

IAN DROSOS'

DISSERTATION DEFENSE

Synthesizing Transparent and Inspectable Technical Workflows

Abstract:

The highly technical and complex workflows found in data science and programming have led to many barriers that hinder even an expert while they wrangle code inside their computational notebooks or write code for their data analysis. Further, when these experts want to share these workflows online and act as content creators, they run into similar barriers to the work being done effectively and efficiently. The goal of this research is to discover and remove barriers and frustrations felt by data scientists and programmers in their daily workflows, including data wrangling and the broadcasting of these workflows online.

My thesis is that being able to see and interact with technical workflows behind the scenes can help novices grow into experts, but existing user-friendly computational systems may hinder growth by obscuring important kinds of complexity.

This dissertation supports my thesis by performing three studies. The first study investigates current issues with data wrangling and tools that assist data scientists in their daily data wrangling workflows, and introduces a unified interaction model based on programming-by-example that generates readable code for a variety of useful data transformations, implemented as a Jupyter notebook extension called Wrex. The second study investigates how technical expert content creators, such as programmers and data scientists, are teaching naturalistic workflows via livestreaming on sites like Twitch and YouTube, and how these might lend themselves to a virtual form of cognitive apprenticeship. Finally, the third study investigates the current state of livestreaming equipment setups and develops a design space defining the range of system setups and the challenges faced by content creators when they must build, run, and maintain these complex systems.

This dissertation found that the ability to read and extend the code that performs data transformations is a critical feature for new wrangling tools that use programming-by-example to be acceptable to users. Then it found that instructional workflow streams can provide a virtual form of cognitive apprenticeship by revealing serendipitous teachable moments that would otherwise be cut from polished tutorials. Finally, this dissertation contributes a design space for streaming setups to provide insights on the complexity dealt with by streamers in building and maintaining these setups and to provide guidance for novel tools to assist content creators in creating these setups.

IAN DROSOS
TO DEFEND

FRIDAY
MAY 27, 2022
AT 10 AM

VIA [ZOOM](#)

COMMITTEE MEMBERS:

Philip Guo, Chair,
Cognitive Science

Jim Hollan, Cognitive Science

Ranjit Jhala, CSE

Nadezhda Polikarpova, CSE

Brad Voytek, Cognitive Science

