Job Description for Software Engineer

The PROSE team under CVP Joseph Sirosh has the mission to build Programming-by-Example (PBE) technologies for data wrangling (DW). Data wrangling is the process of transforming data wherever and in whatever format it is found (in log files, spreadsheets, webpages, emails, XML, pdf, relational DBs, document stores, etc.) to a structured form amenable to analysis/visualization. We live in a world of data and data analysis, and DW is what makes it all possible: from custom field extraction in log files to web scraping, or extraction of user data from emails (flight itineraries, expense reports) to joining data from completely different sources to gain new insights. Unfortunately, DW is challenging: Data scientists spend 80% of their effort in DW, and 99% of computer users do not have the programming background to be effective at DW. The good news is that PBE can be used to create DW scripts from examples, enabling delightful DW experiences for non-programmers. We have developed state-of-the-art PBE technologies at MSR for the past 5 years and worked with product teams to ship them yielding successes like the Flash Fill feature in Excel 2013, the ConvertFrom-String cmdlet in Powershell in Windows 10, and the custom field extraction capability in Azure Operations Management Suite. Now we have moved from MSR to join a bigger Data Wrangling initiative so that we can increase the pace of our innovation and get those innovations into the hands of customers.

The charter of our team is to build a platform and SDK for PBE technologies for various DW tasks (such as string/date/number transformations, extraction from log files/webpages/XML/JSON/semi-structured spreadsheets, sorting, filtering, grouping, parsing dirty CSV files, joins, series creation). This SDK will be used to build and bring the overall data wrangling product to the market. At the same time other product teams within Microsoft will also make use of the SDK including Powershell, Azure OMS, Exchange, and PowerBI among others.

The PROSE team consists of a mix of researchers who have been involved with efforts in MSR, engineers, PMs and UX researchers with experience shipping products. We are now looking to expand the team with another Software Engineer who will work alongside other team members to contribute to the core framework, take charge of UI development related to PBE technologies and its deployment on the playground to facilitate data collection, and help drive build, test and debugging processes for the team. The SDE should be able to become fluent in technical ideas and the vocabulary of PBE research (with guidance from researchers), contribute to brainstorming, and work closely with researchers.

Qualifications:

1. Few years of experience developing and shipping software.
2. Expert in front-end web development technologies including Javascript/Typescript.
3. Knowledgeable about parsing, ASTs, program semantics, symbolic manipulation, search algorithms. University-level course in compilers, formal methods, or program analysis required.
4. Expert in writing efficient managed code. Good understanding of .NET runtime internals (GC, JIT, Framework implementation) and its performance characteristics is a plus.
5. Undergraduate CS degree or equivalent.

Resources:

-PROSE v0.1 non-commercial release
-5 minute video of demo at MLDS (Microsoft Internal only), FlashExtract video
-Sumit's move from MSR to C&E: An approach to accelerated innovation
-Data Wrangling introduction

-FlashFill paper, FlashExtract paper, FlashRelate paper, FlashMeta core framework paper, FlashProg UI paper

Contact:
Sumit Gulwani
sumitg@microsoft.com
http://research.microsoft.com/en-us/um/people/sumitg/