**David Shumway | Project Portfolio** July 2024
Software Engineer & Data Scientist
160 North Elizabeth Street, Unit 1909, Chicago, IL 60607
E-mail: davidshumway@gmail.com | Phone: 1-773-759-5970
Github: @davidshumway | LinkedIn: @david-shumway-0b4661159
(*) available upon request

Related to Microbial Risks to Divers at Coastal Locations – 2021 - 2024

Tech: Python, LaTeX, SPARQL, RDF, OWL, Protégé, WebProtégé, Virtuoso, Dask, rdflib, OBO ROBOT, Overleaf

Visualization & Data Analytics: Pandas, matplotlib, GeoPandas, D3.js, Plotly, Google Colab + Google Drive, Google Sheets

Machine Learning & AI: time-series prediction, domain adaptation, transfer learning, supervised & unsupervised learning, PyTorch, TensorFlow, scikit-learn, Linear Regression, Random Forest, Decision Tree, SVM, MLP, CNN, XGBoost, LSTM, ADAPT, EasyAdapt, subspace alignment, CORAL, GRU, GNN, BERT, LLM

Popular APIs & Datasets: UMLS, Infectious Disease Ontology, WaterQualityData.us, EPA BEACON, OpenStreetMap API, NOAA Global Hourly, NOAA Tides & Currents, NOAA NDBC, EPA ECHO, EN4 Met Office Hadley Centre Ocean Salinity

Publications:

- (*) “A Domain Adaptation Approach for Predicting Recreational Water Quality at Data-Scarce Coastal Locations.” Primary author. Unpublished manuscript, submitted and revised for Environmental Science & Technology. 2023.
- Gautam, Nikita, et al. “Leveraging Existing Literature on the Web and Deep Neural Models to Build a Knowledge Graph Focused on Water Quality and Health Risks.” Proceedings of the ACM Web Conference 2023. https://dl.acm.org/doi/10.1145/3543507.3584185. Co-author.

Files:

- (*) Datasets and code for unpublished manuscript.
- (*) Recreational and Occupational Waterborne Illness Ontology.

Description: Developed waterborne illness ontology using Protégé and WebProtégé, and aided in use of BERT and LLMs for ontology learning. Built spatio-temporal time series datasets using Python (Pandas, GeoPandas), and built and evaluated supervised, unsupervised, and transfer learning models with scikit-learn, PyTorch, and TensorFlow.

FDIC Quarterly Banking App – 2023

Tech: Flask, DataTables.js, FDIC Banks API

Online: http://dlsl.pythonanywhere.com

Files:

- (*) Flask application.

Description: Created a single-page app using Flask, JavaScript, and DataTables.js to analyze FDIC API reporting data from 5,000+ US financial institutions, and implemented mechanisms for automated retrieval, caching, and CSV download of selected view.

**David Shumway | Project Portfolio** July 2024

Software Engineer & Data Scientist
160 North Elizabeth Street, Unit 1909, Chicago, IL 60607
E-mail: davidshumway@gmail.com | Phone: 1-773-759-5970
Github: @davidshumway | LinkedIn: @david-shumway-0b4661159

(*) available upon request

MALDI-DB and supporting research – 2019 - 2021

Tech: Django, PostgreSQL, R, Docker, Docker Compose, Podman, Nginx, RPlumber, NCBI Taxonomy, D3.js, LaTeX

Publications:

- (*) “Storage and Querying of Provenance in Scientific Workflow Management Systems.” Primary author. Unpublished manuscript, completed as part of PhD qualifying exam (UIC). 2020.
- (*) “A public community-driven database for sharing and execution of matrix-assisted laser desorption ionization time-of-flight mass spectrometry bacterial and protein spectral data and workflows.” Primary author. Unpublished manuscript, completed as part of Research Methods graduate course (CS590). 2020.
- (*) “Bacterial identification in a distributed mass spectrometry sensor network using streaming big data.” Primary author. Unpublished manuscript, completed as part of Algorithms for Big Data graduate course (CS594). 2019.
- (*) “ADEPT - An ontology for Atomic Distributed Experiments: Modeling Scientific Workflow Processes as Always-Available Cloud Services to aid in Workflow Reproducibility, Composition, Sharing, Resource Conservation, and Testability.” Primary author. Unpublished manuscript, completed as part of Data and Web Semantics graduate course (CS586). 2018.

Files: https://github.com/idbac/maldidb

Development screenshots: https://github.com/idbac/maldidb/tree/master/mdb/media

Description: Developed a Django web portal hosted on Mass Open Cloud with OpenShift for analysis of bacteria and small molecules from MALDI-TOF spectra, and integrated existing R spectra preprocessing functionality via RPlumber. Utilized separate containers for Django, PostgreSQL, R, and Nginx, and started the system as a whole using Docker Compose.

Notebook Reviewer: 2nd and 3rd Annual EarthCube Call for Notebooks – 2021, 2022

Description: Reviewed earth science-related programming notebooks for the 2nd and 3rd Annual EarthCube Call for Notebooks.

Zoomie: Round-robin breakout scheduling in Zoom – 2020

Tech: Firefox / Chrome browser extension

Files: https://github.com/davidshumway/zoomie

Description: Simple browser extension designed for a meditation group which utilizes Zoom's web-chat client to enable moderators to schedule one-on-one round-robin breakout rooms.

**David Shumway | Project Portfolio** July 2024
Software Engineer & Data Scientist
160 North Elizabeth Street, Unit 1909, Chicago, IL 60607
E-mail: davidshumway@gmail.com | Phone: 1-773-759-5970
Github: @davidshumway | LinkedIn: @david-shumway-0b4661159
(*) available upon request

(*) Manuscript: A Review of QA4IE – 2020

Abstract: This is a review of QA4IE: A Question Answering based Framework for Information Extraction. The authors of QA4IE construct a large information extraction (IE) benchmark and propose the QA4IE framework to reduce triple relation extraction to a question answering task. This article examines QA4IE from two aspects: QA4IE framework and QA4IE benchmark construction. We discuss both the effectiveness and efficiency of the QA4IE framework, as well as the comprehensiveness of evaluation measures. In addition, an extension of the QA4IE framework is explored. For QA4IE benchmark construction, we investigate how to augment the dataset and how to develop a domain-specific IE benchmark.

Description: Co-author. Unpublished manuscript, completed as part of Research Methods graduate course (CS590). 2020.

Chicago Public Schools Recruiting App – 2018

Tech: HTML, CSS, JavaScript, D3.js, Leaflet, Chicago Public Schools dataset

Online: https://dshumw2.people.uic.edu/cs594/CPS/

Files: https://github.com/davidshumway/CPS

Description: Group project. Developed 3 out of 5 visualizations, the search functionality, and tied the final program together into a single page app with linked views.

Related to Co-Owner & Lead Developer, Finger Lakes Business Systems – 2013 - 2018

Tech: JavaScript, PHP, HTML/CSS, MySQL, Python, Google Chrome API, Chrome/Firefox browser extensions, Amazon Mechanical Turk (AMT)

Files:

- (*) AMT marketplace-related browser extensions.
- (*) AMT marketplace-related MySQL databases.
- (*) AMT marketplace-related PHP scripting.
- (*) AMT task-related browser extensions.
- (*) AMT task-related MySQL databases.
- (*) AMT task-related PHP scripting.
- (*) AMT task-related Python scripting (Scrapy).
- (*) Browser extensions, SQL, and PHP scripting related to clients outside of AMT.
- (*) Apache Cordova latest AMT tasks mobile app.

Description: Collaborated with 750+ clients through Amazon MTurk, completing 1000+ unique projects and 3+ million microtasks while maintaining a 99.9% task approval rating, and continued off-site with key clients. Engineered in-house web scraping, data management, workflow optimization, and distributed systems solutions, writing 375K+ lines of in-house code in JavaScript, PHP, HTML, CSS, SQL, and Python, along with extensive use of the Google Chrome API and browser extension programming, including browser tabs, background scripts, event scripts, pop-up scripts, and storage.

**David Shumway | Project Portfolio** July 2024

Software Engineer & Data Scientist
160 North Elizabeth Street, Unit 1909, Chicago, IL 60607
E-mail: davidshumway@gmail.com | Phone: 1-773-759-5970
Github: @davidshumway | LinkedIn: @david-shumway-0b4661159

(*) available upon request

Tools for Amazon's Mechanical Turk – 2013 - 2018

Tech: Firefox / Chrome browser extension, HTML, CSS, JavaScript

Files: https://github.com/davidshumway/amt

Description: Developed and maintained a browser extension to improve worker productivity on the Amazon Mechanical Turk website with up to 10K weekly users.

Trumpocalypse – 2017

Tech: Python, Pygame

Files: https://github.com/davidshumway/Trumpocalypse

Description: Group project. Completed 350 out of roughly 500 commits toward a satirical politically-themed text-based game in the style of Oregon Trail.

Related to Digital Archiver & Technology Specialist, Paul Brunton Philosophic Foundation – 2010 - 2013

Tech: LAMP (Linux, Apache, MySQL, PHP), HTML, CSS, JavaScript, VBA, VB.NET, Tesseract, Bash, Visual Studio, LibreOffice

Files:

- (*) Epson scanning automation tool (VBA).
- (*) Microsoft Word, MySQL database full-text search form (VB.NET).
- (*) Tesseract OCR, MySQL database.
- (*) ImageMagick cropping and crop-review tool (LAMP, HTML, JavaScript, CSS).

Description: Automated the (EPSON) scanning process, resulting in 4x efficiency and saving two years of manual work. Created MySQL database full-text search form within Microsoft Word using VB.NET and Visual Studio. Utilized tesseract open-source OCR library to extract text from typewriter-written documents, saving the extracted text to a MySQL database, and implemented tools for 1) LibreOffice auto-complete using macros, and 2) duplicate hard-copy document checking. Utilized ImageMagick to automatically crop scanned white space from thousands of scanned typewritten documents, and created and used an interactive tool to validate cropped white space using thumbnails (HTML/JavaScript/CSS).

Related to Freelance Technology Specialist – 2005 - 2010

Tech: Linux, Apache, MySQL, PHP, HTML, CSS, JavaScript, Audio transcription

Online: https://www.freelancer.com/u/telosMed (sample work history)

Description: Provided freelance tech services in person and on freelance sites such as Rent-A-Coder, Freelancer.com, and MTurk.com.