Data Engineer

About

Quorate Technology is a specialist developer of high-performance, custom-built Automatic Speech Recognition systems. Quorate is an employee-owned company formed in 2012 as a spin-out from the world-leading Centre for Speech Technology Research at The University of Edinburgh. We are a friendly, relaxed and sociable team who help each other, whilst being highly professional, client-focussed and passionate about what we do.

Opportunity

We are now looking to recruit a Data Engineer to work on our data processing pipeline. This role will involve liaising with clients to obtain large quantities of data for training speech processing systems; organising and documenting training data; using the existing pipeline to create datasets suitable for building speech processing systems; and developing tools and processes to improve the efficiency and portability of the pipeline. You will also have the opportunity to work more generally on our Python codebase, contributing to systems design and development and conducting code reviews.

The role will allow you to develop a thorough understanding of the data requirements for state-of-the-art machine learning systems. Your work will have a direct impact on the quality and performance of our customised speech processing systems. As our client base continues to expand, it is anticipated that you will become the lead developer of a team of data processing engineers and we will support you to develop your leadership potential.

Responsibilities

- Liaise with clients about data (audio and text) needed for training custom systems

- Obtain and organise client data

- Carry out text and audio data processing and normalisation

- Development and maintenance of the data processing pipeline

- Generate training and evaluation data sets

- Support the Machine Learning researchers

What we offer:

We will provide you with:

- A collegial and entrepreneurial working environment

- Real responsibility

- On the job support, training and learning

- External courses and certifications

- Competitive salary

- Share options

- Pension contributions

- Cycle to Work Scheme

- Plenty to do

You must have:

- Excellent Python programming skills

- A degree in Computer Science and/or considerable, relevant work experience

- A good command of regular expressions

- Knowledge of basic Unix command line tools

- Knowledge of information security and privacy best practices

- Excellent written, verbal communication and interpersonal skills

- Attention to detail

- The ability to work effectively in a team environment

- Patience, resilience and the ability to work under pressure

Experience with some or all of the following would be desirable:

- Advanced Unix command line tools

- Unix shell scripting (Bash)

- Git and Bitbucket

- Amazon Web Services, and in particular S3, EC2 and EFS

- XML

- Multiprocessing and parallelisation in Python

- HPC software (SGE or similar)

How to Apply

Please contact info+dataengineer@quoratetechnology.com with your CV and covering letter.

Thank you!