

Samuel Brotherton

CONTACT INFORMATION

Phone: (404) 641-1221
Email: sbrother@gmail.com

LinkedIn: <https://www.linkedin.com/in/sbrother>
GitHub: <http://github.com/sbrother>

EDUCATION

Harvard University, Cambridge, MA

B.A., Mathematics and East Asian Studies

Sep 2008 – May 2012

Received highest honors for senior thesis analyzing over 200,000 Chinese blog posts, algorithmically detecting mutations in the Chinese language in response to censorship. Completed coursework in abstract algebra, Galois theory, topology, real and complex analysis, probability theory, and linguistics.

PROFESSIONAL EXPERIENCE

Google, Venice, CA

Software Engineer III

Jun 2014 – present

Worked on a small team using natural language processing and other machine learning techniques to improve advertisement quality. Led a 20% project related to mining semantic information from web data, which was adopted by several teams across different product areas. Built a named entity recognition system in C++ and a link detection algorithm that runs on very large graphs; contributed to a topic model for clustering semantic entities.

Whisper, Venice, CA

Software Engineer and Data Scientist

Mar 2013 – Apr 2014

Worked as sole data scientist at a rapidly expanding social media startup seeing upwards of three billion monthly pageviews. Designed and built an NLP service to extract topics and tags from posts, predict image searchterms from unstructured text, and target content to users. Implemented a new geographic search system using PostGIS that decreased search time by 90%. Worked closely with the front and backend development teams, writing production code in Erlang and Python.

General Assembly, Santa Monica, CA

Expert in Residence

Feb 2014 – May 2014

Helped teach an introduction to data science class to a group of 15 adult students. The class was based in Python, and included material on ETL techniques, regression and clustering, ensemble learning, natural language processing, and dimensionality reduction techniques.

Cairn Labs, Los Angeles, CA

President and Founder

Sep 2011 – present

Software development, data analysis, and technical writing for clients in the technology, academic, and healthcare industries. Most projects completed by myself, but frontend and DevOps subcontractors brought on as needed. Projects included:

- A scalable API server built in Elixir/Erlang, capable of serving and updating content recommendation streams at 200 QPS from a single commodity server.
- A graphical tool allowing decision-makers within the pharmaceutical industry to optimize total revenue of product launches across many countries. Involved building a fast probabilistic traveling salesman problem solver in F#, and building a user interface around it in C#/WPF.
- A cross-platform time series data analysis tool, including a visualization panel that allows researchers to watch their changes to analysis parameters in real-time.
- A distributed web crawling system using the PlanetLab framework. Harnessed the power of over 100 continuously running nodes across the world to analyze differences in how web pages are served to various countries.
- A user-friendly, scalable text organization and analysis system for researchers in quantitative social science, written in Python and powered by Apache Lucene.

SELECTED PUBLICATIONS

Stephens, M., Brotherton, S., Dunning, S., Emerson, L., Gilbertson, D., McClellan, A., McClellan, W. & Gitlin, M. (in press). “Geographic Disparities in Patient Travel for Dialysis in the United States”, *The Journal of Rural Health*.

PROGRAMMING EXPERIENCE

Languages: Python, C++, C#, F#, Erlang, Elixir, Mathematica, Haskell, Bash, C, L^AT_EX
Data Technology: Cassandra, Redis, PostgreSQL, ElasticSearch/Lucene, MySQL, MongoDB