David BLACK

linkedin.com/in/thedavidwblack +1 801-376-4520 thedavidwblack@gmail.com
stackexchange.com/users/8693193 75 Peregrine Lane #8 Bountiful, UT 84010

MOTIVATION

This position as a Research Scientist in the Advanced Technology Research team at FamilySearch is a confluence of my passions, interests, & beliefs, with a purpose that I find supremely gratifying. Since I can remember, I have wanted to be a scientist at the frontier of research that helps people. From an early age, I tagged along with my dad on his visits to the Family History Library. Now, in this position, I can combine these two interests with two more of my foremost passions & talents – linguistics & programming. The workplace not only has a focus on faith, but it gives a mandate to use it in one's work. In all of this, I know – due to my previous work in helping this team – that I will find two conditions I crave in work & life: the use of my initiative & creativity thanks to the independence granted; & the enjoyment of interactions & teamwork with kind, intelligent, & faithful team members who rely on the guidance of the Savior.

HIGHLIGHTS

- Over 5 years' experience in cutting-edge machine-learning research for automated handwriting & speech recognition
- 11+ years' experience in computation research & software design
- 2 years' experience teaching at a post-secondary level
- At least 6 months' experience in each of the following roles
  - Oversight of data flow & manipulation at a particle accelerator
  - Simulation of physical processes to compare with particle-collision data (27 publications)
  - Programming the underlying shape-detection algorithms for Asian-language transcriptions
  - Substitute teaching for K-12 students, especially science & foreign-language immersion
  - Volunteer work with Dr. Ryan FARRELL (BYU) in computer vision

SKILLS

- Finding, reading, & understanding scientific papers as well as applying their principles to set up experiments & to create practical implementations
- Helping people from varied backgrounds to locate, analyze, & apply learning from historical records so as to lead to the discovery of additional records & information sources
- Initiative & excitement in finding new problems & solutions
- Data compilation, creation, curation, annotation, assimilation, appreciation, utilization, interpretation, & dissemination
- Handling the flow of large data-collection efforts in academic & industry-related research
- Kindness, likeability, social aptitude, & the sincere desire not only to take notice of personal service opportunities that come, but to seek out such opportunities
- Listening
- Using `Linux`® tools & the pipeline of text-based input & output to gather, parse, sort, and, generally, manipulate data; I love to get a problem or task for which my skills in `sed`, `awk`, `grep`, `find`, `sort`, `shuf`, or related tools allow me to provide a quick solution or a solution which allows the correct manipulation on a large amount of data
- Implementing shell-script solutions – usually `bash` shell, but also in other major shells – into `Python`
- Communication with business executives & legal counsel to show the financial implications and ethical concerns, respectively, accompanying research projects
- Lifelong learning, by study & also by faith
- Understanding the underpinnings & applications of Speech Science with a breadth of knowledge including physical characteristics of the vocal tract, phoneme models, language

models, grammar, transcription, & metrics; I have also had recent discussions with a group using transformers to perform summarization & reasoning from transcribed speech
» Paleography in multiple languages from multiple language families
» Coding in many programming languages & using various software packages, APIs, etc.; more of these will be found next to the descriptions of projects where I used them; ability to quickly learn & use new tools

WORK EXPERIENCE

Speech Science & Machine Learning (AI Labs) Team Member *(May 2018 – Present)*
CaptionCall / Sorenson Communications (Salt Lake City, UT)

» As part of the original team, helped set priorities, determine direction, choose tools, & collaborate closely with the other original team members.
» Initially conducted substantial literature review & decided implementation of findings || `Python`, `scikit-learn`, `NumPy`, `SciPy`, `matplotlib`, NIST's `sclite`, `kaldi`
» Taught other members of the team to use `Cygwin`® as our `Linux`-type implementation; this involved debugging & adapting package builds & software || `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 `sed`, `awk`, `grep`, & `Perl` in development; implementation included NIST's `rfliter1` with additional `regex`es & customized algorithms
» Designed, developed, distributed, & maintained the normalization tool as a `Python` package || `setuptools`, `pip`
» Communication with management & pass-offs to engineering with the normalization software || `C#`, `Microsoft Visual Studio` & `Azure DevOps`
» Performed dataset management for purchased, open-source, & created datasets || `sed`, `awk`, `grep`, `find`, `vox`, `ffmpeg`, `mediainfo`
» For the creation & curation of new datasets, used customized hardware, call-recording software – implemented by a German freelancer, & experimental design || `Python`, `Twilio`, `Amazon Elastic Beanstalk`, AWS `NoSQL`
» Silence detection & removal to allow more words per unit time, implementing HMM & SVM models || `pyAudioAnalysis`
» Iterative break testing for changes in the normalization software; used my knowledge of the intricacies of language & its representation along my significant experience to create robust testing material
» Solved the majority of file encoding & decoding issues for text & audio files
» Participated in a recent broadening of ML tools based on speech; attended informational and demo meetings, particularly concerning end-to-end, real-time ASR & concerning conversation summarization along with question answering using transformers

Document Specialist *(August 2017 – April 2018)*
Church History Department – Granite Mountain Vault (Salt Lake City, Utah)

» Digitized microfilm of family-history-related documents using proprietary software worked on by Heath NIELSON & Dr. Bill BARRETT
» Viewed a wide variety of documents & kept notes with DGS of interesting documents. Having just finished my work with Family Search automatic extraction effort with the Advanced Technology Research team, I especially noted things that could help the effort to expand or could cause difficulties for it.

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

» 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, especially Pat SCHONE & Heath NIELSON
» Data curation, especially for the 2-to-20 effort. Family Search catalog use, efficient data storage & retrieval || `bash` tools, `Python`, `Perl`, `Unicode`, & `utf-8`
» Software design for data annotation. Worked closely with Senior & Service missionaries who used the software on which I worked. This software included resources for image classification, image segmentation, segment classification, & transcription. Allowed 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`
» Testing for encoding compatibility, annotation validity, & degree of task completion; included finding files whose annotation had been missed, left incomplete, or corrupted || `bash` tools, `Python`, `Java`

Assistant Professor of Mathematics and Science *(July 2014 – December 2015)*
Snow College (Ephraim, Utah)

» Planned & taught courses in remedial math, college algebra, physical science – including astronomy, Earth science, chemistry, physics, & college physics
» Physical science courses were taught simultaneously as on-campus & distance-learning courses
» Focused on technology in the classroom; encouraged students to use `Wolfram Alpha` to learn concepts & check answers; created simulations of physical processes, notably of Solar System formation from randomly-placed particles, which students could run multiple times || `Adobe Flash ActionScript` (very similar to `JavaScript`)

Undergraduate and 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`, shell tools for parsing & managing data, `Perl`, `PHP`, `HTML` || UCR
» Simulation of physical processes for comparison between theories & experiments, including theoretical & simulation checks for particle identification with already-existing components || `C++`, large & custom `C++` libraries & wrappers, probability distribution functions, Monte Carlo methods || UCR
» Computation of electromagnetic fields & intensities || `MATLAB®` || BYU
» Finding experimentally-discoverable differences between two competing quantum-dynamical theories || `Mathematica`, `MATLAB®` || BYU

EDUCATION

» **Master of Science, Physics**, University of California at Riverside, 2012
» **Bachelor of Science, Physics**, Brigham Young University, 2010
Minors: **French, Mathematics**

MISSIONARY SERVICE: France Paris Mission *(2005 – 2007)*