KEVIN KAO

San Jose, CA 95132 | 408.373.5520 | kkao@berkeley.edu
www.kevkao.com | github.com/kk415kk

EDUCATION

University of California, Berkeley May 2016
B.S. Electrical Engineering & Computer Science 3.63 – CS GPA

- Data Structures
- Machine Architecture
- Algorithms
- Computer Security
- Randomized Algorithms*
- Web Architecture
- Computational Biology*
- Data Science
- Artificial Intelligence
- Machine Learning
- Databases

Awards: Regents & Chancellor Scholar

PUBLICATIONS

[1] Zachary Pardos, Kevin Kao. "moocRP: An Open-Source Analytics Platform". October 2014. LAS 2015.
[2] Kevin Kao, Ananth Sub., Valerie Liu. "Machine Learning with fMRI Brain Images." December 2014. CS194-16 BIDS Poster Session.

SKILLS

Proficient with Java, Python, JavaScript, HTML/CSS, UNIX, Git, jQuery, SQL, MySQL, Redis, nginx
Experience with Scala, Node.js, Sails.js, Flask, Ruby on Rails, C, C++, Bash, Amazon EC2/S3, Mercurial, Underscore.js
Data Analytics Experience: Hadoop (MapReduce, HBase), R, Spark, D3.js, numpy, scikit-learn, OpenCV, BIDMach, Pandas

WORK EXPERIENCE

LogicBlox – Database Engineering Intern August 2014 – January 2015

- Researched cutting-edge academic papers on implementation of **dynamic SCC algorithm** under edge additions/deletions
- Designed and implemented internal graph structure optimized for **incremental graph maintenance**, avoiding full recomputes of SCCs

AutoGrid Systems – Software Engineering Intern May 2014 – August 2014

- Fully developed new vital portion of data analytics pipeline: an **incremental ETL algorithm** using multithreading, the star schema, Redis, MySQL, and MapReduce to allow scalable incremental computations on millions of data points by ML algorithms
- Developed a **communications platform** (email, SMS, voice) Ruby gem with an enhanced templating system as an infrastructure service
- Developed **internal monitoring app** with Python Flask to check config/status of applications, ensuring quick detection of problems

UC Berkeley School of Information – Full Stack Engineering Intern February 2014 – Present

- Developed **Node.js web app** called moocRP that serves as a data distribution and data analytics sharing platform for researchers
- Contributed significant data transformation Python and Bash scripts to **open-source Stanford/Harvard data analytics scripts**
- Designed **analytics sharing system** that allows users to upload D3 visualizations to be automatically scaffolded and shared publicly
- Implemented **secure datasets download pipeline** with SSL, CSRF-protection, CAS authentication module, GPG encryption

Past Positions: Undergraduate Researcher at **UC Berkeley MOOCLab**, SDET Extern at **Microsoft**, Founder of **Paw for Paw** non-profit

PROJECTS

Scribble.ly (Python, Flask, JavaScript, D3.js, HTML/CSS) October 2014 – December 2014

- Designed a single page application (SPA) with Python Flask that shortens URLs, using AJAX requests to dynamically update page content
- Implemented simple real-time analytics visitor tracking feature of shortened URLs with visualizations of the data in graphs through D3.js

fMRI Brain Image Analytics (Python, Scala, scikit-learn, skimage, OpenCV, numpy) October 2014 – December 2014

- Performed exploratory data analysis on fMRI images to classify voxel responses based on features drawn on stimulus images
- Developed classification, clustering, and regression models to predict brain responses, as well as data wrangling processing algorithms.
- Yielded significant results with binary classification of brain region activity, with approximately 90% accuracy and on par with baseline

Karpus-Strong Algorithm(Genetics/DNA Processing) (Python) September 2014 – October 2014

- Implemented a suffix array generation algorithm in log-linear time that can perform fast DNA/text processing, with auxiliary algorithms

SimpleDB Relational Database System (Java) January 2014 – May 2014

- Designed a transactional database that allows concurrency at a page level, with locking to prevent race conditions
- Wrote sort-merge algorithm and DP query optimization algorithm that generated close-to-optimal query plans

Web & Virtual Machine Security Penetration Testing (Python, C, Ruby, Bash) March 2014 – April 2014

- Wrote Python and Bash scripts to pipe hex data and perform buffer overflows, using GDB stack analysis, netcat, and nmap to scout
- Scanned vulnerable C code, writing a Python script to bruteforce RSA private key and Ruby script to spoof DNS responses to bypass SSL

MapReduce Text Processing on Amazon EC2 (Java, Hadoop, Amazon EC2) September 2013

- Utilized the Hadoop MapReduce framework to process large documents and calculate word relationships on Amazon's EC2 clusters
- Implemented efficient MapReduce algorithm, benchmarking a speed of 51 seconds, less than half required time (200% speedup)