Foreword

Arnold Robbins and I are good friends. We were introduced 11 years ago by circumstances—and our favorite programming language, *awk*. The circumstances started a couple of years earlier. I was working at a new job and noticed an unplugged Unix computer sitting in the corner. No one knew how to use it, and neither did I. However, a couple of days later it was running, and I was `root` and the one-and-only user. That day, I began the transition from statistician to Unix programmer.

On one of many trips to the library or bookstore in search of books on Unix, I found the gray *awk* book, a.k.a. Aho, Kernighan, and Weinberger, *The AWK Programming Language* (Addison Wesley, 1988). *awk*'s simple programming paradigm—find a pattern in the input and then perform an action—often reduced complex or tedious data manipulations to few lines of code. I was excited to try my hand at programming in *awk*.

Alas, the *awk* on my computer was a limited version of the language described in the *awk* book. I discovered that my computer had "old awk" and the *awk* book described "new awk." I learned that this was typical; the old version refused to step aside or relinquish its name. If a system had a new *awk*, it was invariably called *nawk*, and few systems had it. The best way to get a new *awk* was to *ftp* the source code for *gawk* from `prep.ai.mit.edu`. *gawk* was a version of new *awk* written by David Trueman and Arnold, and available under the GNU General Public License.

(Incidentally, it's no longer difficult to find a new *awk*. *gawk* ships with Linux, and you can download binaries or source code for almost any system; my wife uses *gawk* on her VMS box.)

My Unix system started out unplugged from the wall; it certainly was not plugged into a network. So, oblivious to the existence of *gawk* and the Unix community in general, and desiring a new *awk*, I wrote my own, called *mawk*. Before I was finished I knew about *gawk*, but it was too late to stop, so I eventually posted to a `comp.sources` newsgroup.

A few days after my posting, I got a friendly email from Arnold introducing himself. He suggested we share designs and algorithms and attached a draft of the POSIX standard so that I could update *mawk* to support language extensions added after publication of the *awk* book.

Frankly, if our roles had been reversed, I would not have been so open and we probably would have never met. I'm glad we did meet. He is an *awk* expert's *awk* expert and a genuinely nice person. Arnold contributes significant amounts of his expertise and time to the Free Software Foundation.

This book is the *gawk* reference manual, but at its core it is a book about *awk* programming that will appeal to a wide audience. It is a definitive reference to the *awk* language as defined by the 1987 Bell Labs release and codified in the 1992 POSIX Utilities standard.

On the other hand, the novice *awk* programmer can study a wealth of practical programs that emphasize the power of *awk*'s basic idioms: data driven control-flow, pattern matching with regular expressions, and associative arrays. Those looking for something new can try out *gawk*'s interface to network protocols via special */inet* files.

The programs in this book make clear that an *awk* program is typically much smaller and faster to develop than a counterpart written in C. Consequently, there is often a payoff to prototyping an algorithm or design in *awk* to get it running quickly and expose problems early. Often, the interpreted performance is adequate and the *awk* prototype becomes the product.

The new *pgawk* (profiling *gawk*) produces program execution counts. I recently experimented with an algorithm that for *n* lines of input exhibited ~ Cn^2 performance, while theory predicted ~ *Cn log n* behavior. A few minutes of poring over the *awkprof.out* profile pinpointed the problem to a single line of code. *pgawk* is a welcome addition to my programmer's toolbox.

Arnold has distilled over a decade of experience writing and using *awk* programs, and developing *gawk*, into this book. If you use awk or want to learn how, then read this book.

<div align="right">

Michael Brennan
Author of *mawk*

</div>