Tomas Petricek — Dissertation Preface

http://tomasp.net | tomas@tomasp.net

Most of my research to date has been focused on the subject of programming languages and systems. The topic can be studied from multiple perspectives and in multiple contexts and my work explored multiple such research directions, each spanning multiple individual projects and papers.

I studied theory of (i) *context-aware programming* (in my PhD thesis), worked on (ii) *functional programming* and the F# language (in parallel to my PhD studies), programming in the context of (iii) *data exploration* (during my two Post-doctoral stays) and researched (iv) *interactive programming systems* (since joining University of Kent). I also devoted a part of my time to (v) *interdisciplinary research*, focusing on the history and philosophy of programming.

In each of those areas, my work has had (or is showing early signs of having) a lasting academic and industrial impact. The annotated selection of the five most important works serves as the evidence for this fact. It includes five significant academic works, each representing one of the five aforementioned research areas. The five selected works have been widely cited, received significant academic awards, inspired international grant applications, influenced main-stream programming languages or have been subject to invited keynotes.

My Habilitation dissertation *Simple programming tools for data exploration* is centred on one of the five research directions outlined above, choosing depth over breadth. It presents my work on data exploration, which is an independent research direction that I pursued following my PhD and for which I secured funding from multiple competitive sources. The research employed theoretical programming language knowledge in the practical area of working with data. As such, the work is of a more applied nature than some of my earlier research.

The dissertation provides a comprehensive summary of my work on data exploration, including key theoretical publications, publications accompanying influential software artifacts, as well as publications that complement the core theory with the practically necessary applied developments. The aim of the dissertation is to show the work in context and highlight connections that may not be apparent when reading individual papers. I also use the longer format of a dissertation to highlight interesting work that has been perhaps less widely recognized.

The key theoretical results in the dissertation rethink established programming language notions or use them in a new way. These have been published in top conferences including CORE A* (Chapters 6 and 13) and CORE A (Chapter 7), indexed journals (Chapters 11, 12) or received best paper awards (Chapter 6 and 8). The open-source software developed as part of the work is arguably of equal importance. One library (F# Data) has become widely adopted and is being further developed by industry. Another package (Fluid) is being used as the basis for further development at the Institute of Computing for Climate Science at University of Cambridge. Finally, two other software packages (The Gamma and Compost.js) have been subject to multiple talks and invited keynotes at industry conferences.

Viewing data exploration as a programming language problem proved to be a mutually beneficial appraoch. On the one hand, it lets us apply a range of established programming language techniques to improve data science tools. On the other hand, it raises interesting questions about those tools—some of which I started exploring in my most recent work, and intend to continue exploring in the future.