/Django、MySQL、Redis、负载均衡、分布式系统等技术, 涉及即时通讯, 消息推送, 短信通知, 云存储, CDN, 微信公众号, 微信支付, 统计分析等开发.
- ◊ 开发了一套自动化、可视化的运营工具.
- ◊ 巨大的成长: 专业技能, 沟通交流, 团队协作, 团队管理, 解决问题.

运维部BFE, 百度  
实习

2014年3月- 2014年8月  
北京

- ◇ 研究网页搜索页面访问次数 (PVs) 的异常检测.
- ◇ 研究并实现多种异常检测算法: 斜率、均值-阈值、时序分解(TSD)、滑动平均、ARIMA、CUSUM、三阶指数平滑(Holt-Winters)、聚类分析(K-Means)、主成分分析(PCA)、奇异值分解(SVD)等.
- ◇ 采用了手动标注、自适应、自学习等机制, 有效降低了误报率、漏报率, 并对节假日等做了特定优化.
- ◇ 设计并实现内部的异常检测系统, 使用了Python/Django、Memcached、并行化等.

edX, OPENedX (<https://open.edx.org/>)  
贡献者

2013年10月- 2013年11月  
开源项目

- ◇ 参与开发国外edX开源项目 (<https://github.com/edx/edx-platform>, 2900+ github stars) 并贡献代码.

学堂在线, 清华大学  
核心开发者

2013年5月- 2013年10月  
北京

- ◇ 学堂在线 (<http://www.xuetangx.com/>) 是清华大学研发的中文MOOC (大规模开放在线课程) 平台.
- ◇ 独立负责视频模块的设计与开发, 涉及视频前端播放器开发, 后端视频上传、存储、同步、与第三方CDN厂商合作开发等工作.
- ◇ 内部多个代码分支的核心开发者和审核者.

## 个人技能

精通	C/C++, Python, Linux, MySQL, Redis, Web Full Stack, Algorithm, Data Structure
熟练	Java, Scala, Shell, PHP, Memcached, MongoDB, SQLite3, HTTP, TCP/IP, Big Data
熟悉	Nginx, Apache, Nagios, Kafka, Spark, Mesos, HBase, OpenTSDB, Hadoop, Docker
工具	Git, SVN, Vim, Maven, yum, rpm, apt-get, pip, npm, Homebrew, Fabric, Markdown
语言	中文, 英语 (基本听说/良好读写)

## 奖项荣誉

新鸿基地产郭氏基金·清华大学励学金	2014年5月
清华之友—黄奕聪伉俪奖学金	2013年9月
清华大学学业单项优秀奖奖学金	2012年9月
清华大学计算机系优秀学生干部	2011年9月
国家奖学金	2011年9月
第25届中国数学奥林匹克 (CMO) 银牌	2010年1月
第24届中国数学奥林匹克 (CMO) 银牌	2009年1月

## 社会活动

计算机系党支部组织委员, 负责院系党支部各项活动的组织工作	2013年8月- 2014年7月
清华大学计算机系25班辅导员, 负责新生学习和生活引导及班级建设等工作	2012年8月- 2014年7月
清华大学团委实践组干部, 负责全校各院系的寒假和暑期实践的组织工作	2011年9月- 2012年8月
计算机系团委实践部干部, 负责计算机系寒假实践和暑期实践的组织工作	2010年9月- 2011年8月

## 个人评价

勤 奋	9 / 10	学习能力	10 / 10
沟通交流	8 / 10	团队协作	8 / 10
解决问题	9 / 10	责 任	9 / 10
专 注	9 / 10	创新能力	7 / 10

# Shiwen Cheng

Master ◇ Computer Science and Technology

(+86) 15201523580 ◇ chengshiwen0103@gmail.com

Room 9-325, East Main Building, Tsinghua University, Beijing, China, 100084

<http://chengshiwen.com/> ◇ <https://github.com/chengshiwen>

## INTERESTS

---

**Big Data:** Big data analysis and processing, e.g., architecture of big data, algorithm based on big data

**Operation:** Intelligent solution of operation, e.g., anomaly detection, performance bottleneck analysis

**Algorithm:** Research and design of algorithm, e.g., C/C++ algorithm, algorithm based on big data

**Backend:** Design and development of the backend, e.g., server, cache, SQL, load balancing, distributed

## EDUCATION

---

### **Tsinghua University**

*Aug. 2014 - Jul. 2017*

Master in Department of Computer Science and Technology

Research: Big data analysis and processing, Anomaly detection, Performance Bottleneck Analysis

### **Tsinghua University**

*Aug. 2010 - Jul. 2014*

Bachelor in Department of Computer Science and Technology

Early admission to Tsinghua in senior 2

Rank: Top 20%

## EXPERIENCE

---

### **Arch of QA, Didi Chuxing**

*Oct. 2016 - Jan. 2017*

*Intern*

*Beijing, China*

- ◇ Cleaning and clustering of large amounts of non-standard log, based on machine learning.
- ◇ Research, implement and improve the log cleaning and clustering algorithm based on K-Means.
- ◇ Design the system architecture, based on JStorm, Kafka, Django, Gunicorn, Redis, InfluxDB and etc.
- ◇ Support a peak total of 80K requests per second, and the size of each log is about 4 KB.

### **Big Data System, Network Lab at Tsinghua**

*Mar. 2016 - Aug. 2016*

*System Architect & Core Developer*

*Beijing, China*

- ◇ Design architecture of the big data system that can deal with streaming analysis and offline batch analysis, using popular opensourced systems, including Kafka, Logstash, Spark streaming, Hadoop, OpenTSDB, HBase, Redis, MangoDB and etc.
- ◇ The big data system consists of data collection, delivery, processing, storage, querying and visualization.
- ◇ Core Developer of the big data system for data processing, storage, querying and visualization.

### **Technology Department, THTrain Technology Co., Ltd**

*Sep. 2014 - Jan. 2016*

*Co-Founder & CTO*

*Beijing, China*

- ◇ Co-Founder and the CTO of an Internet startup company in Beijing, with a major mobile app **ZhiYa**.
- ◇ Lead the technology department and make a decision about all major technical solutions.
- ◇ Designer and developer of the backend, using Python/Django, MySQL, Redis, Load Balancing and etc, involving IM, Message Push, SMS, OSS, CDN, WeChat, Statistical Analysis and etc.
- ◇ Develop an automatic and visual web operation tool with jQuery, Bootstrap, ECharts and etc.
- ◇ Tremendous growth in professional skill, communication, team cooperation, team management.

**Baidu Front End of OP, Baidu Inc***Mar. 2014 - Aug. 2014**Intern**Beijing, China*

- ◇ Research anomaly detection of Web search page views (PVs).
- ◇ Research and implement anomaly detection algorithms: slope, Mean-threshold, TSD, MA, ARIMA, CUSUM, Holt-Winters, K-Means, PCA, SVD and etc.
- ◇ Support manual flag, self-adaption and self-learning to reduce FP and FN rate, and optimize holidays.
- ◇ Design and implement the internal anomaly detection system with Python/Django, Memcached.

**edX, OPENedX** (<https://open.edx.org/>)*Oct. 2013 - Nov. 2013**Contributor**Open Source*

- ◇ Contribute several patches to the open source project edx-platform (<https://github.com/edx/edx-platform>, 2900+ github stars), which is a subproject from OPENedX.

**XuetangX, Tsinghua University***May. 2013 - Oct. 2013**Core Developer**Beijing, China*

- ◇ XuetangX (<http://www.xuetangx.com/>) is a Chinese MOOC (Mass Open Online Course) platform.
- ◇ Independent development of video modules, involving video player, uploading, storage, CDN and etc.
- ◇ Core developer and reviewer of many internal code patches.

**SKILLS**


---

<b>Proficient</b>	C/C++, Python, Linux, MySQL, Redis, Web Full Stack, Algorithm, Data Structure
<b>Experienced</b>	Java, Scala, Shell, PHP, Memcached, MongoDB, SQLite3, HTTP, TCP/IP, Big Data
<b>Familiar</b>	Nginx, Apache, Nagios, Kafka, Spark, Mesos, HBase, OpenTSDB, Hadoop, Docker
<b>Tools</b>	Git, SVN, Vim, Maven, yum, rpm, apt-get, pip, npm, Homebrew, Fabric, Markdown
<b>Languages</b>	Chinese, English (Basic communication / good reading & writing)

**AWARDS**


---

Scholarship of Sun Hung Kai Properties in Tsinghua University	<i>May. 2014</i>
Qinghua's Friend - Huang Yicong Couple Scholarship	<i>Sept. 2013</i>
Scholarship of Academic Excellence in Tsinghua University	<i>Sept. 2012</i>
Excellent Student Cadre in Department of Computer Science and Technology	<i>Sept. 2011</i>
National Scholarship	<i>Sept. 2011</i>
Silver Medal of 25th China Mathematical Olympiad (CMO)	<i>Jan. 2010</i>
Silver Medal of 24th China Mathematical Olympiad (CMO)	<i>Jan. 2009</i>

**SOCIAL ACTIVITIES**


---

Commissary in Charge of Organization in Computer Science and Technology	<i>Aug. 2013 - Jul. 2014</i>
College Counselor of Class CST 25 in Tsinghua University	<i>Aug. 2012 - Jul. 2014</i>
Cadre of School League Practice Department in Tsinghua University	<i>Sept. 2011 - Aug. 2012</i>
Cadre of Practice Department in Computer Science and Technology	<i>Sept. 2010 - Aug. 2011</i>

**SELF-EVALUATION**


---

<b>Diligence</b>	9 / 10	<b>Learning</b>	10 / 10
<b>Communication</b>	8 / 10	<b>Cooperation</b>	8 / 10
<b>Problem Solving</b>	9 / 10	<b>Responsibility</b>	9 / 10
<b>Concentration</b>	9 / 10	<b>Innovation</b>	7 / 10