|  |  |  |
| --- | --- | --- |
| **David BLACK** |  |  |
|  |  |
| [github.com/bballdave025](https://github.com/bballdave025) | [linkedin.com/in/thedavidwblack](https://linkedin.com/in/thedavidwblack) | | +1 801-376-4520 | | [thedavidwblack@gmail.com](mailto:thedavidwblack@gmail.com) |
| [stackexchange.com/users/8693193](https://stackexchange.com/users/8693193) | | 75 Peregrine Ln. #8 Bountiful, UT 84010 USA | |

MOTIVATION

This position, as a Data Scientist using Computer Vision & Natural Language Processing at Ancestry, is an intersection of my passions & interests. I have always intended to be a scientist at the frontier of research. Also from an early age, I've assisted my father in his effort to source & verify our family history. In this role, I can combine these two interests & talents with three more – data extraction from older documents, linguistics, & coding. At a previous job, I supported work on automated information extraction from manuscript & machine-created text. In my most-recent job, I was able to expand my skills in NLP as I worked on the Speech Science & Machine Learning team. I would love to use my initiative & creativity in CV/NLP tasks for family history & especially to continue with the use of Neural Networks for data extraction.

HIGHLIGHTS

» Over 5 years' experience in cutting-edge machine-learning research for automated

handwriting & speech recognition

» 11+ years' experience in computation research & software design

» At least 6 months' experience in each of the following roles

» Programming the underlying shape-detection algorithms for Asian handwriting recognition

& participating in Neural Network model creation & coding; work mostly done in Keras /

TensorFlow but also learned from PyTorch implementations

» Volunteer work with Dr. Ryan Farrell (BYU) in computer vision

» Simulation of physical processes to compare with particle-collision data (27 publications)

SKILLS

» Finding, reading, & understanding scientific papers as well as applying their principles to set

up experiments & to create practical implementations

» Helping patrons to successfully locate, analyze, & apply learning from historical records so as to lead to the discovery of additional records & information sources

» Initiative & satisfaction in finding new problems & solutions

» Handling the flow of large data-collection efforts in academic & industry-related research

» Interpersonal interactions

» Implementing shell-script solutions – usually bash, but also other major shells – into Python to increase portability &, for some projects, expand into an object-oriented framework

» Communication with business executives & legal counsel to show the financial implications & ethical concerns involved with research projects

» Paleography - i.e., the understanding & transcription of handwriting in multiple languages & from various time periods

» Coding in many programming languages using various software packages, APIs, etc.

» Ability to quickly learn & use new concepts, software, languages, & tools

RELEVANT WORK EXPERIENCE

**AI Labs Team Member** *(May 2018 – Mar 2023) CaptionCall / Sorenson Communications (Salt Lake City, UT)*

» As part of the original Speech Science & Machine Learning team, helped set priorities & determine direction & tools, & collaborated closely with the other original team members

» Initially conducted substantial literature review & decided implementation of findings ||

Python, scikit-learn, NumPy, SciPy, matplotlib, PIL, Pandas, NIST's sclite, kaldi

» Taught other team members to use Linux® & Linux-type implementations; debugged & adapted package builds & software for our needs || bash, C, C++, make, CMake, Perl

» Automated scoring || wrote Python wrapper for NIST's sclite executable

» Normalization of different human & machine transcribers' output so as to have only

words as pronounced || extensive use of regexes & customized algorithms

» Designed, developed, distributed, & maintained normalization tool as a Python package

» Used a Microsoft-based version control system at the start but used git the last 3 years

» Participated in a transition to an Agile-based system accompanied by a move to Jira

» Broad project to use the best of annotation from humans & machines to increase accuracy

» Created customized datasets leveraging call-recording software on which I worked through managing a freelancer || Python, Twilio, Amazon Elastic Beanstalk, AWS NoSQL

» Silence detection & removal to allow more words per unit time in data collection, implemented Hidden Markov & Support Vector Machine models || pyAudioAnalysis

» Iterative break testing for changes in software; used my knowledge of the intricacies of language & its representation along with coding experience to create robust testing material

» U.S. Patent No. 11017778, Issued May 25, 2021; One of five inventors

» Participated in a recent broadening of ML tools based on speech; attended informational

& demo meetings, particularly concerning end-to-end, real-time ASR & concerning conversation summarization along with question answering using transformers

» Solved the majority of the team's file encoding & decoding issues for text & audio files

**Document Specialist** *(Aug 2017 – April 2018) Church History Dept. – Granite Mountain Vault (Salt Lake City, UT)*

» Digitized microfilm of family-history-related documents using proprietary software

» Viewed a wide variety of documents & kept notes of interesting documents; Having just finished my work with automated image classification, zoning, & textual continuity - all for automated extraction for individual & relationship information with the research team - I focused on potentially-useful document types that could further such extraction

**Data Analyst, Contractor** *(July 2016 – July 2017) FamilySearch (Salt Lake City, UT)*

» Provided material for the Advanced Technology Research team as part of a group that provided a bridge between it & the Records Division

» Found, prepared, analyzed & helped facilitate annotation of records. Discussed algorithms & analyzed data patterns with the research team

» Data curation, reviewed & selected document images for datasets, particularly a corpus of ~20k documents in 12 languages that cover patron demand. Family Search catalog use. Efficient data storage & retrieval || bash tools, Python, Perl, Unicode, & utf-8

» Software design for data annotation. Object-oriented Java programs which allowed image classification, image segmentation, segment classification, & transcription. Worked closely with volunteers doing the annotation. Facilitated major acceleration of the zoning & classification || Java, Windows CMD Scripting

» Software design for evaluating Asian language OCR software || Java, JNI (Java Native Interface), C++, NIST's sclite

**Undergraduate & Graduate Research Assistant** *(2008 – 2013)*

*Brigham Young University (Provo, UT)* *(2008 – 2010)*

*University of California at Riverside (Riverside, CA)* *(2010 – 2013)*

» Data flow management & real-time data quality assurance at the Relativistic Heavy Ion

Collider || Postgre SQL, csh/tcsh, ksh, zsh, bash, Perl, PHP, HTML || UCR

» Simulation of physical processes for comparison between theory & experiment || C++, large & custom C++ libraries & wrappers, Monte Carlo methods || UCR

» Computation of electromagnetic fields & intensities || MATLAB®, Mathematica || BYU

EDUCATION

» **Master of Science, Physics**, University of California at Riverside, 2012

» **Bachelor of Science, Physics**, Brigham Young University, 2010 || Minors: **French, Mathematics**