

FOURTH EDITION

Effective awk Programming

Arnold Robbins

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'REILLY[®]

Table of Contents

Foreword to the Third Edition.....	xv
Foreword to the Fourth Edition.....	xvii
Preface.....	xix

Part I. The awk Language

1. Getting Started with awk.....	3
How to Run awk Programs	3
One-Shot Throwaway awk Programs	4
Running awk Without Input Files	4
Running Long Programs	5
Executable awk Programs	6
Comments in awk Programs	7
Shell Quoting Issues	8
Datafiles for the Examples	11
Some Simple Examples	12
An Example with Two Rules	14
A More Complex Example	15
awk Statements Versus Lines	16
Other Features of awk	18
When to Use awk	18
Summary	19
2. Running awk and gawk.....	21
Invoking awk	21
Command-Line Options	22
Other Command-Line Arguments	29
Naming Standard Input	30
The Environment Variables gawk Uses	30

The AWKPATH Environment Variable	31
The AWKLIBPATH Environment Variable	32
Other Environment Variables	32
gawk's Exit Status	34
Including Other Files into Your Program	35
Loading Dynamic Extensions into Your Program	36
Obsolete Options and/or Features	37
Undocumented Options and Features	37
Summary	37
3. Regular Expressions.....	39
How to Use Regular Expressions	39
Escape Sequences	40
Regular Expression Operators	43
Using Bracket Expressions	46
How Much Text Matches?	48
Using Dynamic Regexp	49
gawk-Specific Regexp Operators	50
Case Sensitivity in Matching	52
Summary	54
4. Reading Input Files.....	55
How Input Is Split into Records	55
Record Splitting with Standard awk	55
Record Splitting with gawk	58
Examining Fields	60
Nonconstant Field Numbers	61
Changing the Contents of a Field	62
Specifying How Fields Are Separated	65
Whitespace Normally Separates Fields	66
Using Regular Expressions to Separate Fields	66
Making Each Character a Separate Field	67
Setting FS from the Command Line	68
Making the Full Line Be a Single Field	69
Field-Splitting Summary	70
Reading Fixed-Width Data	71
Defining Fields by Content	73
Multiple-Line Records	75
Explicit Input with getline	77
Using getline with No Arguments	78
Using getline into a Variable	79
Using getline from a File	80

Using getline into a Variable from a File	80
Using getline from a Pipe	81
Using getline into a Variable from a Pipe	82
Using getline from a Coprocess	83
Using getline into a Variable from a Coprocess	83
Points to Remember About getline	83
Summary of getline Variants	85
Reading Input with a Timeout	85
Directories on the Command Line	87
Summary	87
5. Printing Output.....	89
The print Statement	89
print Statement Examples	90
Output Separators	91
Controlling Numeric Output with print	92
Using printf Statements for Fancier Printing	93
Introduction to the printf Statement	93
Format-Control Letters	94
Modifiers for printf Formats	96
Examples Using printf	99
Redirecting Output of print and printf	100
Special Files for Standard Preopened Data Streams	103
Special Filenames in gawk	104
Accessing Other Open Files with gawk	104
Special Files for Network Communications	104
Special Filename Caveats	105
Closing Input and Output Redirections	105
Summary	108
6. Expressions.....	111
Constants, Variables, and Conversions	111
Constant Expressions	111
Using Regular Expression Constants	114
Variables	115
Conversion of Strings and Numbers	117
Operators: Doing Something with Values	119
Arithmetic Operators	119
String Concatenation	121
Assignment Expressions	122
Increment and Decrement Operators	125
Truth Values and Conditions	127

True and False in awk	127
Variable Typing and Comparison Expressions	128
Boolean Expressions	132
Conditional Expressions	134
Function Calls	134
Operator Precedence (How Operators Nest)	136
Where You Are Makes a Difference	138
Summary	139
7. Patterns, Actions, and Variables.....	141
Pattern Elements	141
Regular Expressions as Patterns	142
Expressions as Patterns	142
Specifying Record Ranges with Patterns	143
The BEGIN and END Special Patterns	145
The BEGINFILE and ENDFILE Special Patterns	147
The Empty Pattern	148
Using Shell Variables in Programs	148
Actions	149
Control Statements in Actions	150
The if-else Statement	150
The while Statement	151
The do-while Statement	151
The for Statement	152
The switch Statement	154
The break Statement	155
The continue Statement	156
The next Statement	157
The nextfile Statement	158
The exit Statement	159
Predefined Variables	160
Built-in Variables That Control awk	160
Built-in Variables That Convey Information	163
Using ARGV and ARGV	170
Summary	172
8. Arrays in awk.....	173
The Basics of Arrays	173
Introduction to Arrays	173
Referring to an Array Element	175
Assigning Array Elements	176
Basic Array Example	177

Scanning All Elements of an Array	178
Using Predefined Array Scanning Orders with gawk	179
Using Numbers to Subscript Arrays	182
Using Uninitialized Variables as Subscripts	183
The delete Statement	184
Multidimensional Arrays	185
Scanning Multidimensional Arrays	187
Arrays of Arrays	187
Summary	190
9. Functions.....	191
Built-in Functions	191
Calling Built-in Functions	191
Numeric Functions	192
String-Manipulation Functions	194
Input/Output Functions	208
Time Functions	211
Bit-Manipulation Functions	217
Getting Type Information	219
String-Translation Functions	220
User-Defined Functions	220
Function Definition Syntax	221
Function Definition Examples	223
Calling User-Defined Functions	224
The return Statement	229
Functions and Their Effects on Variable Typing	230
Indirect Function Calls	231
Summary	236

Part II. Problem Solving with awk

10. A Library of awk Functions.....	241
Naming Library Function Global Variables	242
General Programming	243
Converting Strings to Numbers	244
Assertions	245
Rounding Numbers	247
The Cliff Random Number Generator	247
Translating Between Characters and Numbers	248
Merging an Array into a String	250
Managing the Time of Day	250

Reading a Whole File at Once	252
Quoting Strings to Pass to the Shell	253
Datafile Management	254
Noting Datafile Boundaries	254
Rereading the Current File	256
Checking for Readable Datafiles	257
Checking for Zero-Length Files	257
Treating Assignments as Filenames	258
Processing Command-Line Options	259
Reading the User Database	264
Reading the Group Database	269
Traversing Arrays of Arrays	273
Summary	275
11. Practical awk Programs	277
Running the Example Programs	277
Reinventing Wheels for Fun and Profit	278
Cutting Out Fields and Columns	278
Searching for Regular Expressions in Files	283
Printing Out User Information	287
Splitting a Large File into Pieces	289
Duplicating Output into Multiple Files	290
Printing Nonduplicated Lines of Text	292
Counting Things	296
A Grab Bag of awk Programs	299
Finding Duplicated Words in a Document	299
An Alarm Clock Program	300
Transliterating Characters	302
Printing Mailing Labels	304
Generating Word-Usage Counts	306
Removing Duplicates from Unsorted Text	308
Extracting Programs from Texinfo Source Files	309
A Simple Stream Editor	312
An Easy Way to Use Library Functions	314
Finding Anagrams from a Dictionary	321
And Now for Something Completely Different	323
Summary	323

Part III. Moving Beyond Standard awk with gawk

12. Advanced Features of gawk.....	327
Allowing Nondecimal Input Data	327
Controlling Array Traversal and Array Sorting	328
Controlling Array Traversal	328
Sorting Array Values and Indices with gawk	333
Two-Way Communications with Another Process	334
Using gawk for Network Programming	337
Profiling Your awk Programs	338
Summary	342
13. Internationalization with gawk.....	345
Internationalization and Localization	345
GNU gettext	346
Internationalizing awk Programs	348
Translating awk Programs	350
Extracting Marked Strings	350
Rearranging printf Arguments	351
awk Portability Issues	352
A Simple Internationalization Example	353
gawk Can Speak Your Language	355
Summary	355
14. Debugging awk Programs.....	357
Introduction to the gawk Debugger	357
Debugging in General	357
Debugging Concepts	358
awk Debugging	359
Sample gawk Debugging Session	359
How to Start the Debugger	359
Finding the Bug	360
Main Debugger Commands	363
Control of Breakpoints	364
Control of Execution	365
Viewing and Changing Data	367
Working with the Stack	369
Obtaining Information About the Program and the Debugger State	369
Miscellaneous Commands	372
Readline Support	374
Limitations	374
Summary	375

15. Arithmetic and Arbitrary-Precision Arithmetic with gawk.....	377
A General Description of Computer Arithmetic	377
Other Stuff to Know	379
Arbitrary-Precision Arithmetic Features in gawk	381
Floating-Point Arithmetic: Caveat Emptor!	382
Floating-Point Arithmetic Is Not Exact	382
Getting the Accuracy You Need	384
Try a Few Extra Bits of Precision and Rounding	385
Setting the Precision	385
Setting the Rounding Mode	387
Arbitrary-Precision Integer Arithmetic with gawk	388
Standards Versus Existing Practice	389
Summary	391
16. Writing Extensions for gawk.....	393
Introduction	393
Extension Licensing	394
How It Works at a High Level	394
API Description	396
Introduction	397
General-Purpose Data Types	399
Memory Allocation Functions and Convenience Macros	402
Constructor Functions	403
Registration Functions	404
Printing Messages	413
Updating ERRNO	413
Requesting Values	413
Accessing and Updating Parameters	414
Symbol Table Access	414
Array Manipulation	419
API Variables	428
Boilerplate Code	429
How gawk Finds Extensions	431
Example: Some File Functions	431
Using chdir() and stat()	431
C Code for chdir() and stat()	434
Integrating the Extensions	440
The Sample Extensions in the gawk Distribution	442
File-Related Functions	442
Interface to fnmatch()	445
Interface to fork(), wait(), and waitpid()	446
Enabling In-Place File Editing	447

Character and Numeric values: ord() and chr()	448
Reading Directories	448
Reversing Output	449
Two-Way I/O Example	449
Dumping and Restoring an Array	450
Reading an Entire File	450
Extension Time Functions	451
API Tests	452
The gawkextlib Project	452
Summary	453

Part IV. Appendices

A. The Evolution of the awk Language.....	457
B. Installing gawk.....	469
C. GNU General Public License.....	491
Index.....	505