

Stephen Reid McLaughlin

Data Scientist, Researcher, Teacher

1511 Faro Drive, Apt. 65 | Austin, TX, 78741
+1 (609) 217-4540 | steve.mclaugh@gmail.com
<https://github.com/stevemclaugh>

Education

- Master of Science in Information Studies (2018), University of Texas at Austin
- Bachelor of Arts in English (2008), University of Pennsylvania

Skills

- Python, Docker, SciPy stack, scikit-learn, Linux server administration, MongoDB, MySQL, writing, copyediting; some Java and JavaScript

Experience

- Data Analyst | University of Texas at Austin
April 2018 to present
I handle web scraping, data management, and data analysis tasks for faculty in UT's Communication department and School of Information. I am currently assembling a collection of 300+ million transaction records from the mobile payment service Venmo.
- Programming Instructor | Hello World Studio
August 2018
I co-taught a one-week course for high school students called "Amazon Alexa & Artificial Intelligence," held at Austin's Capital Factory. Students learned natural language processing techniques in Python, then designed and deployed custom Alexa skills using AWS Lambda.
- Assistant Instructor | University of Texas at Austin School of Information
January 2018 to May 2018
I designed and taught the undergraduate course "Programming for Cultural Data Analysis." We covered the basics of Python and the Linux shell, as well as web scraping, statistical analysis, and the fundamentals of machine learning.
- Research Assistant | University of Texas at Austin School of Information
August 2016 to December 2017
I developed software to index archival audio collections using machine learning techniques, in collaboration with WGBH and the American Archive of Public Broadcasting. I wrote a Docker-based tool for high-speed audio labeling, a Python package for managing audio labels, and several classifier pipelines for speaker recognition and event detection.

Software

- Audio Labeler, an application for rapidly generating audio training data. <https://github.com/hipstas/audio-labeler/>
- Audio Tagging Toolkit, a Python package for wrangling time-based labels and audio data. <https://github.com/hipstas/audio-tagging-toolkit/>
- SIDA (Speaker Identification for Archives), a set of template pipelines for training audio classifiers. <https://github.com/hipstas/sida/>
- Audio ML Lab, a portable environment for audio research using Docker. <https://github.com/hipstas/audio-ml-lab/>

Papers

- “Sci-Hub Provides Access to Nearly All Scholarly Literature.” 2018. Himmelstein, Romero, Levernier, Munro, McLaughlin, Tzovaras, and Greene. *eLife*. <https://elifesciences.org/articles/32822>
- “Collaborative Syllabus Design for Studying Information Work.” 2017. Buchanan, Bullard, Aspray, Bailey, Barker, Carter, Clement, Gottschlich, Howison, McLaughlin, Ocepek, Sholler, and Trace. *Proceedings of the Association for Information Science and Technology* 54 (1): 630–32. <https://doi.org/10.1002/pra2.2017.14505401094>
- “Shadow Libraries and You: Sci-Hub Usage and the Future of ILL.” 2017. Gardner, McLaughlin, and Asher. *ACRL 2017 Proceedings*. <http://eprints.rclis.org/30981/1/ShadowLibrariesandYou.pdf>
- “Measured Applause: Toward a Cultural Analysis of Audio Collections.” 2016. Clement and McLaughlin. *CA: Cultural Analytics*. <http://culturalanalytics.org/2016/05/measured-applause-toward-a-cultural-analysis-of-audio-collections/>

Selected Talks

- “Sci-Hub and LibGen in Perspective.” Perry-Castañeda Library, Austin, Texas. February 21, 2018.
- “Online Shadow Libraries and the Future of Humanities Scholarship.” Digital Humanities 2017, Montreal, Canada. August 10, 2017.
- “Applause in the Archive: Gauging Audience Response in Recorded Poetry.” Modern Language Association Convention, Austin, Texas. January 7, 2016.
- “Sound in Space: Tools for Regression-Based Audio Analysis.” High Performance Sound Technologies for Access and Scholarship Meeting, University of Texas at Austin. May 27, 2014.

Selected Workshops

- “Large-Scale Digital Preservation Under Adversarial Conditions.” UbuWeb in Athens, Athens, Greece. March 17, 2018.
- “Audio Analysis for Spoken Text Collections.” Shaping Humanities Data, hosted by the Collections as Data project. Digital Humanities 2017 preconference event, Montreal, Canada. August 7, 2017.
- “Text Analysis with Python.” Alice Kaplan Institute for the Humanities, Northwestern University. September 7, 2016.