Tektronix®

8560
MULTI-USER SOFTWARE DEVELOPMENT UNIT

AUXILIARY UTILITIES PACKAGE

USERS MANUAL

Tektronix®
COMMITTED TO EXCELLENCE

This Manual supports the following TEKTRONIX products:

8560 Option	Product
4C	8560U03

These modules are compatible with:

TNIX Version 1 (8560)

PLEASE CHECK FOR CHANGE INFORMATION AT THE REAR OF THIS MANUAL.

8560
MULTI-USER SOFTWARE DEVELOPMENT UNIT
AUXILIARY UTILITIES PACKAGE
USERS MANUAL

Tektronix, Inc.
P.O. Box 500
Beaverton, Oregon 97077

070-4270-00
Product Group 61

Serial Number ____________

First Printing FEB 1982

ABOUT WARRANTY AND SUPPORT FOR THIS PRODUCT

This product is provided by Tektronix as Category C software.

NOTE

Licensed Software for which the software support is specified as Category C is furnished without warranty of any kind, and without any representation regarding quality, performance, or suitability.

TEKTRONIX SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Tektronix' liability for damages, if any, whether based upon contract, negligence, strict liability in tort, warranty, or any other basis, shall not exceed the fee paid by the Customer for the Licensed Software.

Category C software is provided on an "as is" basis. Any software services, if available, will be provided at the then current charges.

CONTENTS

Section 1
INTRODUCTION

TABLES

Section 1
INTRODUCTION

ABOUT THIS PRODUCT

The 8560 MUSDU Auxiliary Utilities Package is a set of miscellaneous software utilities. The Auxiliary Utilities Package includes: pattern scanning, "computer-aided instruction" on TNIX, calculators, macro processing, simple graphics, backup media transfer, and various file manipulation utilities.

ABOUT THIS MANUAL

This users manual provides tutorial and reference material for use with the 8560 Auxiliary Utilities Package. The following sections are included:

Installation. Tells you how to install the Auxiliary Utilities Package.

Technical Notes. Describes any limitations or special instructions for the programs, and any changes made to the programs by Tektronix.

BC—An Arbitrary Precision Desk-Calculator Language. Explains the usage of the BC calculator program.

DC—An Interactive Desk-Calculator. Explains the usage of the DC calculator program.

LEARN—Computer-Aided Instruction on UNIX TM. Describes the program for interpreting Computer-Aided Instruction scripts on the TNIX operating system. This package also includes a set of scripts that provide a computerized introduction to the system.

The M4 Macro Processor. Explains the usage of the M4 macro processor.

SED—A Non-Interactive Text Editor. Explains the usage of the SED stream-oriented editor.

AWK—A Pattern Scanning and Processing Language. Describes the usage of the AWK programming language.

SOURCE OF DOCUMENTS

The tutorial and reference documents contained in Sections 4 through 9 of this manual are reprinted by permission of Bell Laboratories.

LIST OF COMMANDS

Table 1-1 contains a list of the commands included in this package, a brief description of the command's function, and a reference to more detailed information about the command.

Table 1-1
8560 Auxiliary Utilities Package Commands

Command	Description	Reference
at	Executes a command at a later time.	8560 MUSDU Reference Manual Section 6.
awk	Pattern scanning and processing language	See section 9 of this manual; also see 8560 MUSDU Reference Manual Section 6.
basename	Strips filename prefixes and suffixes.	8560 MUSDU Reference Manual Section 6.
bc	Arbitrary-precision binary calculator.	See section 4 of this manual; also see 8560 MUSDU Reference Manual Section 6.
cal	Print a calendar.	8560 MUSDU Reference Manual Section 6.
calendar	Maintains a reminder service.	8560 MUSDU Reference Manual Section 6.
crypt	Encode/decode files.	8560 MUSDU Reference Manual Section 6.
dc	Desk calculator.	See section 5 of this manual; also see 8560 MUSDU Reference Manual Section 6.
dd	File conversion utility.	8560 MUSDU Reference Manual Section 6.
diff3	Perform 3-way differential file comparison.	8560 MUSDU Reference Manual Section 6.
enroll	Establish secret mail password.	8560 MUSDU Reference Manual Section 6.
factor	Factor a number.	8560 MUSDU Reference Manual Section 6.
file	Identify a file type.	8560 MUSDU Reference Manual Section 6.
graph	Draw a graph (works with plot).	8560 MUSDU Reference Manual Section 6.
join	Relational data-base operator.	8560 MUSDU Reference Section 6.
learn	Computer-aided instruction.	See section 6 of this manual; also see 8560 MUSDU Reference Manual Section 6.

Table 1-1 (cont)

Command	Description	Reference
m4	Macro processor.	See section 7 of this manual; also see 8560 MUSDU Reference Manual Section 6.
plot	Graphics filter.	8560 MUSDU Reference Manual Section 6.
prep	Prepare text for statistical processing.	8560 MUSDU Reference Manual Section 6.
primes	Generate large primes.	8560 MUSDU Reference Manual Section 6.
pstat	Print system status.	8560 MUSDU Reference Manual Section 6.
quot	Summarize file system ownership.	8560 MUSDU Reference Manual Section 6.
rev	Reverse the character sequence of each line	8560 MUSDU Reference Manual Section 6.
sed	Stream-oriented editor.	See section 8 of this manual; also see 8560 MUSDU Reference Manual Section 6.
spline	Interpolate smooth curve.	8560 MUSDU Reference Manual Section 6.
split	Split a file into pieces.	8560 MUSDU Reference Manual Section 6.
sum	Count and sum blocks in a file.	8560 MUSDU Reference Manual Section 6.
tabs	Set terminal tabs.	8560 MUSDU Reference Manual Section 6.
tar	Save and restore files on magnetic tape or flexible disks.	8560 MUSDU Reference Manual Section 6.
tk	Paginator for TEKTRONIX 4014 terminal	8560 MUSDU Reference Manual Section 6.
tsort	Topological sort.	8560 MUSDU Reference Manual Section 6.
units	Metric/English conversion program.	8560 MUSDU Reference Manual Section 6.
xget	Receive secret mail.	8560 MUSDU Reference Manual Section 6.
xsend	Send secret mail.	8560 MUSDU Reference Manual Section 6.

Section 2
INSTALLATION

TABLES

Section 2
INSTALLATION

INTRODUCTION

This section explains the procedure for installing the 8560 Auxiliary Utilities Package on your 8560 system. The following information is included: an explanation of the format of the installation disks, installation procedures, and a list of the files needed by each of the Auxiliary Utilities Package commands.

INSTALLATION PROCEDURES

The Auxiliary Utilities Package software resides on two flexible disks. The information on the disks consists of executable binary files in **fbr** format. You can load these programs onto your 8560 system disk as a group, or you can install individual programs. To load the whole package, use the 8560 command **install**. The **install** command takes all of the information from a **fbr** format disk and loads it to the system disk. If you want to install a single program from the disk, the command install -f -x program loads the specified file from the **fbr** disk to the system disk.

For each of the Auxiliary Utilities Package programs to execute properly, certain files must be on the system disk. Refer to the "Dependency Files" discussion later in this section for a complete list of these files. In order for these programs to be installed as system commands, they must be loaded while you are logged in as root.

Installing the Auxiliary Utilities Package

The general procedure for installing the Auxiliary Utilities Package is:

1. Log in to the 8560 as root. You must have superuser status to perform the installation.
2. Load the first software installation disk into the disk drive.
3. Enter the following command to install the software:

   # install

4. When the system returns a prompt (#), Remove the first disk and load the second software installation disk into the disk drive.
5. Enter the following command to install the software:

   # install

Installing an Individual Program

The general procedure for installing a particular program from the installation disks is:

1. Log in to the 8560 as root. You must have superuser status to perform the installation.
2. If you know which disk contains the program that you want to load, place that disk in the disk drive.
3. Enter the following command to install the particular program:

   # install -f -x program
4. If the first disk did not contain the desired program, the system will display the following error message:

   fbr: (warning) **program** not in archive.
5. Remove the first disk and place the second disk in the disk drive.
6. Enter the following command to install the particular program:

   # install -f -x program

For example, to install **learn** you would enter:

# install -f -x learn

DEPENDENCY FILES

Table 2-1 lists each command and the files that it needs for execution. These files may be installed separately to rebuild a command.

Table 2-1
Files Required for Auxiliary Utilities Package Commands

Command	Files Required	
at	/bin/at /bin/sh /etc/cron	/bin/pwd /usr/lib/atrun /dev/null
awk	/bin/awk	
basename	/bin/basename	
bc	/bin/bc /usr/lib/lib.b	/bin/dc
cal	/bin/cal	

Table 2-1 (cont)

Command	Files Required	
calendar	/bin/calendar /etc/cron /bin/sed	/usr/lib/calendar /bin/egrep /bin/mail
crypt	/bin/crypt /dev/tty	/user/lib/makekey
dc	/bin/dc	/bin/sh
dd	/bin/dd	
diff3	/bin/diff3 /bin/test /bin/diff	/usr/lib/diff3 /bin/echo /bin/rm
enroll	/bin/enroll /dev/tty /etc/utmp /dev	/usr/lib/makekey /etc/passwd /etc/ttys /usr/spool/secretmail/notice
factor	/bin/factor	
file	/bin/file	
graph	/bin/graph	
join	/bin/join	
learn	/bin/learn /usr/lib/learn/tee /usr/lib/learn/editor /usr/lib/learn/Linfo /usr/lib/learn/play /usr/lib/learn/macros /bin/sh	/usr/lib/learn/lcount /usr/lib/learn/C /usr/lib/learn/eqn /usr/lib/learn/Xinfo /usr/lib/learn/files /usr/lib/learn/morefiles /bin/rm
m4	/bin/m4	
plot	/bin/plot /bin/t300s /bin/tek	/bin/t300 /bin/t450
prep	/bin/prep	/usr/lib/eign
primes	/bin/primes	
pstat	/bin/pstat	/dev/mem
quot	/bin/quot	/etc/passwd
rev	/bin/rev	
sed	/bin/sed	

Table 2-1 (cont)

Command	Files Required	
spline	/bin/spline	
split	/bin/split	
sum	/bin/sum	
tabs	/bin/tabs	
tar	/bin/tar /bin/sort /tmp /bin/mkdir /bin/pwd /bin/sh	
tk	/bin/tk /bin/sh /dev/null /dev/tty	
tsort	/bin/tsort	
units	/bin/units	/usr/lib/units
xget	/bin/xget /dev/tty	/usr/lib/makekey /usr/spool/secretmail/notice
xsend	/bin/xsend	/usr/spool/secretmail/notice

Section 3

TECHNICAL NOTES

This section is reserved for technical information about the 8560 MUSDU Auxiliary Utilities Package. At the time of this writing, no technical notes are included. Technical notes will be incorporated into later versions of this manual, as needed.

Section 4

BC—AN ARBITRARY PRECISION DESK-CALCULATOR LANGUAGE

INTRODUCTION

bc, an arbitrary precision desk-calculator language, was developed at Bell Laboratories and is licensed by Western Electric for use on the 8560. The remainder of this section is a reprint of an article describing *bc*. The Technical Notes section of this manual describes the limitations of this program and any changes made to this program by Tektronix.

BC – An Arbitrary Precision Desk-Calculator Language

Lorinda Cherry

Robert Morris

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

BC is a language and a compiler for doing arbitrary precision arithmetic on the PDP-11 under the UNIX† time-sharing system. The output of the compiler is interpreted and executed by a collection of routines which can input, output, and do arithmetic on indefinitely large integers and on scaled fixed-point numbers.

These routines are themselves based on a dynamic storage allocator. Overflow does not occur until all available core storage is exhausted.

The language has a complete control structure as well as immediate-mode operation. Functions can be defined and saved for later execution.

Two five hundred-digit numbers can be multiplied to give a thousand digit result in about ten seconds.

A small collection of library functions is also available, including sin, cos, arctan, log, exponential, and Bessel functions of integer order.

Some of the uses of this compiler are

– to do computation with large integers,
– to do computation accurate to many decimal places,
– conversion of numbers from one base to another base.

November 12, 1978

†UNIX is a Trademark of Bell Laboratories.

BC – An Arbitrary Precision Desk-Calculator Language

Lorinda Cherry

Robert Morris

Bell Laboratories
Murray Hill, New Jersey 07974

Introduction

BC is a language and a compiler for doing arbitrary precision arithmetic on the UNIX† time-sharing system [1]. The compiler was written to make conveniently available a collection of routines (called DC [5]) which are capable of doing arithmetic on integers of arbitrary size. The compiler is by no means intended to provide a complete programming language. It is a minimal language facility.

There is a scaling provision that permits the use of decimal point notation. Provision is made for input and output in bases other than decimal. Numbers can be converted from decimal to octal by simply setting the output base to equal 8.

The actual limit on the number of digits that can be handled depends on the amount of storage available on the machine. Manipulation of numbers with many hundreds of digits is possible even on the smallest versions of UNIX.

The syntax of BC has been deliberately selected to agree substantially with the C language [2]. Those who are familiar with C will find few surprises in this language.

Simple Computations with Integers

The simplest kind of statement is an arithmetic expression on a line by itself. For instance, if you type in the line:

```
142857 + 285714
```

the program responds immediately with the line

```
428571
```

The operators –, *, /, %, and ^ can also be used; they indicate subtraction, multiplication, division, remaindering, and exponentiation, respectively. Division of integers produces an integer result truncated toward zero. Division by zero produces an error comment.

Any term in an expression may be prefixed by a minus sign to indicate that it is to be negated (the 'unary' minus sign). The expression

```
7+–3
```

is interpreted to mean that –3 is to be added to 7.

More complex expressions with several operators and with parentheses are interpreted just as in Fortran, with ^ having the greatest binding power, then * and % and /, and finally + and –. Contents of parentheses are evaluated before material outside the parentheses. Exponentiations are performed from right to left and the other operators from left to right. The two expressions

†UNIX is a Trademark of Bell Laboratories.

```
        a^b^c  and  a^(b^c)
```

are equivalent, as are the two expressions

```
        a*b*c  and  (a*b)*c
```

BC shares with Fortran and C the undesirable convention that

```
        a/b*c  is equivalent to  (a/b)*c
```

Internal storage registers to hold numbers have single lower-case letter names. The value of an expression can be assigned to a register in the usual way. The statement

```
        x = x + 3
```

has the effect of increasing by three the value of the contents of the register named x. When, as in this case, the outermost operator is an =, the assignment is performed but the result is not printed. Only 26 of these named storage registers are available.

There is a built-in square root function whose result is truncated to an integer (but see scaling below). The lines

```
        x = sqrt(191)
        x
```

produce the printed result

```
        13
```

Bases

There are special internal quantities, called 'ibase' and 'obase'. The contents of 'ibase', initially set to 10, determines the base used for interpreting numbers read in. For example, the lines

```
        ibase = 8
        11
```

will produce the output line

```
        9
```

and you are all set up to do octal to decimal conversions. Beware, however of trying to change the input base back to decimal by typing

```
        ibase = 10
```

Because the number 10 is interpreted as octal, this statement will have no effect. For those who deal in hexadecimal notation, the characters A—F are permitted in numbers (no matter what base is in effect) and are interpreted as digits having values 10—15 respectively. The statement

```
        ibase = A
```

will change you back to decimal input base no matter what the current input base is. Negative and large positive input bases are permitted but useless. No mechanism has been provided for the input of arbitrary numbers in bases less than 1 and greater than 16.

The contents of 'obase', initially set to 10, are used as the base for output numbers. The lines

```
        obase = 16
        1000
```

will produce the output line

```
    3E8
```

which is to be interpreted as a 3-digit hexadecimal number. Very large output bases are permitted, and they are sometimes useful. For example, large numbers can be output in groups of five digits by setting 'obase' to 100000. Strange (i.e. 1, 0, or negative) output bases are handled appropriately.

Very large numbers are split across lines with 70 characters per line. Lines which are continued end with \. Decimal output conversion is practically instantaneous, but output of very large numbers (i.e., more than 100 digits) with other bases is rather slow. Non-decimal output conversion of a one hundred digit number takes about three seconds.

It is best to remember that 'ibase' and 'obase' have no effect whatever on the course of internal computation or on the evaluation of expressions, but only affect input and output conversion, respectively.

Scaling

A third special internal quantity called 'scale' is used to determine the scale of calculated quantities. Numbers may have up to 99 decimal digits after the decimal point. This fractional part is retained in further computations. We refer to the number of digits after the decimal point of a number as its scale.

When two scaled numbers are combined by means of one of the arithmetic operations, the result has a scale determined by the following rules. For addition and subtraction, the scale of the result is the larger of the scales of the two operands. In this case, there is never any truncation of the result. For multiplications, the scale of the result is never less than the maximum of the two scales of the operands, never more than the sum of the scales of the operands and, subject to those two restrictions, the scale of the result is set equal to the contents of the internal quantity 'scale'. The scale of a quotient is the contents of the internal quantity 'scale'. The scale of a remainder is the sum of the scales of the quotient and the divisor. The result of an exponentiation is scaled as if the implied multiplications were performed. An exponent must be an integer. The scale of a square root is set to the maximum of the scale of the argument and the contents of 'scale'.

All of the internal operations are actually carried out in terms of integers, with digits being discarded when necessary. In every case where digits are discarded, truncation and not rounding is performed.

The contents of 'scale' must be no greater than 99 and no less than 0. It is initially set to 0. In case you need more than 99 fraction digits, you may arrange your own scaling.

The internal quantities 'scale', 'ibase', and 'obase' can be used in expressions just like other variables. The line

```
    scale = scale + 1
```

increases the value of 'scale' by one, and the line

```
    scale
```

causes the current value of 'scale' to be printed.

The value of 'scale' retains its meaning as a number of decimal digits to be retained in internal computation even when 'ibase' or 'obase' are not equal to 10. The internal computations (which are still conducted in decimal, regardless of the bases) are performed to the specified number of decimal digits, never hexadecimal or octal or any other kind of digits.

Functions

The name of a function is a single lower-case letter. Function names are permitted to collide with simple variable names. Twenty-six different defined functions are permitted in addition to the twenty-six variable names. The line

```
        define a(x){
```

begins the definition of a function with one argument. This line must be followed by one or more statements, which make up the body of the function, ending with a right brace }. Return of control from a function occurs when a return statement is executed or when the end of the function is reached. The return statement can take either of the two forms

```
    return
    return(x)
```

In the first case, the value of the function is 0, and in the second, the value of the expression in parentheses.

Variables used in the function can be declared as automatic by a statement of the form

```
    auto x,y,z
```

There can be only one 'auto' statement in a function and it must be the first statement in the definition. These automatic variables are allocated space and initialized to zero on entry to the function and thrown away on return. The values of any variables with the same names outside the function are not disturbed. Functions may be called recursively and the automatic variables at each level of call are protected. The parameters named in a function definition are treated in the same way as the automatic variables of that function with the single exception that they are given a value on entry to the function. An example of a function definition is

```
        define a(x,y){
                auto z
                z = x*y
                return(z)
        }
```

The value of this function, when called, will be the product of its two arguments.

A function is called by the appearance of its name followed by a string of arguments enclosed in parentheses and separated by commas. The result is unpredictable if the wrong number of arguments is used.

Functions with no arguments are defined and called using parentheses with nothing between them: b().

If the function *a* above has been defined, then the line

```
    a(7,3.14)
```

would cause the result 21.98 to be printed and the line

```
    x = a(a(3,4),5)
```

would cause the value of x to become 60.

Subscripted Variables

A single lower-case letter variable name followed by an expression in brackets is called a subscripted variable (an array element). The variable name is called the array name and the expression in brackets is called the subscript. Only one-dimensional arrays are permitted. The names of arrays are permitted to collide with the names of simple variables and function names. Any fractional part of a subscript is discarded before use. Subscripts must be greater than or equal to zero and less than or equal to 2047.

Subscripted variables may be freely used in expressions, in function calls, and in return statements.

An array name may be used as an argument to a function, or may be declared as automatic in a function definition by the use of empty brackets:

```
f(a[])
define f(a[])
auto a[]
```

When an array name is so used, the whole contents of the array are copied for the use of the function, and thrown away on exit from the function. Array names which refer to whole arrays cannot be used in any other contexts.

Control Statements

The 'if', the 'while', and the 'for' statements may be used to alter the flow within programs or to cause iteration. The range of each of them is a statement or a compound statement consisting of a collection of statements enclosed in braces. They are written in the following way

```
if(relation) statement
while(relation) statement
for(expression1; relation; expression2) statement
```

or

```
if(relation) {statements}
while(relation) {statements}
for(expression1; relation; expression2) {statements}
```

A relation in one of the control statements is an expression of the form

```
x>y
```

where two expressions are related by one of the six relational operators <, >, <=, >=, ==, or !=. The relation == stands for 'equal to' and != stands for 'not equal to'. The meaning of the remaining relational operators is clear.

BEWARE of using = instead of == in a relational. Unfortunately, both of them are legal, so you will not get a diagnostic message, but = really will not do a comparison.

The 'if' statement causes execution of its range if and only if the relation is true. Then control passes to the next statement in sequence.

The 'while' statement causes execution of its range repeatedly as long as the relation is true. The relation is tested before each execution of its range and if the relation is false, control passes to the next statement beyond the range of the while.

The 'for' statement begins by executing 'expression1'. Then the relation is tested and, if true, the statements in the range of the 'for' are executed. Then 'expression2' is executed. The relation is tested, and so on. The typical use of the 'for' statement is for a controlled iteration, as in the statement

```
for(i=1; i<=10; i=i+1) i
```

which will print the integers from 1 to 10. Here are some examples of the use of the control statements.

```
define f(n){
auto i, x
x=1
for(i=1; i<=n; i=i+1) x=x*i
return(x)
}
```

The line

```
        f(a)
```

will print *a* factorial if *a* is a positive integer. Here is the definition of a function which will compute values of the binomial coefficient (m and n are assumed to be positive integers).

```
define b(n,m){
auto x, j
x=1
for(j=1; j<=m; j=j+1) x=x*(n-j+1)/j
return(x)
}
```

The following function computes values of the exponential function by summing the appropriate series without regard for possible truncation errors:

```
scale = 20
define e(x){
        auto a, b, c, d, n
        a = 1
        b = 1
        c = 1
        d = 0
        n = 1
        while(1==1){
                a = a*x
                b = b*n
                c = c + a/b
                n = n + 1
                if(c==d) return(c)
                d = c
        }
}
```

Some Details

There are some language features that every user should know about even if he will not use them.

Normally statements are typed one to a line. It is also permissible to type several statements on a line separated by semicolons.

If an assignment statement is parenthesized, it then has a value and it can be used anywhere that an expression can. For example, the line

```
(x=y+17)
```

not only makes the indicated assignment, but also prints the resulting value.

Here is an example of a use of the value of an assignment statement even when it is not parenthesized.

```
x = a[i=i+1]
```

causes a value to be assigned to x and also increments i before it is used as a subscript.

The following constructs work in BC in exactly the same manner as they do in the C language. Consult the appendix or the C manuals [2] for their exact workings.

x=y=z is the same as	x=(y=z)
x =+ y	x = x+y
x =− y	x = x−y
x =* y	x = x*y
x =/ y	x = x/y
x =% y	x = x%y
x =^ y	x = x^y
x++	(x=x+1)−1
x−−	(x=x−1)+1
++x	x = x+1
−−x	x = x−1

Even if you don't intend to use the constructs, if you type one inadvertently, something correct but unexpected may happen.

WARNING! In some of these constructions, spaces are significant. There is a real difference between x=−y and x= −y. The first replaces x by x−y and the second by −y.

Three Important Things

1. To exit a BC program, type 'quit'.

2. There is a comment convention identical to that of C and of PL/I. Comments begin with '/*' and end with '*/'.

3. There is a library of math functions which may be obtained by typing at command level

```
bc −l
```

This command will load a set of library functions which, at the time of writing, consists of sine (named 's'), cosine ('c'), arctangent ('a'), natural logarithm ('l'), exponential ('e') and Bessel functions of integer order ('j(n,x)'). Doubtless more functions will be added in time. The library sets the scale to 20. You can reset it to something else if you like. The design of these mathematical library routines is discussed elsewhere [3].

If you type

```
bc file ...
```

BC will read and execute the named file or files before accepting commands from the keyboard. In this way, you may load your favorite programs and function definitions.

Acknowledgement

The compiler is written in YACC [4]; its original version was written by S. C. Johnson.

References

[1] K. Thompson and D. M. Ritchie, *UNIX Programmer's Manual*, Bell Laboratories, 1978.

[2] B. W. Kernighan and D. M. Ritchie, *The C Programming Language*, Prentice-Hall, 1978.

[3] R. Morris, *A Library of Reference Standard Mathematical Subroutines*, Bell Laboratories internal memorandum, 1975.

[4] S. C. Johnson, *YACC — Yet Another Compiler-Compiler*. Bell Laboratories Computing Science Technical Report #32, 1978.

[5] R. Morris and L. L. Cherry, *DC — An Interactive Desk Calculator*.

Appendix

1. Notation

In the following pages syntactic categories are in *italics*; literals are in **bold**; material in brackets [] is optional.

2. Tokens

Tokens consist of keywords, identifiers, constants, operators, and separators. Token separators may be blanks, tabs or comments. Newline characters or semicolons separate statements.

2.1. Comments

Comments are introduced by the characters /* and terminated by */.

2.2. Identifiers

There are three kinds of identifiers — ordinary identifiers, array identifiers and function identifiers. All three types consist of single lower-case letters. Array identifiers are followed by square brackets, possibly enclosing an expression describing a subscript. Arrays are singly dimensioned and may contain up to 2048 elements. Indexing begins at zero so an array may be indexed from 0 to 2047. Subscripts are truncated to integers. Function identifiers are followed by parentheses, possibly enclosing arguments. The three types of identifiers do not conflict; a program can have a variable named **x**, an array named **x** and a function named **x**, all of which are separate and distinct.

2.3. Keywords

The following are reserved keywords:

ibase	**if**
obase	**break**
scale	**define**
sqrt	**auto**
length	**return**
while	**quit**
for	

2.4. Constants

Constants consist of arbitrarily long numbers with an optional decimal point. The hexadecimal digits **A**—**F** are also recognized as digits with values 10—15, respectively.

3. Expressions

The value of an expression is printed unless the main operator is an assignment. Precedence is the same as the order of presentation here, with highest appearing first. Left or right associativity, where applicable, is discussed with each operator.

3.1. Primitive expressions

3.1.1. Named expressions

Named expressions are places where values are stored. Simply stated, named expressions are legal on the left side of an assignment. The value of a named expression is the value stored in the place named.

3.1.1.1. *identifiers*

Simple identifiers are named expressions. They have an initial value of zero.

3.1.1.2. *array-name* [*expression*]

Array elements are named expressions. They have an initial value of zero.

3.1.1.3. scale, ibase and obase

The internal registers **scale**, **ibase** and **obase** are all named expressions. **scale** is the number of digits after the decimal point to be retained in arithmetic operations. **scale** has an initial value of zero. **ibase** and **obase** are the input and output number radix respectively. Both **ibase** and **obase** have initial values of 10.

3.1.2. Function calls

3.1.2.1. *function-name* ([*expression* [,*expression*...]])

A function call consists of a function name followed by parentheses containing a comma-separated list of expressions, which are the function arguments. A whole array passed as an argument is specified by the array name followed by empty square brackets. All function arguments are passed by value. As a result, changes made to the formal parameters have no effect on the actual arguments. If the function terminates by executing a return statement, the value of the function is the value of the expression in the parentheses of the return statement or is zero if no expression is provided or if there is no return statement.

3.1.2.2. sqrt (*expression*)

The result is the square root of the expression. The result is truncated in the least significant decimal place. The scale of the result is the scale of the expression or the value of **scale**, whichever is larger.

3.1.2.3. length (*expression*)

The result is the total number of significant decimal digits in the expression. The scale of the result is zero.

3.1.2.4. scale (*expression*)

The result is the scale of the expression. The scale of the result is zero.

3.1.3. Constants

Constants are primitive expressions.

3.1.4. Parentheses

An expression surrounded by parentheses is a primitive expression. The parentheses are used to alter the normal precedence.

3.2. Unary operators

The unary operators bind right to left.

3.2.1. − *expression*

The result is the negative of the expression.

3.2.2. ++ *named-expression*

The named expression is incremented by one. The result is the value of the named expression after incrementing.

3.2.3. −− *named-expression*

The named expression is decremented by one. The result is the value of the named expression after decrementing.

3.2.4. *named-expression* ++

The named expression is incremented by one. The result is the value of the named expression before incrementing.

3.2.5. *named-expression* −−

The named expression is decremented by one. The result is the value of the named expression before decrementing.

3.3. Exponentiation operator

The exponentiation operator binds right to left.

3.3.1. *expression* ^ *expression*

The result is the first expression raised to the power of the second expression. The second expression must be an integer. If *a* is the scale of the left expression and *b* is the absolute value of the right expression, then the scale of the result is:

min ($a \times b$, max (**scale**, a))

3.4. Multiplicative operators

The operators *, /, % bind left to right.

3.4.1. *expression* * *expression*

The result is the product of the two expressions. If *a* and *b* are the scales of the two expressions, then the scale of the result is:

min ($a+b$, max (**scale**, a, b))

3.4.2. *expression* / *expression*

The result is the quotient of the two expressions. The scale of the result is the value of **scale**.

3.4.3. *expression* % *expression*

The % operator produces the remainder of the division of the two expressions. More precisely, *a%b* is $a-a/b*b$.

The scale of the result is the sum of the scale of the divisor and the value of **scale**

3.5. Additive operators

The additive operators bind left to right.

3.5.1. *expression + expression*

The result is the sum of the two expressions. The scale of the result is the maximun of the scales of the expressions.

3.5.2. *expression − expression*

The result is the difference of the two expressions. The scale of the result is the maximum of the scales of the expressions.

3.6. assignment operators

The assignment operators bind right to left.

3.6.1. *named-expression = expression*

This expression results in assigning the value of the expression on the right to the named expression on the left.

3.6.2. *named-expression =+ expression*

3.6.3. *named-expression =− expression*

3.6.4. *named-expression =* expression*

3.6.5. *named-expression =/ expression*

3.6.6. *named-expression =% expression*

3.6.7. *named-expression =^ expression*

The result of the above expressions is equivalent to "named expression = named expression OP expression", where OP is the operator after the = sign.

4. Relations

Unlike all other operators, the relational operators are only valid as the object of an **if**, **while**, or inside a **for** statement.

4.1. *expression < expression*

4.2. *expression > expression*

4.3. *expression <= expression*

4.4. *expression >= expression*

4.5. *expression == expression*

4.6. *expression != expression*

5. Storage classes

There are only two storage classes in BC, global and automatic (local). Only identifiers that are to be local to a function need be declared with the **auto** command. The arguments to a function are local to the function. All other identifiers are assumed to be global and available to all functions. All identifiers, global and local, have initial values of zero. Identifiers declared as **auto** are allocated on entry to the function and released on returning from the function. They therefore do not retain values between function calls. **auto** arrays are specified by the array name followed by empty square brackets.

Automatic variables in BC do not work in exactly the same way as in either C or PL/I. On entry to a function, the old values of the names that appear as parameters and as automatic variables are pushed onto a stack. Until return is made from the function, reference to these names refers only to the new values.

6. Statements

Statements must be separated by semicolon or newline. Except where altered by control statements, execution is sequential.

6.1. Expression statements

When a statement is an expression, unless the main operator is an assignment, the value of the expression is printed, followed by a newline character.

6.2. Compound statements

Statements may be grouped together and used when one statement is expected by surrounding them with { }.

6.3. Quoted string statements

"any string"

This statement prints the string inside the quotes.

6.4. If statements

if (*relation*) *statement*

The substatement is executed if the relation is true.

6.5. While statements

while (*relation*) *statement*

The statement is executed while the relation is true. The test occurs before each execution of the statement.

6.6. For statements

for (*expression*; *relation*; *expression*) *statement*

The for statement is the same as

```
first-expression
while (relation) {
        statement
        last-expression
}
```

All three expressions must be present.

6.7. Break statements

break

break causes termination of a **for** or **while** statement.

6.8. Auto statements

auto *identifier* [*,identifier*]

The auto statement causes the values of the identifiers to be pushed down. The identifiers can be ordinary identifiers or array identifiers. Array identifiers are specified by following the array name by empty square brackets. The auto statement must be the first statement in a function definition.

6.9. Define statements

define([*parameter* [*,parameter*. . .]]) {
statements }

The define statement defines a function. The parameters may be ordinary identifiers or array names. Array names must be followed by empty square brackets.

6.10. Return statements

return

return(*expression*)

The return statement causes termination of a function, popping of its auto variables, and specifies the result of the function. The first form is equivalent to **return(0)**. The result of the function is the result of the expression in parentheses.

6.11. Quit

The quit statement stops execution of a BC program and returns control to UNIX when it is first encountered. Because it is not treated as an executable statement, it cannot be used in a function definition or in an **if, for,** or **while** statement.

Section 5

DC—AN INTERACTIVE DESK-CALCULATOR

INTRODUCTION

dc, an interactive desk-calculator, was developed at Bell Laboratories and is licensed by Western Electric for use on the 8560. The remainder of this section is a reprint of an article describing *dc*. The Technical Notes section of this manual describes the limitations of this program and any changes made to this program by Tektronix.

DC — An Interactive Desk Calculator

Robert Morris

Lorinda Cherry

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

DC is an interactive desk calculator program implemented on the UNIX† time-sharing system to do arbitrary-precision integer arithmetic. It has provision for manipulating scaled fixed-point numbers and for input and output in bases other than decimal.

The size of numbers that can be manipulated is limited only by available core storage. On typical implementations of UNIX, the size of numbers that can be handled varies from several hundred digits on the smallest systems to several thousand on the largest.

November 15, 1978

†UNIX is a Trademark of Bell Laboratories.

DC — An Interactive Desk Calculator

Robert Morris

Lorinda Cherry

Bell Laboratories
Murray Hill, New Jersey 07974

DC is an arbitrary precision arithmetic package implemented on the UNIX† time-sharing system in the form of an interactive desk calculator. It works like a stacking calculator using reverse Polish notation. Ordinarily DC operates on decimal integers, but one may specify an input base, output base, and a number of fractional digits to be maintained.

A language called BC [1] has been developed which accepts programs written in the familiar style of higher-level programming languages and compiles output which is interpreted by DC. Some of the commands described below were designed for the compiler interface and are not easy for a human user to manipulate.

Numbers that are typed into DC are put on a push-down stack. DC commands work by taking the top number or two off the stack, performing the desired operation, and pushing the result on the stack. If an argument is given, input is taken from that file until its end, then from the standard input.

SYNOPTIC DESCRIPTION

Here we describe the DC commands that are intended for use by people. The additional commands that are intended to be invoked by compiled output are described in the detailed description.

Any number of commands are permitted on a line. Blanks and new-line characters are ignored except within numbers and in places where a register name is expected.

The following constructions are recognized:

number

The value of the number is pushed onto the main stack. A number is an unbroken string of the digits 0-9 and the capital letters A−F which are treated as digits with values 10−15 respectively. The number may be preceded by an underscore to input a negative number. Numbers may contain decimal points.

+ − * % ^

The top two values on the stack are added (+), subtracted (−), multiplied (*), divided (/), remaindered (%), or exponentiated (^). The two entries are popped off the stack; the result is pushed on the stack in their place. The result of a division is an integer truncated toward zero. See the detailed description below for the treatment of numbers with decimal points. An exponent must not have any digits after the decimal point.

†UNIX is a Trademark of Bell Laboratories.

s*x*

The top of the main stack is popped and stored into a register named *x*, where *x* may be any character. If the s is capitalized, *x* is treated as a stack and the value is pushed onto it. Any character, even blank or new-line, is a valid register name.

l*x*

The value in register *x* is pushed onto the stack. The register *x* is not altered. If the **l** is capitalized, register *x* is treated as a stack and its top value is popped onto the main stack.

All registers start with empty value which is treated as a zero by the command **l** and is treated as an error by the command **L**.

d

The top value on the stack is duplicated.

p

The top value on the stack is printed. The top value remains unchanged.

f

All values on the stack and in registers are printed.

x

treats the top element of the stack as a character string, removes it from the stack, and executes it as a string of DC commands.

[...]

puts the bracketed character string onto the top of the stack.

q

exits the program. If executing a string, the recursion level is popped by two. If **q** is capitalized, the top value on the stack is popped and the string execution level is popped by that value.

<*x* >*x* =*x* !<*x* !>*x* !=*x*

The top two elements of the stack are popped and compared. Register *x* is executed if they obey the stated relation. Exclamation point is negation.

v

replaces the top element on the stack by its square root. The square root of an integer is truncated to an integer. For the treatment of numbers with decimal points, see the detailed description below.

!

interprets the rest of the line as a UNIX command. Control returns to DC when the UNIX command terminates.

c

All values on the stack are popped; the stack becomes empty.

i

The top value on the stack is popped and used as the number radix for further input. If **i** is capitalized, the value of the input base is pushed onto the stack. No mechanism has been provided for the input of arbitrary numbers in bases less than 1 or greater than 16.

o

The top value on the stack is popped and used as the number radix for further output. If **o** is capitalized, the value of the output base is pushed onto the stack.

k

The top of the stack is popped, and that value is used as a scale factor that influences the number of decimal places that are maintained during multiplication, division, and exponentiation. The scale factor must be greater than or equal to zero and less than 100. If **k** is capitalized, the value of the scale factor is pushed onto the stack.

z

The value of the stack level is pushed onto the stack.

?

A line of input is taken from the input source (usually the console) and executed.

DETAILED DESCRIPTION

Internal Representation of Numbers

Numbers are stored internally using a dynamic storage allocator. Numbers are kept in the form of a string of digits to the base 100 stored one digit per byte (centennial digits). The string is stored with the low-order digit at the beginning of the string. For example, the representation of 157 is 57,1. After any arithmetic operation on a number, care is taken that all digits are in the range 0—99 and that the number has no leading zeros. The number zero is represented by the empty string.

Negative numbers are represented in the 100's complement notation, which is analogous to two's complement notation for binary numbers. The high order digit of a negative number is always −1 and all other digits are in the range 0—99. The digit preceding the high order −1 digit is never a 99. The representation of −157 is 43,98,−1. We shall call this the canonical form of a number. The advantage of this kind of representation of negative numbers is ease of addition. When addition is performed digit by digit, the result is formally correct. The result need only be modified, if necessary, to put it into canonical form.

Because the largest valid digit is 99 and the byte can hold numbers twice that large, addition can be carried out and the handling of carries done later when that is convenient, as it sometimes is.

An additional byte is stored with each number beyond the high order digit to indicate the number of assumed decimal digits after the decimal point. The representation of .001 is 1,*3* where the scale has been italicized to emphasize the fact that it is not the high order digit. The value of this extra byte is called the **scale factor** of the number.

The Allocator

DC uses a dynamic string storage allocator for all of its internal storage. All reading and writing of numbers internally is done through the allocator. Associated with each string in the allocator is a four-word header containing pointers to the beginning of the string, the end of the string, the next place to write, and the next place to read. Communication between the allocator and DC is done via pointers to these headers.

The allocator initially has one large string on a list of free strings. All headers except the one pointing to this string are on a list of free headers. Requests for strings are made by size. The size of the string actually supplied is the next higher power of 2. When a request for a string is made, the allocator first checks the free list to see if there is a string of the desired size. If none is found, the allocator finds the next larger free string and splits it repeatedly until it has a string of the right size. Left-over strings are put on the free list. If there are no larger strings, the allocator tries to coalesce smaller free strings into larger ones. Since all strings are the result of splitting large strings, each string has a neighbor that is next to it in core and, if free, can be combined with it to make a string twice as long. This is an implementation of the 'buddy system' of allocation described in [2].

Failing to find a string of the proper length after coalescing, the allocator asks the system for more space. The amount of space on the system is the only limitation on the size and number of strings in DC. If at any time in the process of trying to allocate a string, the allocator runs out of headers, it also asks the system for more space.

There are routines in the allocator for reading, writing, copying, rewinding, forward-spacing, and backspacing strings. All string manipulation is done using these routines.

The reading and writing routines increment the read pointer or write pointer so that the characters of a string are read or written in succession by a series of read or write calls. The write pointer is interpreted as the end of the information-containing portion of a string and a call to read beyond that point returns an end-of-string indication. An attempt to write beyond the end of a string causes the allocator to allocate a larger space and then copy the old string into the larger block.

Internal Arithmetic

All arithmetic operations are done on integers. The operands (or operand) needed for the operation are popped from the main stack and their scale factors stripped off. Zeros are added or digits removed as necessary to get a properly scaled result from the internal arithmetic routine. For example, if the scale of the operands is different and decimal alignment is required, as it is for addition, zeros are appended to the operand with the smaller scale. After performing the required arithmetic operation, the proper scale factor is appended to the end of the number before it is pushed on the stack.

A register called **scale** plays a part in the results of most arithmetic operations. **scale** is the bound on the number of decimal places retained in arithmetic computations. **scale** may be set to the number on the top of the stack truncated to an integer with the **k** command. **K** may be used to push the value of **scale** on the stack. **scale** must be greater than or equal to 0 and less than 100. The descriptions of the individual arithmetic operations will include the exact effect of **scale** on the computations.

Addition and Subtraction

The scales of the two numbers are compared and trailing zeros are supplied to the number with the lower scale to give both numbers the same scale. The number with the smaller scale is multiplied by 10 if the difference of the scales is odd. The scale of the result is then set to the larger of the scales of the two operands.

Subtraction is performed by negating the number to be subtracted and proceeding as in addition.

Finally, the addition is performed digit by digit from the low order end of the number. The carries are propagated in the usual way. The resulting number is brought into canonical form, which may require stripping of leading zeros, or for negative numbers replacing the high-order configuration 99,−1 by the digit −1. In any case, digits which are not in the range 0−99 must be brought into that range, propagating any carries or borrows that result.

Multiplication

The scales are removed from the two operands and saved. The operands are both made positive. Then multiplication is performed in a digit by digit manner that exactly mimics the hand method of multiplying. The first number is multiplied by each digit of the second number, beginning with its low order digit. The intermediate products are accumulated into a partial sum which becomes the final product. The product is put into the canonical form and its sign is computed from the signs of the original operands.

The scale of the result is set equal to the sum of the scales of the two operands. If that scale is larger than the internal register **scale** and also larger than both of the scales of the two operands, then the scale of the result is set equal to the largest of these three last quantities.

Division

The scales are removed from the two operands. Zeros are appended or digits removed from the dividend to make the scale of the result of the integer division equal to the internal quantity **scale**. The signs are removed and saved.

Division is performed much as it would be done by hand. The difference of the lengths of the two numbers is computed. If the divisor is longer than the dividend, zero is returned. Otherwise the top digit of the divisor is divided into the top two digits of the dividend. The result is used as the first (high-order) digit of the quotient. It may turn out be one unit too low, but if it is, the next trial quotient will be larger than 99 and this will be adjusted at the end of the process. The trial digit is multiplied by the divisor and the result subtracted from the dividend and the process is repeated to get additional quotient digits until the remaining dividend is smaller than the divisor. At the end, the digits of the quotient are put into the canonical form, with propagation of carry as needed. The sign is set from the sign of the operands.

Remainder

The division routine is called and division is performed exactly as described. The quantity returned is the remains of the dividend at the end of the divide process. Since division truncates toward zero, remainders have the same sign as the dividend. The scale of the remainder is set to the maximum of the scale of the dividend and the scale of the quotient plus the scale of the divisor.

Square Root

The scale is stripped from the operand. Zeros are added if necessary to make the integer result have a scale that is the larger of the internal quantity **scale** and the scale of the operand.

The method used to compute sqrt(y) is Newton's method with successive approximations by the rule

$$x_{n+1} = \tfrac{1}{2}(x_n + \frac{y}{x_n})$$

The initial guess is found by taking the integer square root of the top two digits.

Exponentiation

Only exponents with zero scale factor are handled. If the exponent is zero, then the result is 1. If the exponent is negative, then it is made positive and the base is divided into one. The scale of the base is removed.

The integer exponent is viewed as a binary number. The base is repeatedly squared and the result is obtained as a product of those powers of the base that correspond to the positions of the one-bits in the binary representation of the exponent. Enough digits of the result are removed to make the scale of the result the same as if the indicated multiplication had been performed.

Input Conversion and Base

Numbers are converted to the internal representation as they are read in. The scale stored with a number is simply the number of fractional digits input. Negative numbers are indicated by preceding the number with a _. The hexadecimal digits A—F correspond to the numbers 10—15 regardless of input base. The **i** command can be used to change the base of the input numbers. This command pops the stack, truncates the resulting number to an integer, and uses it as the input base for all further input. The input base is initialized to 10 but may, for example be changed to 8 or 16 to do octal or hexadecimal to decimal conversions. The command **I** will push the value of the input base on the stack.

Output Commands

The command **p** causes the top of the stack to be printed. It does not remove the top of the stack. All of the stack and internal registers can be output by typing the command **f**. The **o** command can be used to change the output base. This command uses the top of the stack, truncated to an integer as the base for all further output. The output base in initialized to 10. It will work correctly for any base. The command **O** pushes the value of the output base on the stack.

Output Format and Base

The input and output bases only affect the interpretation of numbers on input and output; they have no effect on arithmetic computations. Large numbers are output with 70 characters per line; a \ indicates a continued line. All choices of input and output bases work correctly, although not all are useful. A particularly useful output base is 100000, which has the effect of grouping digits in fives. Bases of 8 and 16 can be used for decimal-octal or decimal-hexadecimal conversions.

Internal Registers

Numbers or strings may be stored in internal registers or loaded on the stack from registers with the commands **s** and **l**. The command **s***x* pops the top of the stack and stores the result in register **x**. *x* can be any character. **l***x* puts the contents of register **x** on the top of the stack. The **l** command has no effect on the contents of register *x*. The **s** command, however, is destructive.

Stack Commands

The command **c** clears the stack. The command **d** pushes a duplicate of the number on the top of the stack on the stack. The command **z** pushes the stack size on the stack. The command **X** replaces the number on the top of the stack with its scale factor. The command **Z** replaces the top of the stack with its length.

Subroutine Definitions and Calls

Enclosing a string in **[]** pushes the ascii string on the stack. The **q** command quits or in executing a string, pops the recursion levels by two.

Internal Registers — Programming DC

The load and store commands together with **[]** to store strings, **x** to execute and the testing commands '<', '>', '=', '!<', '!>', '!=' can be used to program DC. The **x** command assumes the top of the stack is an string of DC commands and executes it. The testing commands compare the top two elements on the stack and if the relation holds, execute the register that follows the relation. For example, to print the numbers 0-9,

```
[lipl+ si li10>a]sa
0si lax
```

Push-Down Registers and Arrays

These commands were designed for used by a compiler, not by people. They involve push-down registers and arrays. In addition to the stack that commands work on, DC can be thought of as having individual stacks for each register. These registers are operated on by the commands S and L. S*x* pushes the top value of the main stack onto the stack for the register *x*. L*x* pops the stack for register *x* and puts the result on the main stack. The commands s and l also work on registers but not as push-down stacks. l doesn't effect the top of the register stack, and s destroys what was there before.

The commands to work on arrays are : and ;. :*x* pops the stack and uses this value as an index into the array *x*. The next element on the stack is stored at this index in *x*. An index must be greater than or equal to 0 and less than 2048. ;*x* is the command to load the main stack from the array *x*. The value on the top of the stack is the index into the array *x* of the value to be loaded.

Miscellaneous Commands

The command ! interprets the rest of the line as a UNIX command and passes it to UNIX to execute. One other compiler command is **Q**. This command uses the top of the stack as the number of levels of recursion to skip.

DESIGN CHOICES

The real reason for the use of a dynamic storage allocator was that a general purpose program could be (and in fact has been) used for a variety of other tasks. The allocator has some value for input and for compiling (i.e. the bracket [...] commands) where it cannot be known in advance how long a string will be. The result was that at a modest cost in execution time, all considerations of string allocation and sizes of strings were removed from the remainder of the program and debugging was made easier. The allocation method used wastes approximately 25% of available space.

The choice of 100 as a base for internal arithmetic seemingly has no compelling advantage. Yet the base cannot exceed 127 because of hardware limitations and at the cost of 5% in space, debugging was made a great deal easier and decimal output was made much faster.

The reason for a stack-type arithmetic design was to permit all DC commands from addition to subroutine execution to be implemented in essentially the same way. The result was a considerable degree of logical separation of the final program into modules with very little communication between modules.

The rationale for the lack of interaction between the scale and the bases was to provide an understandable means of proceeding after a change of base or scale when numbers had already been entered. An earlier implementation which had global notions of scale and base did not work out well. If the value of **scale** were to be interpreted in the current input or output base, then a change of base or scale in the midst of a computation would cause great confusion in the interpretation of the results. The current scheme has the advantage that the value of the input and output bases are only used for input and output, respectively, and they are ignored in all other operations. The value of scale is not used for any essential purpose by any part of the program and it is used only to prevent the number of decimal places resulting from the arithmetic operations from growing beyond all bounds.

The design rationale for the choices for the scales of the results of arithmetic were that in no case should any significant digits be thrown away if, on appearances, the user actually wanted them. Thus, if the user wants to add the numbers 1.5 and 3.517, it seemed reasonable to give him the result 5.017 without requiring him to unnecessarily specify his rather obvious requirements for precision.

On the other hand, multiplication and exponentiation produce results with many more digits than their operands and it seemed reasonable to give as a minimum the number of decimal places in the operands but not to give more than that number of digits unless the user asked for them by specifying a value for **scale**. Square root can be handled in just the same way as multiplication. The operation of division gives arbitrarily many decimal places and there is simply no way to guess how many places the user wants. In this case only, the user must specify a **scale** to get any decimal places at all.

The scale of remainder was chosen to make it possible to recreate the dividend from the quotient and remainder. This is easy to implement; no digits are thrown away.

References

[1] L. L. Cherry, R. Morris, *BC — An Arbitrary Precision Desk-Calculator Language.*

[2] K. C. Knowlton, *A Fast Storage Allocator,* Comm. ACM **8**, pp. 623-625 (Oct. 1965).

Section 6

LEARN—COMPUTER-AIDED INSTRUCTION ON UNIX™

INTRODUCTION

learn, a program for computer-aided instruction on UNIX, was developed at Bell Laboratories and is licensed by Western Electric for use on the 8560. The remainder of this section is a reprint of an article describing *learn*. The Technical Notes section of this manual describes the limitations of this program and any changes made to this program by Tektronix.

™UNIX is a Trademark of Bell Laboratories.

LEARN — Computer-Aided Instruction on UNIX (Second Edition)

Brian W. Kernighan

Michael E. Lesk

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

This paper describes the second version of the *learn* program for interpreting CAI scripts on the UNIX† operating system, and a set of scripts that provide a computerized introduction to the system.

Six current scripts cover basic commands and file handling, the editor, additional file handling commands, the *eqn* program for mathematical typing, the "−ms" package of formatting macros, and an introduction to the C programming language. These scripts now include a total of about 530 lessons.

Many users from a wide variety of backgrounds have used *learn* to acquire basic UNIX skills. Most usage involves the first two scripts, an introduction to files and commands, and the text editor.

The second version of *learn* is about four times faster than the previous one in CPU utilization, and much faster in perceived time because of better overlap of computing and printing. It also requires less file space than the first version. Many of the lessons have been revised; new material has been added to reflect changes and enhancements in the UNIX system itself. Script-writing is also easier because of revisions to the script language.

January 30, 1979

†UNIX is a Trademark of Bell Laboratories.

LEARN — Computer-Aided Instruction on UNIX (Second Edition)

Brian W. Kernighan

Michael E. Lesk

Bell Laboratories
Murray Hill, New Jersey 07974

1. Introduction.

Learn is a driver for CAI scripts. It is intended to permit the easy composition of lessons and lesson fragments to teach people computer skills. Since it is teaching the same system on which it is implemented, it makes direct use of UNIX† facilities to create a controlled UNIX environment. The system includes two main parts: (1) a driver that interprets the lesson scripts; and (2) the lesson scripts themselves. At present there are six scripts:

— basic file handling commands

— the UNIX text editor *ed*

— advanced file handling

— the *eqn* language for typing mathematics

— the "−ms" macro package for document formatting

— the C programming language

The purported advantages of CAI scripts for training in computer skills include the following:

(a) students are forced to perform the exercises that are in fact the basis of training in any case;

(b) students receive immediate feedback and confirmation of progress;

(c) students may progress at their own rate;

(d) no schedule requirements are imposed; students may study at any time convenient for them;

(e) the lessons may be improved individually and the improvements are immediately available to new users;

(f) since the student has access to a computer for the CAI script there is a place to do exercises;

(g) the use of high technology will improve student motivation and the interest of their management.

Opposed to this, of course, is the absence of anyone to whom the student may direct questions. If CAI is used without a "counselor" or other assistance, it should properly be compared to a textbook, lecture series, or taped course, rather than to a seminar. CAI has been used for many years in a variety of educational areas.[1,2,3] The use of a computer to teach itself, however, offers unique advantages. The skills developed to get through the script are exactly those needed to use the computer; there is no waste effort.

The scripts written so far are based on some familiar assumptions about education; these

†UNIX is a Trademark of Bell Laboratories.

assumptions are outlined in the next section. The remaining sections describe the operation of the script driver and the particular scripts now available. The driver puts few restrictions on the script writer, but the current scripts are of a rather rigid and stereotyped form in accordance with the theory in the next section and practical limitations.

2. Educational Assumptions and Design.

First, the way to teach people how to do something is to have them do it. Scripts should not contain long pieces of explanation; they should instead frequently ask the student to do some task. So teaching is always by example: the typical script fragment shows a small example of some technique and then asks the user to either repeat that example or produce a variation on it. All are intended to be easy enough that most students will get most questions right, reinforcing the desired behavior.

Most lessons fall into one of three types. The simplest presents a lesson and asks for a yes or no answer to a question. The student is given a chance to experiment before replying. The script checks for the correct reply. Problems of this form are sparingly used.

The second type asks for a word or number as an answer. For example a lesson on files might say

How many files are there in the current directory? Type "answer N", where N is the number of files.

The student is expected to respond (perhaps after experimenting) with

answer 17

or whatever. Surprisingly often, however, the idea of a substitutable argument (i.e., replacing *N* by 17) is difficult for non-programmer students, so the first few such lessons need real care.

The third type of lesson is open-ended — a task is set for the student, appropriate parts of the input or output are monitored, and the student types *ready* when the task is done. Figure 1 shows a sample dialog that illustrates the last of these, using two lessons about the *cat* (concatenate, i.e., print) command taken from early in the script that teaches file handling. Most *learn* lessons are of this form.

After each correct response the computer congratulates the student and indicates the lesson number that has just been completed, permitting the student to restart the script after that lesson. If the answer is wrong, the student is offered a chance to repeat the lesson. The "speed" rating of the student (explained in section 5) is given after the lesson number when the lesson is completed successfully; it is printed only for the aid of script authors checking out possible errors in the lessons.

It is assumed that there is no foolproof way to determine if the student truly "understands" what he or she is doing; accordingly, the current *learn* scripts only measure performance, not comprehension. If the student can perform a given task, that is deemed to be "learning."[4]

The main point of using the computer is that what the student does is checked for correctness immediately. Unlike many CAI scripts, however, these scripts provide few facilities for dealing with wrong answers. In practice, if most of the answers are not right the script is a failure; the universal solution to student error is to provide a new, easier script. Anticipating possible wrong answers is an endless job, and it is really easier as well as better to provide a simpler script.

Along with this goes the assumption that anything can be taught to anybody if it can be broken into sufficiently small pieces. Anything not absorbed in a single chunk is just subdivided.

To avoid boring the faster students, however, an effort is made in the files and editor scripts to provide three tracks of different difficulty. The fastest sequence of lessons is aimed at roughly the bulk and speed of a typical tutorial manual and should be adequate for review and for well-prepared students. The next track is intended for most users and is roughly twice as

Figure 1: Sample dialog from basic files script

(Student responses in italics; '$' is the prompt)

```
A file can be printed on your terminal
by using the "cat" command. Just say
"cat file" where "file" is the file name.
For example, there is a file named
"food" in this directory. List it
by saying "cat food"; then type "ready".
$ cat food
  this is the file
  named food.
$ ready

Good. Lesson 3.3a (1)

Of course, you can print any file with "cat".
In particular, it is common to first use
"ls" to find the name of a file and then "cat"
to print it. Note the difference between
"ls", which tells you the name of the file,
and "cat", which tells you the contents.
One file in the current directory is named for
a President. Print the file, then type "ready".
$ cat President
cat: can't open President
$ ready

Sorry, that's not right. Do you want to try again? yes
Try the problem again.
$ ls
.ocopy
X1
roosevelt
$ cat roosevelt
  this file is named roosevelt
  and contains three lines of
  text.
$ ready

Good. Lesson 3.3b (0)

The "cat" command can also print several files
at once. In fact, it is named "cat" as an abbreviation
for "concatenate"....
```

long. Typically, for example, the fast track might present an idea and ask for a variation on the example shown; the normal track will first ask the student to repeat the example that was shown before attempting a variation. The third and slowest track, which is often three or four times the length of the fast track, is intended to be adequate for anyone. (The lessons of Figure 1 are from the third track.) The multiple tracks also mean that a student repeating a course is unlikely to hit the same series of lessons; this makes it profitable for a shaky user to back up and try again, and many students have done so.

The tracks are not completely distinct, however. Depending on the number of correct answers the student has given for the last few lessons, the program may switch tracks. The driver is actually capable of following an arbitrary directed graph of lesson sequences, as discussed in section 5. Some more structured arrangement, however, is used in all current scripts to aid the script writer in organizing the material into lessons. It is sufficiently difficult to write lessons that the three-track theory is not followed very closely except in the files and editor scripts. Accordingly, in some cases, the fast track is produced merely by skipping lessons from the slower track. In others, there is essentially only one track.

The main reason for using the *learn* program rather than simply writing the same material as a workbook is not the selection of tracks, but actual hands-on experience. Learning by doing is much more effective than pencil and paper exercises.

Learn also provides a mechanical check on performance. The first version in fact would not let the student proceed unless it received correct answers to the questions it set and it would not tell a student the right answer. This somewhat Draconian approach has been moderated in version 2. Lessons are sometimes badly worded or even just plain wrong; in such cases, the student has no recourse. But if a student is simply unable to complete one lesson, that should not prevent access to the rest. Accordingly, the current version of *learn* allows the student to skip a lesson that he cannot pass; a "no" answer to the "Do you want to try again?" question in Figure 1 will pass to the next lesson. It is still true that *learn* will not tell the student the right answer.

Of course, there are valid objections to the assumptions above. In particular, some students may object to not understanding what they are doing; and the procedure of smashing everything into small pieces may provoke the retort "you can't cross a ditch in two jumps." Since writing CAI scripts is considerably more tedious than ordinary manuals, however, it is safe to assume that there will always be alternatives to the scripts as a way of learning. In fact, for a reference manual of 3 or 4 pages it would not be surprising to have a tutorial manual of 20 pages and a (multi-track) script of 100 pages. Thus the reference manual will exist long before the scripts.

3. Scripts.

As mentioned above, the present scripts try at most to follow a three-track theory. Thus little of the potential complexity of the possible directed graph is employed, since care must be taken in lesson construction to see that every necessary fact is presented in every possible path through the units. In addition, it is desirable that every unit have alternate successors to deal with student errors.

In most existing courses, the first few lessons are devoted to checking prerequisites. For example, before the student is allowed to proceed through the editor script the script verifies that the student understands files and is able to type. It is felt that the sooner lack of student preparation is detected, the easier it will be on the student. Anyone proceeding through the scripts should be getting mostly correct answers; otherwise, the system will be unsatisfactory both because the wrong habits are being learned and because the scripts make little effort to deal with wrong answers. Unprepared students should not be encouraged to continue with scripts.

There are some preliminary items which the student must know before any scripts can be tried. In particular, the student must know how to connect to a UNIX system, set the terminal properly, log in, and execute simple commands (e.g., *learn* itself). In addition, the character erase and line kill conventions (# and @) should be known. It is hard to see how this much could be taught by computer-aided instruction, since a student who does not know these basic skills will not be able to run the learning program. A brief description on paper is provided (see Appendix A), although assistance will be needed for the first few minutes. This assistance, however, need not be highly skilled.

The first script in the current set deals with files. It assumes the basic knowledge above and teaches the student about the *ls*, *cat*, *mv*, *rm*, *cp* and *diff* commands. It also deals with the abbreviation characters *, ?, and [ ] in file names. It does not cover pipes or I/O redirection, nor does it present the many options on the *ls* command.

This script contains 31 lessons in the fast track; two are intended as prerequisite checks, seven are review exercises. There are a total of 75 lessons in all three tracks, and the instructional passages typed at the student to begin each lesson total 4,476 words. The average lesson thus begins with a 60-word message. In general, the fast track lessons have somewhat longer introductions, and the slow tracks somewhat shorter ones. The longest message is 144 words and the shortest 14.

The second script trains students in the use of the context editor *ed*, a sophisticated editor using regular expressions for searching.[5] All editor features except encryption, mark names and ';' in addressing are covered. The fast track contains 2 prerequisite checks, 93 lessons, and a review lesson. It is supplemented by 146 additional lessons in other tracks.

A comparison of sizes may be of interest. The *ed* description in the reference manual is 2,572 words long. The *ed* tutorial[6] is 6,138 words long. The fast track through the *ed* script is 7,407 words of explanatory messages, and the total *ed* script, 242 lessons, has 15,615 words. The average *ed* lesson is thus also about 60 words; the largest is 171 words and the smallest 10. The original *ed* script represents about three man-weeks of effort.

The advanced file handling script deals with *ls* options, I/O diversion, pipes, and supporting programs like *pr*, *wc*, *tail*, *spell* and *grep*. (The basic file handling script is a prerequisite.) It is not as refined as the first two scripts; this is reflected at least partly in the fact that it provides much less of a full three-track sequence than they do. On the other hand, since it is perceived as "advanced," it is hoped that the student will have somewhat more sophistication and be better able to cope with it at a reasonably high level of performance.

A fourth script covers the *eqn* language for typing mathematics. This script must be run on a terminal capable of printing mathematics, for instance the DASI 300 and similar Diablo-based terminals, or the nearly extinct Model 37 teletype. Again, this script is relatively short of tracks: of 76 lessons, only 17 are in the second track and 2 in the third track. Most of these provide additional practice for students who are having trouble in the first track.

The *−ms* script for formatting macros is a short one-track only script. The macro package it describes is no longer the standard, so this script will undoubtedly be superseded in the future. Furthermore, the linear style of a single learn script is somewhat inappropriate for the macros, since the macro package is composed of many independent features, and few users need all of them. It would be better to have a selection of short lesson sequences dealing with the features independently.

The script on C is in a state of transition. It was originally designed to follow a tutorial on C, but that document has since become obsolete. The current script has been partially converted to follow the order of presentation in *The C Programming Language*,[7] but this job is not complete. The C script was never intended to teach C; rather it is supposed to be a series of exercises for which the computer provides checking and (upon success) a suggested solution.

This combination of scripts covers much of the material which any user will need to know to make effective use of the UNIX system. With enlargement of the advanced files course to include more on the command interpreter, there will be a relatively complete introduction to UNIX available via *learn*. Although we make no pretense that *learn* will replace other instructional materials, it should provide a useful supplement to existing tutorials and reference manuals.

4. Experience with Students.

Learn has been installed on many different UNIX systems. Most of the usage is on the first two scripts, so these are more thoroughly debugged and polished. As a (random) sample of user experience, the *learn* program has been used at Bell Labs at Indian Hill for 10,500 lessons in a four month period. About 3600 of these are in the files script, 4100 in the editor, and 1400 in advanced files. The passing rate is about 80%, that is, about 4 lessons are passed for every one failed. There have been 86 distinct users of the files script, and 58 of the editor. On our system at Murray Hill, there have been nearly 4000 lessons over four weeks that include Christmas and New Year. Users have ranged in age from six up.

It is difficult to characterize typical sessions with the scripts; many instances exist of someone doing one or two lessons and then logging out, as do instances of someone pausing in a script for twenty minutes or more. In the earlier version of *learn*, the average session in the files course took 32 minutes and covered 23 lessons. The distribution is quite broad and skewed, however; the longest session was 130 minutes and there were five sessions shorter than five minutes. The average lesson took about 80 seconds. These numbers are roughly typical for non-programmers; a UNIX expert can do the scripts at approximately 30 seconds per lesson, most of which is the system printing.

At present working through a section of the middle of the files script took about 1.4 seconds of processor time per lesson, and a system expert typing quickly took 15 seconds of real time per lesson. A novice would probably take at least a minute. Thus, as a rough approximation, a UNIX system could support ten students working simultaneously with some spare capacity.

5. The Script Interpreter.

The *learn* program itself merely interprets scripts. It provides facilities for the script writer to capture student responses and their effects, and simplifies the job of passing control to and recovering control from the student. This section describes the operation and usage of the driver program, and indicates what is required to produce a new script. Readers only interested in the existing scripts may skip this section.

The file structure used by *learn* is shown in Figure 2. There is one parent directory (named *lib*) containing the script data. Within this directory are subdirectories, one for each subject in which a course is available, one for logging (named *log*), and one in which user subdirectories are created (named *play*). The subject directory contains master copies of all lessons, plus any supporting material for that subject. In a given subdirectory, each lesson is a single text file. Lessons are usually named systematically; the file that contains lesson *n* is called *Ln*.

When *learn* is executed, it makes a private directory for the user to work in, within the *learn* portion of the file system. A fresh copy of all the files used in each lesson (mostly data for the student to operate upon) is made each time a student starts a lesson, so the script writer may assume that everything is reinitialized each time a lesson is entered. The student directory is deleted after each session; any permanent records must be kept elsewhere.

The script writer must provide certain basic items in each lesson:

(1) the text of the lesson;

(2) the set-up commands to be executed before the user gets control;

(3) the data, if any, which the user is supposed to edit, transform, or otherwise process;

(4) the evaluating commands to be executed after the user has finished the lesson, to decide whether the answer is right; and

(5) a list of possible successor lessons.

Learn tries to minimize the work of bookkeeping and installation, so that most of the effort involved in script production is in planning lessons, writing tutorial paragraphs, and coding tests of student performance.

Figure 2: Directory structure for *learn*

```
lib
     play
                    student1
                                files for student1...
                    student2
                                files for student2...
     files
                    L0.1a       lessons for files course
                    L0.1b
                    ...
     editor
                    ...
     (other courses)
     log
```

The basic sequence of events is as follows. First, *learn* creates the working directory. Then, for each lesson, *learn* reads the script for the lesson and processes it a line at a time. The lines in the script are: (1) commands to the script interpreter to print something, to create a files, to test something, etc.; (2) text to be printed or put in a file; (3) other lines, which are sent to the shell to be executed. One line in each lesson turns control over to the user; the user can run any UNIX commands. The user mode terminates when the user types *yes*, *no*, *ready*, or *answer*. At this point, the user's work is tested; if the lesson is passed, a new lesson is selected, and if not the old one is repeated.

Let us illustrate this with the script for the second lesson of Figure 1; this is shown in Figure 3.

Lines which begin with # are commands to the *learn* script interpreter. For example,

#print

causes printing of any text that follows, up to the next line that begins with a sharp.

#print file

prints the contents of *file*; it is the same as *cat file* but has less overhead. Both forms of *#print* have the added property that if a lesson is failed, the *#print* will not be executed the second time through; this avoids annoying the student by repeating the preamble to a lesson.

#create filename

creates a file of the specified name, and copies any subsequent text up to a # to the file. This is used for creating and initializing working files and reference data for the lessons.

#user

gives control to the student; each line he or she types is passed to the shell for execution. The *#user* mode is terminated when the student types one of *yes*, *no*, *ready* or *answer*. At that time, the driver resumes interpretation of the script.

#copyin
#uncopyin

Anything the student types between these commands is copied onto a file called *.copy*. This lets the script writer interrogate the student's responses upon regaining control.

Figure 3: Sample Lesson

```
#print
Of course, you can print any file with "cat".
In particular, it is common to first use
"ls" to find the name of a file and then "cat"
to print it. Note the difference between
"ls", which tells you the name of the files,
and "cat", which tells you the contents.
One file in the current directory is named for
a President. Print the file, then type "ready".
#create roosevelt
  this file is named roosevelt
  and contains three lines of
  text.
#copyout
#user
#uncopyout
tail −3 .ocopy >X1
#cmp X1 roosevelt
#log
#next
3.2b 2
```

#copyout
#uncopyout

Between these commands, any material typed at the student by any program is copied to the file *.ocopy*. This lets the script writer interrogate the effect of what the student typed, which true believers in the performance theory of learning usually prefer to the student's actual input.

#pipe
#unpipe

Normally the student input and the script commands are fed to the UNIX command interpreter (the "shell") one line at a time. This won't do if, for example, a sequence of editor commands is provided, since the input to the editor must be handed to the editor, not to the shell. Accordingly, the material between *#pipe* and *#unpipe* commands is fed continuously through a pipe so that such sequences work. If *copyout* is also desired the *copyout* brackets must include the *pipe* brackets.

There are several commands for setting status after the student has attempted the lesson.

#cmp file1 file2

is an in-line implementation of *cmp*, which compares two files for identity.

#match stuff

The last line of the student's input is compared to *stuff*, and the success or fail status is set according to it. Extraneous things like the word *answer* are stripped before the comparison is made. There may be several *#match* lines; this provides a convenient mechanism for handling multiple "right" answers. Any text up to a # on subsequent lines after a successful *#match* is printed; this is illustrated in Figure 4, another sample lesson.

#bad stuff

This is similar to *#match*, except that it corresponds to specific failure answers; this can be used to produce hints for particular wrong answers that have been anticipated by the script

Figure 4: Another Sample Lesson

```
#print
What command will move the current line
to the end of the file? Type
"answer COMMAND", where COMMAND is the command.
#copyin
#user
#uncopyin
#match m$
#match .m$
"m$" is easier.
#log
#next
63.1d 10
```

writer.

#succeed
#fail

print a message upon success or failure (as determined by some previous mechanism).

When the student types one of the "commands" *yes*, *no*, *ready*, or *answer*, the driver terminates the *#user* command, and evaluation of the student's work can begin. This can be done either by the built-in commands above, such as *#match* and *#cmp*, or by status returned by normal UNIX commands, typically *grep* and *test*. The last command should return status true (0) if the task was done successfully and false (non-zero) otherwise; this status return tells the driver whether or not the student has successfully passed the lesson.

Performance can be logged:

#log file

writes the date, lesson, user name and speed rating, and a success/failure indication on *file*. The command

#log

by itself writes the logging information in the logging directory within the *learn* hierarchy, and is the normal form.

#next

is followed by a few lines, each with a successor lesson name and an optional speed rating on it. A typical set might read

```
25.1a  10
25.2a  5
25.3a  2
```

indicating that unit 25.1a is a suitable follow-on lesson for students with a speed rating of 10 units, 25.2a for student with speed near 5, and 25.3a for speed near 2. Speed ratings are maintained for each session with a student; the rating is increased by one each time the student gets a lesson right and decreased by four each time the student gets a lesson wrong. Thus the driver tries to maintain a level such that the users get 80% right answers. The maximum rating is limited to 10 and the minimum to 0. The initial rating is zero unless the student specifies a different rating when starting a session.

If the student passes a lesson, a new lesson is selected and the process repeats. If the student fails, a false status is returned and the program reverts to the previous lesson and tries another alternative. If it can not find another alternative, it skips forward a lesson. The student can terminate a session at any time by typing *bye*, which causes a graceful exit from *learn*. Hanging up is the usual novice's way out.

The lessons may form an arbitrary directed graph, although the present program imposes a limitation on cycles in that it will not present a lesson twice in the same session. If the student is unable to answer one of the exercises correctly, the driver searches for a previous lesson with a set of alternatives as successors (following the *#next* line). From the previous lesson with alternatives one route was taken earlier; the program simply tries a different one.

It is perfectly possible to write sophisticated scripts that evaluate the student's speed of response, or try to estimate the elegance of the answer, or provide detailed analysis of wrong answers. Lesson writing is so tedious already, however, that most of these abilities are likely to go unused.

The driver program depends heavily on features of the UNIX system that are not available on many other operating systems. These include the ease of manipulating files and directories, file redirection, the ability to use the command interpreter as just another program (even in a pipeline), command status testing and branching, the ability to catch signals like interrupts, and of course the pipeline mechanism itself. Although some parts of *learn* might be transferable to other systems, some generality will probably be lost.

A bit of history: The first version of *learn* had fewer built-in commands in the driver program, and made more use of the facilities of the UNIX system itself. For example, file comparison was done by creating a *cmp* process, rather than comparing the two files within *learn*. Lessons were not stored as text files, but as archives. There was no concept of the in-line document; even *#print* had to be followed by a file name. Thus the initialization for each lesson was to extract the archive into the working directory (typically 4-8 files), then *#print* the lesson text.

The combination of such things made *learn* rather slow and demanding of system resources. The new version is about 4 or 5 times faster, because fewer files and processes are created. Furthermore, it appears even faster to the user because in a typical lesson, the printing of the message comes first, and file setup with *#create* can be overlapped with printing, so that when the program finishes printing, it is really ready for the user to type at it.

It is also a great advantage to the script maintainer that lessons are now just ordinary text files, rather than archives. They can be edited without any difficulty, and UNIX text manipulation tools can be applied to them. The result has been that there is much less resistance to going in and fixing substandard lessons.

6. Conclusions

The following observations can be made about secretaries, typists, and other non-programmers who have used *learn*:

(a) A novice must have assistance with the mechanics of communicating with the computer to get through to the first lesson or two; once the first few lessons are passed people can proceed on their own.

(b) The terminology used in the first few lessons is obscure to those inexperienced with computers. It would help if there were a low level reference card for UNIX to supplement the existing programmer oriented bulky manual and bulky reference card.

(c) The concept of "substitutable argument" is hard to grasp, and requires help.

(d) They enjoy the system for the most part. Motivation matters a great deal, however.

It takes an hour or two for a novice to get through the script on file handling. The total time for a reasonably intelligent and motivated novice to proceed from ignorance to a reasonable ability to create new files and manipulate old ones seems to be a few days, with perhaps half of each day spent on the machine.

The normal way of proceeding has been to have students in the same room with someone who knows the UNIX system and the scripts. Thus the student is not brought to a halt by difficult questions. The burden on the counselor, however, is much lower than that on a teacher of a course. Ideally, the students should be encouraged to proceed with instruction immediately prior to their actual use of the computer. They should exercise the scripts on the same computer and the same kind of terminal that they will later use for their real work, and their first few jobs for the computer should be relatively easy ones. Also, both training and initial work should take place on days when the hardware and software are working reliably. Rarely is all of this possible, but the closer one comes the better the result. For example, if it is known that the hardware is shaky one day, it is better to attempt to reschedule training for another one. Students are very frustrated by machine downtime; when nothing is happening, it takes some sophistication and experience to distinguish an infinite loop, a slow but functioning program, a program waiting for the user, and a broken machine.*

One disadvantage of training with *learn* is that students come to depend completely on the CAI system, and do not try to read manuals or use other learning aids. This is unfortunate, not only because of the increased demands for completeness and accuracy of the scripts, but because the scripts do not cover all of the UNIX system. New users should have manuals (appropriate for their level) and read them; the scripts ought to be altered to recommend suitable documents and urge students to read them.

There are several other difficulties which are clearly evident. From the student's viewpoint, the most serious is that lessons still crop up which simply can't be passed. Sometimes this is due to poor explanations, but just as often it is some error in the lesson itself — a botched setup, a missing file, an invalid test for correctness, or some system facility that doesn't work on the local system in the same way it did on the development system. It takes knowledge and a certain healthy arrogance on the part of the user to recognize that the fault is not his or hers, but the script writer's. Permitting the student to get on with the next lesson regardless does alleviate this somewhat, and the logging facilities make it easy to watch for lessons that no one can pass, but it is still a problem.

The biggest problem with the previous *learn* was speed (or lack thereof) — it was often excruciatingly slow and a significant drain on the system. The current version so far does not seem to have that difficulty, although some scripts, notably *eqn*, are intrinsically slow. *eqn*, for example, must do a lot of work even to print its introductions, let alone check the student responses, but delay is perceptible in all scripts from time to time.

Another potential problem is that it is possible to break *learn* inadvertently, by pushing interrupt at the wrong time, or by removing critical files, or any number of similar slips. The defenses against such problems have steadily been improved, to the point where most students should not notice difficulties. Of course, it will always be possible to break *learn* maliciously, but this is not likely to be a problem.

One area is more fundamental — some commands are sufficiently global in their effect that *learn* currently does not allow them to be executed at all. The most obvious is *cd*, which changes to another directory. The prospect of a studen ho is learning about directories inadvertently moving to some random directory and removing iles has deterred us from even writing lessons on *cd*, but ultimately lessons on such topics probably should be added.

7. Acknowledgments

We are grateful to all those who have tried *learn*, for we have benefited greatly from their suggestions and criticisms. In particular, M. E. Bittrich, J. L. Blue, S. I. Feldman, P. A. Fox, and M. J. McAlpin have provided substantial feedback. Conversations with E. Z. Rothkopf also provided many of the ideas in the system. We are also indebted to Don Jackowski for serving

* We have even known an expert programmer to decide the computer was broken when he had simply left his terminal in local mode. Novices have great difficulties with such problems.

as a guinea pig for the second version, and to Tom Plum for his efforts to improve the C script.

References

1. D. L. Bitzer and D. Skaperdas, "The Economics of a Large Scale Computer Based Education System: Plato IV," pp. 17-29 in *Computer Assisted Instruction, Testing and Guidance*, ed. Wayne Holtzman, Harper and Row, New York (1970).
2. D. C. Gray, J. P. Hulskamp, J. H. Kumm, S. Lichtenstein, and N. E. Nimmervoll, "COALA - A Minicomputer CAI System," *IEEE Trans. Education* **E-20**(1), pp.73-77 (Feb. 1977).
3. P. Suppes, "On Using Computers to Individualize Instruction," pp. 11-24 in *The Computer in American Education*, ed. D. D. Bushnell and D. W. Allen, John Wiley, New York (1967).
4. B. F. Skinner, "Why We Need Teaching Machines," *Harv. Educ. Review* **31**, pp.377-398, Reprinted in *Educational Technology*, ed. J. P. DeCecco, Holt, Rinehart & Winston (New York, 1964). (1961).
5. K. Thompson and D. M. Ritchie, *UNIX Programmer's Manual*, Bell Laboratories (1978). See section *ed* (I).
6. B. W. Kernighan, *A tutorial introduction to the UNIX text editor*, Bell Laboratories internal memorandum (1974).
7. B. W. Kernighan and D. M. Ritchie, *The C Programming Language*, Prentice-Hall, Englewood Cliffs, New Jersey (1978).

APPENDIX A – Page given to new users

How to Get Started

Absolutely basic information for using the UNIX system from DASI, Terminet, or HP terminals

First time. BRING A FRIEND. Anyone who has used UNIX before, however briefly, will be of enormous help for the first fifteen minutes to show you where all the switches are and supply information missing from this page.

Terminals. Turn the power on. There are many kinds of terminals. Look at the telephone used with the terminal to distinguish them. Terminals may have

– *old style datasets* (if the phone set is a small gray box with "talk" and "data" buttons at the right above the handset)

– *new style datasets* (if the phone set is a black six button phone with a red "data" button on the left, sitting on a rectangular box with a glass front)

– *acoustic couplers* (if an ordinary telephone is used to call and the terminal has rubber receptacles that the handset fits into) or

– *modems* (if the phone used for calling has a white button for the left button of the pair of buttons the handset usually rests on).

– *none of the above* (in which case there is probably a switch somewhere that should be flipped to signal the computer).

Calling in. For your local UNIX call ______________.

– If the terminal doesn't use a phone, ignore this section, and proceed to *Login.*.

– On terminals with datasets you must push the "talk" button to get a dial tone.

– If the terminal has a separate coupler turn the coupler power on.

– If the line is busy UNIX is probably full.

– If there is no answer UNIX is broken.

Usually the phone rings only once; UNIX answers and whistles at you.

Connecting the terminal. Remember what kind of terminal you have. If it uses a

– *dataset,* push down the "data" button, let it spring back up, and then hang up the handset (IN THAT ORDER).

– *coupler,* place the handset in the rubber receptacles. There will be an indication of where the phone cord should be (it matters). You may get better results by placing the handset in the receptacles as you dial.

– *modem,* pull up the white button on the telephone and put the handset down somewhere (but don't hang up the phone!).

Login. UNIX should type "login:". If it does not:

– Your terminal may be in "local" mode – check that the "local/line" switch is on "line". Also, Terminets may have their "interrupt" light on – turn it off by pushing "ready."

– If the message is garbled, the speed is wrong. Somewhere on the terminal is a switch labeled "rate" or "baud" with positions of either "10,15,30" or "110,150,300". Set it to 30 or 300. Push the break or interrupt button slowly a few times. If "login:" doesn't appear, call for help.

– UNIX may be broken (call ext. ______ to check on that).

Type your userid, followed by "return". Your userid is __________.

– If each letter appears twice, find the switch labeled "full/half duplex" and set it to "full".

– If the computer typed back your userid in upper case, find the "all caps" switch or "shift lock" and turn it off. Then dial in again.

Normally UNIX says "Password:" and you should enter your password; printing will be turned off while you do.

If you misspell it, UNIX will say "Login incorrect. login:" and you can then retype your userid and password correctly.

UNIX will say "$". You have successfully logged in.

Commands. When UNIX has typed "$" you can type commands, one per line. For example, you can type "date" to find out what day and time it is, or "who" to find out who is logged on. Every command must end with a "return". After typing a command, wait for the next "$" to see what happens. For example, your terminal paper might look like this (what the computer typed is in italics):

```
login: myid
Password: <you can't see it>
$ date
Thu Jan 15 10:58:21 EST 1979
$
```

There are a great many other commands you can type (see the guides below) and in particular the *learn* command can help you learn some features of UNIX.

— If you make a mistake typing: the character # will erase the previous character, so that typing

```
dax#te
```

is the same as typing

```
date
```

and the character @ will erase the entire line; typing

```
xxxxx@
date
```

is the same as typing "date". UNIX supplies the carriage return after the @.

— You must hit return if you expect the computer to notice what you typed; otherwise it will wait patiently and silently for you to do so. When in doubt, type return and see what happens.

— If you make a typing error and don't correct it with # or @ before hitting return, the computer will typically say

```
datr: not found
```

where "datr" is the erroneous input line.

— Other messages that may arise from mistyping include *"cannot execute"* or *"No match"* or just *"?"*. The cure is almost always to retype the offending line correctly.

Terminology. Everything stored on the computer is saved in *files*. A file might contain, for example, a memo or a chapter of a book or a letter. Every file has a name, which is used whenever you want to refer to it. Sample names might be "chap3" or "memo2". The files are grouped into *directories;* each directory contains the names of several files. All users have directories containing their own files.

Logging out. Just hang up. On a terminal with a data set, push the "talk" button. On other terminals hang up the handset. Turn the terminal power off.

Guides. You should have copies of *UNIX For Beginners* and *A Tutorial Introduction to the UNIX Text Editor.*

Section 7

THE M4 MACRO PROCESSOR

INTRODUCTION

m4, a macro processor, was developed at Bell Laboratories and is licensed by Western Electric for use on the 8560. The remainder of this section is a reprint of an article describing *m4*. The Technical Notes section of this manual describes the limitations of this program and any changes made to this program by Tektronix.

The M4 Macro Processor

Brian W. Kernighan

Dennis M. Ritchie

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

M4 is a macro processor available on UNIX† and GCOS. Its primary use has been as a front end for Ratfor for those cases where parameterless macros are not adequately powerful. It has also been used for languages as disparate as C and Cobol. M4 is particularly suited for functional languages like Fortran, PL/I and C since macros are specified in a functional notation.

M4 provides features seldom found even in much larger macro processors, including

- arguments
- condition testing
- arithmetic capabilities
- string and substring functions
- file manipulation

This paper is a user's manual for M4.

July 1, 1977

†UNIX is a Trademark of Bell Laboratories.

The M4 Macro Processor

Brian W. Kernighan

Dennis M. Ritchie

Bell Laboratories
Murray Hill, New Jersey 07974

Introduction

A macro processor is a useful way to enhance a programming language, to make it more palatable or more readable, or to tailor it to a particular application. The **#define** statement in C and the analogous **define** in Ratfor are examples of the basic facility provided by any macro processor — replacement of text by other text.

The M4 macro processor is an extension of a macro processor called M3 which was written by D. M. Ritchie for the AP-3 minicomputer; M3 was in turn based on a macro processor implemented for [1]. Readers unfamiliar with the basic ideas of macro processing may wish to read some of the discussion there.

M4 is a suitable front end for Ratfor and C, and has also been used successfully with Cobol. Besides the straightforward replacement of one string of text by another, it provides macros with arguments, conditional macro expansion, arithmetic, file manipulation, and some specialized string processing functions.

The basic operation of M4 is to copy its input to its output. As the input is read, however, each alphanumeric "token" (that is, string of letters and digits) is checked. If it is the name of a macro, then the name of the macro is replaced by its defining text, and the resulting string is pushed back onto the input to be rescanned. Macros may be called with arguments, in which case the arguments are collected and substituted into the right places in the defining text before it is rescanned.

M4 provides a collection of about twenty built-in macros which perform various useful operations; in addition, the user can define new macros. Built-ins and user-defined macros work exactly the same way, except that some of the built-in macros have side effects on the state of the process.

Usage

On UNIX, use

```
m4 [files]
```

Each argument file is processed in order; if there are no arguments, or if an argument is '−', the standard input is read at that point. The processed text is written on the standard output, which may be captured for subsequent processing with

```
m4 [files] >outputfile
```

On GCOS, usage is identical, but the program is called **./m4**.

Defining Macros

The primary built-in function of M4 is **define**, which is used to define new macros. The input

```
define(name, stuff)
```

causes the string **name** to be defined as **stuff**. All subsequent occurrences of **name** will be replaced by **stuff**. **name** must be alphanumeric and must begin with a letter (the underscore _ counts as a letter). **stuff** is any text that contains balanced parentheses; it may stretch over multiple lines.

Thus, as a typical example,

```
define(N, 100)
...
if (i > N)
```

defines N to be 100, and uses this "symbolic constant'' in a later **if** statement.

The left parenthesis must immediately follow the word **define**, to signal that **define** has arguments. If a macro or built-in name is not followed immediately by \`(', it is assumed to have no arguments. This is the situation for N above; it is actually a macro with no arguments, and thus when it is used there need be no (...) following it.

You should also notice that a macro name is only recognized as such if it appears surrounded by non-alphanumerics. For example, in

```
define(N, 100)
...
if (NNN > 100)
```

the variable NNN is absolutely unrelated to the defined macro N, even though it contains a lot of N's.

Things may be defined in terms of other things. For example,

```
define(N, 100)
define(M, N)
```

defines both M and N to be 100.

What happens if N is redefined? Or, to say it another way, is M defined as N or as 100? In M4, the latter is true — **M** is 100, so even if N subsequently changes, **M** does not.

This behavior arises because M4 expands macro names into their defining text as soon as it possibly can. Here, that means that when the string N is seen as the arguments of **define** are being collected, it is immediately replaced by 100; it's just as if you had said

```
define(M, 100)
```

in the first place.

If this isn't what you really want, there are two ways out of it. The first, which is specific to this situation, is to interchange the order of the definitions:

```
define(M, N)
define(N, 100)
```

Now M is defined to be the string N, so when you ask for M later, you'll always get the value of N at that time (because the M will be replaced by N which will be replaced by 100).

Quoting

The more general solution is to delay the expansion of the arguments of **define** by *quoting* them. Any text surrounded by the single quotes \` and ´ is not expanded immediately, but has the quotes stripped off. If you say

```
define(N, 100)
define(M, `N´)
```

the quotes around the N are stripped off as the argument is being collected, but they have served their purpose, and M is defined as the string N, not 100. The general rule is that M4 always strips off one level of single quotes whenever it evaluates something. This is true even outside of macros. If you want the word **define** to appear in the output, you have to quote it in the input, as in

```
`define´ = 1;
```

As another instance of the same thing, which is a bit more surprising, consider redefining N:

```
define(N, 100)
...
define(N, 200)
```

Perhaps regrettably, the N in the second definition is evaluated as soon as it's seen; that is, it is replaced by 100, so it's as if you had written

```
define(100, 200)
```

This statement is ignored by M4, since you can only define things that look like names, but it obviously doesn't have the effect you wanted. To really redefine N, you must delay the evaluation by quoting:

```
define(N, 100)
...
define(`N´, 200)
```

In M4, it is often wise to quote the first argument of a macro.

If \` and ´ are not convenient for some reason, the quote characters can be changed with the built-in **changequote**:

```
changequote([, ])
```

makes the new quote characters the left and right brackets. You can restore the original characters with just

changequote

There are two additional built-ins related to **define**. **undefine** removes the definition of some macro or built-in:

```
undefine(`N')
```

removes the definition of N. (Why are the quotes absolutely necessary?) Built-ins can be removed with **undefine**, as in

```
undefine(`define')
```

but once you remove one, you can never get it back.

The built-in **ifdef** provides a way to determine if a macro is currently defined. In particular, M4 has pre-defined the names **unix** and **gcos** on the corresponding systems, so you can tell which one you're using:

```
ifdef(`unix', `define(wordsize,16)' )
ifdef(`gcos', `define(wordsize,36)' )
```

makes a definition appropriate for the particular machine. Don't forget the quotes!

ifdef actually permits three arguments; if the name is undefined, the value of **ifdef** is then the third argument, as in

```
ifdef(`unix', on UNIX, not on UNIX)
```

Arguments

So far we have discussed the simplest form of macro processing — replacing one string by another (fixed) string. User-defined macros may also have arguments, so different invocations can have different results. Within the replacement text for a macro (the second argument of its **define**) any occurrence of **$n** will be replaced by the **n**th argument when the macro is actually used. Thus, the macro **bump**, defined as

```
define(bump, $1 = $1 + 1)
```

generates code to increment its argument by 1:

```
bump(x)
```

is

```
x = x + 1
```

A macro can have as many arguments as you want, but only the first nine are accessible, through **$1** to **$9**. (The macro name itself is **$0**, although that is less commonly used.) Arguments that are not supplied are replaced by null strings, so we can define a macro **cat** which simply concatenates its arguments, like this:

```
define(cat, $1$2$3$4$5$6$7$8$9)
```

Thus

```
cat(x, y, z)
```

is equivalent to

```
xyz
```

$4 through **$9** are null, since no corresponding arguments were provided.

Leading unquoted blanks, tabs, or newlines that occur during argument collection are discarded. All other white space is retained. Thus

```
define(a,   b   c)
```

defines **a** to be **b   c**.

Arguments are separated by commas, but parentheses are counted properly, so a comma "protected" by parentheses does not terminate an argument. That is, in

```
define(a, (b,c))
```

there are only two arguments; the second is literally **(b,c)**. And of course a bare comma or parenthesis can be inserted by quoting it.

Arithmetic Built-ins

M4 provides two built-in functions for doing arithmetic on integers (only). The simplest is **incr**, which increments its numeric argument by 1. Thus to handle the common programming situation where you want a variable to be defined as "one more than N", write

```
define(N, 100)
define(N1, `incr(N)')
```

Then **N1** is defined as one more than the current value of N.

The more general mechanism for arithmetic is a built-in called **eval**, which is capable of arbitrary arithmetic on integers. It provides the operators (in decreasing order of precedence)

```
unary + and -
** or ^       (exponentiation)
* / % (modulus)
+ -
== != < <= > >=
!             (not)
& or &&       (logical and)
| or ||       (logical or)
```

Parentheses may be used to group operations where needed. All the operands of an expression given to **eval** must ultimately be numeric. The numeric value of a true relation (like 1>0) is 1, and false is 0. The precision in **eval** is 32 bits on UNIX and 36 bits on GCOS.

As a simple example, suppose we want **M** to be **2**N+1**. Then

```
define(N, 3)
define(M, `eval(2**N+1)')
```

As a matter of principle, it is advisable to quote the defining text for a macro unless it is very simple indeed (say just a number); it usually gives the result you want, and is a good habit to get into.

File Manipulation

You can include a new file in the input at any time by the built-in function **include**:

```
include(filename)
```

inserts the contents of **filename** in place of the **include** command. The contents of the file is often a set of definitions. The value of **include** (that is, its replacement text) is the contents of the file; this can be captured in definitions, etc.

It is a fatal error if the file named in **include** cannot be accessed. To get some control over this situation, the alternate form **sinclude** can be used; **sinclude** (``silent include'') says nothing and continues if it can't access the file.

It is also possible to divert the output of M4 to temporary files during processing, and output the collected material upon command. M4 maintains nine of these diversions, numbered 1 through 9. If you say

```
divert(n)
```

all subsequent output is put onto the end of a temporary file referred to as **n**. Diverting to this file is stopped by another **divert** command; in particular, **divert** or **divert(0)** resumes the normal output process.

Diverted text is normally output all at once at the end of processing, with the diversions output in numeric order. It is possible, however, to bring back diversions at any time, that is, to append them to the current diversion.

```
undivert
```

brings back all diversions in numeric order, and **undivert** with arguments brings back the selected diversions in the order given. The act of undiverting discards the diverted stuff, as does diverting into a diversion whose number is not between 0 and 9 inclusive.

The value of **undivert** is *not* the diverted stuff. Furthermore, the diverted material is *not* rescanned for macros.

The built-in **divnum** returns the number of the currently active diversion. This is zero during normal processing.

System Command

You can run any program in the local operating system with the **syscmd** built-in. For example,

```
syscmd(date)
```

on UNIX runs the **date** command. Normally **syscmd** would be used to create a file for a subsequent **include**.

To facilitate making unique file names, the built-in **maketemp** is provided, with specifications identical to the system function *mktemp:* a string of XXXXX in the argument is replaced by the process id of the current process.

Conditionals

There is a built-in called **ifelse** which enables you to perform arbitrary conditional testing. In the simplest form,

```
ifelse(a, b, c, d)
```

compares the two strings **a** and **b**. If these are identical, **ifelse** returns the string **c**; otherwise it returns **d**. Thus we might define a macro called **compare** which compares two strings and returns ``yes'' or ``no'' if they are the same or different.

```
define(compare, `ifelse($1, $2, yes, no)')
```

Note the quotes, which prevent too-early evaluation of **ifelse**.

If the fourth argument is missing, it is treated as empty.

ifelse can actually have any number of arguments, and thus provides a limited form of multi-way decision capability. In the input

```
ifelse(a, b, c, d, e, f, g)
```

if the string **a** matches the string **b**, the result is **c**. Otherwise, if **d** is the same as **e**, the result is **f**. Otherwise the result is **g**. If the final argument is omitted, the result is null, so

```
ifelse(a, b, c)
```

is **c** if **a** matches **b**, and null otherwise.

String Manipulation

The built-in **len** returns the length of the string that makes up its argument. Thus

```
len(abcdef)
```

is 6, and **len((a,b))** is 5.

The built-in **substr** can be used to produce substrings of strings. **substr(s, i, n)** returns the substring of s that starts at the **i**th position (origin zero), and is **n** characters long. If **n** is omitted, the rest of the string is returned, so

```
substr(`now is the time', 1)
```

is

```
ow is the time
```

If **i** or **n** are out of range, various sensible things happen.

index(s1, s2) returns the index (position) in **s1** where the string **s2** occurs, or −1 if it doesn't occur. As with **substr**, the origin for strings is 0.

The built-in **translit** performs character transliteration.

```
translit(s, f, t)
```

modifies s by replacing any character found in **f** by the corresponding character of **t**. That is,

```
translit(s, aeiou, 12345)
```

replaces the vowels by the corresponding digits. If **t** is shorter than **f**, characters which don't have an entry in **t** are deleted; as a limiting case, if **t** is not present at all, characters from **f** are deleted from s. So

```
translit(s, aeiou)
```

deletes vowels from s.

There is also a built-in called **dnl** which deletes all characters that follow it up to and including the next newline; it is useful mainly for throwing away empty lines that otherwise tend to clutter up M4 output. For example, if you say

```
define(N, 100)
define(M, 200)
define(L, 300)
```

the newline at the end of each line is not part of the definition, so it is copied into the output, where it may not be wanted. If you add **dnl** to each of these lines, the newlines will disappear.

Another way to achieve this, due to J. E. Weythman, is

```
divert(−1)
        define(...)
        ...
divert
```

Printing

The built-in **errprint** writes its arguments out on the standard error file. Thus you can say

```
errprint(`fatal error')
```

dumpdef is a debugging aid which dumps the current definitions of defined terms. If there are no arguments, you get everything; otherwise you get the ones you name as arguments. Don't forget to quote the names!

Summary of Built-ins

Each entry is preceded by the page number where it is described.

3	changequote(L, R)
1	define(name, replacement)
4	divert(number)
4	divnum
5	dnl
5	dumpdef(\`name', \`name', ...)
5	errprint(s, s, ...)
4	eval(numeric expression)
3	ifdef(\`name', this if true, this if false)
5	ifelse(a, b, c, d)
4	include(file)
3	incr(number)
5	index(s1, s2)
5	len(string)
4	maketemp(...XXXXX...)
4	sinclude(file)
5	substr(string, position, number)
4	syscmd(s)
5	translit(str, from, to)
3	undefine(\`name')
4	undivert(number,number,...)

Acknowledgements

We are indebted to Rick Becker, John Chambers, Doug McIlroy, and especially Jim Weythman, whose pioneering use of M4 has led to several valuable improvements. We are also deeply grateful to Weythman for several substantial contributions to the code.

References

[1] B. W. Kernighan and P. J. Plauger, *Software Tools*, Addison-Wesley, Inc., 1976.

Section 8

SED—A NON-INTERACTIVE TEXT EDITOR

INTRODUCTION

sed, a non-interactive text editor, was developed at Bell Laboratories and is licensed by Western Electric for use on the 8560. The remainder of this section is a reprint of an article describing *sed*. The Technical Notes section of this manual describes the limitations of this program and any changes made to this program by Tektronix.

SED — A Non-interactive Text Editor

Lee E. McMahon

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

Sed is a non-interactive context editor that runs on the UNIX† operating system. *Sed* is designed to be especially useful in three cases:

1) To edit files too large for comfortable interactive editing;
2) To edit any size file when the sequence of editing commands is too complicated to be comfortably typed in interactive mode.
3) To perform multiple 'global' editing functions efficiently in one pass through the input.

This memorandum constitutes a manual for users of *sed.*

August 15, 1978

†UNIX is a Trademark of Bell Laboratories.

SED — A Non-interactive Text Editor

Lee E. McMahon

Bell Laboratories
Murray Hill, New Jersey 07974

Introduction

Sed is a non-interactive context editor designed to be especially useful in three cases:

1) To edit files too large for comfortable interactive editing;
2) To edit any size file when the sequence of editing commands is too complicated to be comfortably typed in interactive mode;
3) To perform multiple 'global' editing functions efficiently in one pass through the input.

Since only a few lines of the input reside in core at one time, and no temporary files are used, the effective size of file that can be edited is limited only by the requirement that the input and output fit simultaneously into available secondary storage.

Complicated editing scripts can be created separately and given to *sed* as a command file. For complex edits, this saves considerable typing, and its attendant errors. *Sed* running from a command file is much more efficient than any interactive editor known to the author, even if that editor can be driven by a pre-written script.

The principal loss of functions compared to an interactive editor are lack of relative addressing (because of the line-at-a-time operation), and lack of immediate verification that a command has done what was intended.

Sed is a lineal descendant of the UNIX editor, *ed*. Because of the differences between interactive and non-interactive operation, considerable changes have been made between *ed* and *sed;* even confirmed users of *ed* will frequently be surprised (and probably chagrined), if they rashly use *sed* without reading Sections 2 and 3 of this document. The most striking family resemblance between the two editors is in the class of patterns ('regular expressions') they recognize; the code for matching patterns is copied almost verbatim from the code for *ed,* and the description of regular expressions in Section 2 is copied almost verbatim from the UNIX Programmer's Manual[1]. (Both code and description were written by Dennis M. Ritchie.)

1. Overall Operation

Sed by default copies the standard input to the standard output, perhaps performing one or more editing commands on each line before writing it to the output. This behavior may be modified by flags on the command line; see Section 1.1 below.

The general format of an editing command is:

```
[address1,address2][function][arguments]
```

One or both addresses may be omitted; the format of addresses is given in Section 2. Any number of blanks or tabs may separate the addresses from the function. The function must be present; the available commands are discussed in Section 3. The arguments may be required or optional, according to which function is given; again, they are discussed in Section 3 under each individual function.

Tab characters and spaces at the beginning of lines are ignored.

1.1. Command-line Flags

Three flags are recognized on the command line:

-n: tells *sed* not to copy all lines, but only those specified by *p* functions or *p* flags after *s* functions (see Section 3.3);

-e: tells *sed* to take the next argument as an editing command;

-f: tells *sed* to take the next argument as a file name; the file should contain editing commands, one to a line.

1.2. Order of Application of Editing Commands

Before any editing is done (in fact, before any input file is even opened), all the editing commands are compiled into a form which will be moderately efficient during the execution phase (when the commands are actually applied to lines of the input file). The commands are compiled in the order in which they are encountered; this is generally the order in which they will be attempted at execution time. The commands are applied one at a time; the input to each command is the output of all preceding commands.

The default linear order of application of editing commands can be changed by the flow-of-control commands, *t* and *b* (see Section 3). Even when the order of application is changed by these commands, it is still true that the input line to any command is the output of any previously applied command.

1.3. Pattern-space

The range of pattern matches is called the pattern space. Ordinarily, the pattern space is one line of the input text, but more than one line can be read into the pattern space by using the *N* command (Section 3.6.).

1.4. Examples

Examples are scattered throughout the text. Except where otherwise noted, the examples all assume the following input text:

```
In Xanadu did Kubla Khan
A stately pleasure dome decree:
Where Alph, the sacred river, ran
Through caverns measureless to man
Down to a sunless sea.
```

(In no case is the output of the *sed* commands to be considered an improvement on Coleridge.)

Example:

The command

```
2q
```

will quit after copying the first two lines of the input. The output will be:

```
In Xanadu did Kubla Khan
A stately pleasure dome decree:
```

2. ADDRESSES: Selecting lines for editing

Lines in the input file(s) to which editing commands are to be applied can be selected by addresses. Addresses may be either line numbers or context addresses.

The application of a group of commands can be controlled by one address (or address-pair) by grouping the commands with curly braces ('{ }')(Sec. 3.6.).

2.1. Line-number Addresses

A line number is a decimal integer. As each line is read from the input, a line-number counter is incremented; a line-number address matches (selects) the input line which causes the internal counter to equal the address line-number. The counter runs cumulatively through multiple input files; it is not reset when a new input file is opened.

As a special case, the character $ matches the last line of the last input file.

2.2. Context Addresses

A context address is a pattern ('regular expression') enclosed in slashes ('/'). The regular expressions recognized by *sed* are constructed as follows:

1) An ordinary character (not one of those discussed below) is a regular expression, and matches that character.
2) A circumflex '^' at the beginning of a regular expression matches the null character at the beginning of a line.
3) A dollar-sign '$' at the end of a regular expression matches the null character at the end of a line.
4) The characters '\n' match an imbedded newline character, but not the newline at the end of the pattern space.
5) A period '.' matches any character except the terminal newline of the pattern space.
6) A regular expression followed by an asterisk '*' matches any number (including 0) of adjacent occurrences of the regular expression it follows.
7) A string of characters in square brackets '[ ]' matches any character in the string, and no others. If, however, the first character of the string is circumflex '^', the regular expression matches any character *except* the characters in the string and the terminal newline of the pattern space.
8) A concatenation of regular expressions is a regular expression which matches the concatenation of strings matched by the components of the regular expression.
9) A regular expression between the sequences '\(' and '\)' is identical in effect to the unadorned regular expression, but has side-effects which are described under the *s* command below and specification 10) immediately below.
10) The expression '\d' means the same string of characters matched by an expression enclosed in '\(' and '\)' earlier in the same pattern. Here *d* is a single digit; the string specified is that beginning with the *d*th occurrence of '\(' counting from the left. For example, the expression '^\(.*\)\1' matches a line beginning with two repeated occurrences of the same string.
11) The null regular expression standing alone (e.g., '//') is equivalent to the last regular expression compiled.

To use one of the special characters (^ $. * [ ] \ /) as a literal (to match an occurrence of itself in the input), precede the special character by a backslash '\'.

For a context address to 'match' the input requires that the whole pattern within the address match some portion of the pattern space.

2.3. Number of Addresses

The commands in the next section can have 0, 1, or 2 addresses. Under each command the maximum number of allowed addresses is given. For a command to have more addresses than the maximum allowed is considered an error.

If a command has no addresses, it is applied to every line in the input.

If a command has one address, it is applied to all lines which match that address.

If a command has two addresses, it is applied to the first line which matches the first address, and to all subsequent lines until (and including) the first subsequent line which matches the second address. Then an attempt is made on subsequent lines to again match the first address, and the process is repeated.

Two addresses are separated by a comma.

Examples:

/an/	matches lines 1, 3, 4 in our sample text
/an.*an/	matches line 1
/^an/	matches no lines
/./	matches all lines
/\./	matches line 5
/r*an/	matches lines 1,3, 4 (number = zero!)
/\(an\).*\1/	matches line 1

3. FUNCTIONS

All functions are named by a single character. In the following summary, the maximum number of allowable addresses is given enclosed in parentheses, then the single character function name, possible arguments enclosed in angles (< >), an expanded English translation of the single-character name, and finally a description of what each function does. The angles around the arguments are *not* part of the argument, and should not be typed in actual editing commands.

3.1. Whole-line Oriented Functions

(2)d -- delete lines

The *d* function deletes from the file (does not write to the output) all those lines matched by its address(es).

It also has the side effect that no further commands are attempted on the corpse of a deleted line; as soon as the *d* function is executed, a new line is read from the input, and the list of editing commands is re-started from the beginning on the new line.

(2)n -- next line

The *n* function reads the next line from the input, replacing the current line. The current line is written to the output if it should be. The list of editing commands is continued following the *n* command.

(1)a\
<text> -- append lines

The *a* function causes the argument <text> to be written to the output after the line matched by its address. The *a* command is inherently multi-line; *a* must appear at the end of a line, and <text> may contain any number of lines. To preserve the one-command-to-a-line fiction, the interior newlines must be hidden by a backslash character ('\') immediately preceding the newline. The <text> argument is terminated by the first unhidden newline (the first one not immediately preceded by backslash).

Once an *a* function is successfully executed, <text> will be written to the output regardless of what later commands do to the line which triggered it. The triggering line may be deleted entirely; <text> will still be written to the output.

The <text> is not scanned for address matches, and no editing commands are attempted on it. It does not cause any change in the line-number counter.

(1)i\
<text> -- insert lines

The *i* function behaves identically to the *a* function, except that <text> is written to the output *before* the matched line. All other comments about the *a* function apply to the *i* function as well.

(2)c\
<text> -- change lines

The *c* function deletes the lines selected by its address(es), and replaces them with the lines in <text>. Like *a* and *i*, *c* must be followed by a newline hidden by a backslash; and interior new lines in <text> must be hidden by backslashes.

The *c* command may have two addresses, and therefore select a range of lines. If it does, all the lines in the range are deleted, but only one copy of <text> is written to the output, *not* one copy per line deleted. As with *a* and *i*, <text> is not scanned for address matches, and no editing commands are attempted on it. It does not change the line-number counter.

After a line has been deleted by a *c* function, no further commands are attempted on the corpse.

If text is appended after a line by *a* or *r* functions, and the line is subsequently changed, the text inserted by the *c* function will be placed *before* the text of the *a* or *r* functions. (The *r* function is described in Section 3.4.)

Note: Within the text put in the output by these functions, leading blanks and tabs will disappear, as always in *sed* commands. To get leading blanks and tabs into the output, precede the first desired blank or tab by a backslash; the backslash will not appear in the output.

Example:

The list of editing commands:

```
n
a\
XXXX
d
```

applied to our standard input, produces:

```
In Xanadu did Kubhla Khan
XXXX
Where Alph, the sacred river, ran
XXXX
Down to a sunless sea.
```

In this particular case, the same effect would be produced by either of the two following command lists:

```
n       n
i\      c\
XXXX    XXXX
d
```

3.2. Substitute Function

One very important function changes parts of lines selected by a context search within the line.

(2)s<pattern><replacement><flags> -- substitute

The *s* function replaces *part* of a line (selected by <pattern>) with <replacement>. It can best be read:

Substitute for <pattern>, <replacement>

The <pattern> argument contains a pattern, exactly like the patterns in addresses (see 2.2 above). The only difference between <pattern> and a context address is that the context address must be delimited by slash ('/') characters; <pattern> may be delimited by any character other than space or newline.

By default, only the first string matched by <pattern> is replaced, but see the *g* flag below.

The <replacement> argument begins immediately after the second delimiting character of <pattern>, and must be followed immediately by another instance of the delimiting character. (Thus there are exactly *three* instances of the delimiting character.)

The <replacement> is not a pattern, and the characters which are special in patterns do not have special meaning in <replacement>. Instead, other characters are special:

& is replaced by the string matched by <pattern>

\d (where *d* is a single digit) is replaced by the *d*th substring matched by parts of <pattern> enclosed in '\\(' and '\\)'. If nested substrings occur in <pattern>, the *d*th is determined by counting opening delimiters ('\\(').

As in patterns, special characters may be made literal by preceding them with backslash ('\\').

The <flags> argument may contain the following flags:

g -- substitute <replacement> for all (non-overlapping) instances of <pattern> in the line. After a successful substitution, the scan for the next instance of <pattern> begins just after the end of the inserted characters; characters put into the line from <replacement> are not rescanned.

p -- print the line if a successful replacement was done. The *p* flag causes the line to be written to the output if and only if a substitution was actually made by the *s* function. Notice that if several *s* functions, each followed by a *p* flag, successfully substitute in the same input line, multiple copies of the line will be written to the output: one for each successful substitution.

w <filename> -- write the line to a file if a successful replacement was done. The *w* flag causes lines which are actually substituted by the *s* function to be written to a file named by <filename>. If <filename> exists before *sed* is run, it is overwritten; if not, it is created.

A single space must separate *w* and <filename>.

The possibilities of multiple, somewhat different copies of one input line being written are the same as for *p*.

A maximum of 10 different file names may be mentioned after *w* flags and *w* functions (see below), combined.

Examples:

The following command, applied to our standard input,

```
s/to/by/w changes
```

produces, on the standard output:

```
In Xanadu did Kubhla Khan
A stately pleasure dome decree:
Where Alph, the sacred river, ran
Through caverns measureless by man
Down by a sunless sea.
```

and, on the file 'changes':

```
Through caverns measureless by man
Down by a sunless sea.
```

If the nocopy option is in effect, the command:

```
s/[.,;?:]/*P&*/gp
```

produces:

```
A stately pleasure dome decree*P:*
Where Alph*P,* the sacred river*P,* ran
Down to a sunless sea*P.*
```

Finally, to illustrate the effect of the *g* flag, the command:

```
/X/s/an/AN/p
```

produces (assuming nocopy mode):

```
In XANadu did Kubhla Khan
```

and the command:

```
/X/s/an/AN/gp
```

produces:

```
In XANadu did Kubhla KhAN
```

3.3. Input-output Functions

(2)p -- print

The print function writes the addressed lines to the standard output file. They are written at the time the *p* function is encountered, regardless of what succeeding editing commands may do to the lines.

(2)w <filename> -- write on <filename>

The write function writes the addressed lines to the file named by <filename>. If the file previously existed, it is overwritten; if not, it is created. The lines are written exactly as they exist when the write function is encountered for each line, regardless of what subsequent editing commands may do to them.

Exactly one space must separate the *w* and <filename>.

A maximum of ten different files may be mentioned in write functions and *w* flags after *s* functions, combined.

(1)r <filename> -- read the contents of a file

The read function reads the contents of <filename>, and appends them after the line matched by the address. The file is read and appended regardless of what subsequent editing commands do to the line which matched its address. If *r* and *a* functions are executed on the same line, the text from the *a* functions and the *r* functions is written to the output in the order that the functions are executed.

Exactly one space must separate the *r* and <filename>. If a file mentioned by a *r* function cannot be opened, it is considered a null file, not an error, and no diagnostic is given.

NOTE: Since there is a limit to the number of files that can be opened simultaneously, care should be taken that no more than ten files be mentioned in *w* functions or flags; that number is reduced by one if any *r* functions are present. (Only one read file is open at one time.)

Examples

Assume that the file 'note1' has the following contents:

```
                Note: Kubla Khan (more properly Kublai Khan; 1216-1294) was the grandson
                and most eminent successor of Genghiz (Chingiz) Khan, and founder of the
                Mongol dynasty in China.
```

Then the following command:

```
        /Kubla/r note1
```

produces:

```
        In Xanadu did Kubla Khan
                Note: Kubla Khan (more properly Kublai Khan; 1216-1294) was the grandson
                and most eminent successor of Genghiz (Chingiz) Khan, and founder of the
                Mongol dynasty in China.
        A stately pleasure dome decree:
        Where Alph, the sacred river, ran
        Through caverns measureless to man
        Down to a sunless sea.
```

3.4. Multiple Input-line Functions

Three functions, all spelled with capital letters, deal specially with pattern spaces containing imbedded newlines; they are intended principally to provide pattern matches across lines in the input.

(2)N -- Next line

The next input line is appended to the current line in the pattern space; the two input lines are separated by an imbedded newline. Pattern matches may extend across the imbedded newline(s).

(2)D -- Delete first part of the pattern space

Delete up to and including the first newline character in the current pattern space. If the pattern space becomes empty (the only newline was the terminal newline), read another line from the input. In any case, begin the list of editing commands again from its beginning.

(2)P -- Print first part of the pattern space

Print up to and including the first newline in the pattern space.

The *P* and *D* functions are equivalent to their lower-case counterparts if there are no imbedded newlines in the pattern space.

3.5. Hold and Get Functions

Four functions save and retrieve part of the input for possible later use.

(2)h -- hold pattern space

> The *h* functions copies the contents of the pattern space into a hold area (destroying the previous contents of the hold area).

(2)H -- Hold pattern space

> The *H* function appends the contents of the pattern space to the contents of the hold area; the former and new contents are separated by a newline.

(2)g -- get contents of hold area

> The *g* function copies the contents of the hold area into the pattern space (destroying the previous contents of the pattern space).

(2)G -- Get contents of hold area

> The *G* function appends the contents of the hold area to the contents of the pattern space; the former and new contents are separated by a newline.

(2)x -- exchange

> The exchange command interchanges the contents of the pattern space and the hold area.

Example

The commands

```
1h
1s/ did.*//
1x
G
s/\n/  :/
```

applied to our standard example, produce:

```
In Xanadu did Kubla Khan  :In Xanadu
A stately pleasure dome decree:  :In Xanadu
Where Alph, the sacred river, ran  :In Xanadu
Through caverns measureless to man  :In Xanadu
Down to a sunless sea.  :In Xanadu
```

3.6. Flow-of-Control Functions

These functions do no editing on the input lines, but control the application of functions to the lines selected by the address part.

(2)! -- Don't

> The *Don't* command causes the next command (written on the same line), to be applied to all and only those input lines *not* selected by the adress part.

(2){ -- Grouping

> The grouping command '{' causes the next set of commands to be applied (or not applied) as a block to the input lines selected by the addresses of the grouping command. The first of the commands under control of the grouping may appear on the same line as the '{' or on the next line.

The group of commands is terminated by a matching '}' standing on a line by itself.

Groups can be nested.

(0):<label> -- place a label

The label function marks a place in the list of editing commands which may be referred to by *b* and *t* functions. The <label> may be any sequence of eight or fewer characters; if two different colon functions have identical labels, a compile time diagnostic will be generated, and no execution attempted.

(2)b<label> -- branch to label

The branch function causes the sequence of editing commands being applied to the current input line to be restarted immediately after the place where a colon function with the same <label> was encountered. If no colon function with the same label can be found after all the editing commands have been compiled, a compile time diagnostic is produced, and no execution is attempted.

A *b* function with no <label> is taken to be a branch to the end of the list of editing commands; whatever should be done with the current input line is done, and another input line is read; the list of editing commands is restarted from the beginning on the new line.

(2)t<label> -- test substitutions

The *t* function tests whether *any* successful substitutions have been made on the current input line; if so, it branches to <label>; if not, it does nothing. The flag which indicates that a successful substitution has been executed is reset by:

1) reading a new input line, or
2) executing a *t* function.

3.7. Miscellaneous Functions

(1)= -- equals

The = function writes to the standard output the line number of the line matched by its address.

(1)q -- quit

The *q* function causes the current line to be written to the output (if it should be), any appended or read text to be written, and execution to be terminated.

Reference

[1] Ken Thompson and Dennis M. Ritchie, *The UNIX Programmer's Manual.* Bell Laboratories, 1978.

Section 9

AWK—A PATTERN SCANNING AND PROCESSING LANGUAGE

INTRODUCTION

awk, a pattern scanning and processing language, was developed at Bell Laboratories and is licensed by Western Electric for use on the 8560. The remainder of this section is a reprint of an article describing *awk*. The Technical Notes section of this manual describes the limitations of this program and any changes made to this program by Tektronix.

Awk — A Pattern Scanning and Processing Language (Second Edition)

Alfred V. Aho

Brian W. Kernighan

Peter J. Weinberger

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

Awk is a programming language whose basic operation is to search a set of files for patterns, and to perform specified actions upon lines or fields of lines which contain instances of those patterns. *Awk* makes certain data selection and transformation operations easy to express; for example, the *awk* program

```
length > 72
```

prints all input lines whose length exceeds 72 characters; the program

```
NF % 2 == 0
```

prints all lines with an even number of fields; and the program

```
{ $1 = log($1); print }
```

replaces the first field of each line by its logarithm.

Awk patterns may include arbitrary boolean combinations of regular expressions and of relational operators on strings, numbers, fields, variables, and array elements. Actions may include the same pattern-matching constructions as in patterns, as well as arithmetic and string expressions and assignments, **if-else**, **while**, **for** statements, and multiple output streams.

This report contains a user's guide, a discussion of the design and implementation of *awk*, and some timing statistics.

September 1, 1978

Awk — A Pattern Scanning and Proce sing Language (Second Edition)

Alfred V. Aho

Brian W. Kernighan

Peter J. Weinberger

Bell Laboratories
Murray Hill, New Jersey 07974

1. Introduction

Awk is a programming language designed to make many common information retrieval and text manipulation tasks easy to state and to perform.

The basic operation of *awk* is to scan a set of input lines in order, searching for lines which match any of a set of patterns which the user has specified. For each pattern, an action can be specified; this action will be performed on each line that matches the pattern.

Readers familiar with the UNIX† program *grep*[1] will recognize the approach, although in *awk* the patterns may be more general than in *grep*, and the actions allowed are more involved than merely printing the matching line. For example, the *awk* program

```
{print $3, $2}
```

prints the third and second columns of a table in that order. The program

```
$2 ~ /A|B|C/
```

prints all input lines with an A, B, or C in the second field. The program

```
$1 != prev   { print; prev = $1 }
```

prints all lines in which the first field is different from the previous first field.

1.1. Usage

The command

```
awk program [files]
```

executes the *awk* commands in the string **program** on the set of named files, or on the standard input if there are no files. The statements can also be placed in a file **pfile**, and executed by the command

```
awk -f pfile [files]
```

1.2. Program Structure

An *awk* program is a sequence of statements of the form:

```
pattern    { action }
pattern    { action }
...
```

Each line of input is matched against each of the patterns in turn. For each pattern that matches, the associated action is executed. When all the patterns have been tested, the next line is fetched and the matching starts over.

Either the pattern or the action may be left out, but not both. If there is no action for a pattern, the matching line is simply copied to the output. (Thus a line which matches several patterns can be printed several times.) If there is no pattern for an action, then the action is performed for every input line. A line which matches no pattern is ignored.

Since patterns and actions are both optional, actions must be enclosed in braces to distinguish them from patterns.

1.3. Records and Fields

Awk input is divided into "records" terminated by a record separator. The default record separator is a newline, so by default *awk* processes its input a line at a time. The number of the current record is available in a variable named **NR**.

Each input record is considered to be divided into "fields." Fields are normally separated by white space — blanks or tabs — but the input field separator may be changed, as described below. Fields are referred to as **$1**, **$2**, and so forth, where **$1** is the first field, and **$0** is the whole input record itself. Fields may

†UNIX is a Trademark of Bell Laboratories.

be assigned to. The number of fields in the current record is available in a variable named **NF**.

The variables **FS** and **RS** refer to the input field and record separators; they may be changed at any time to any single character. The optional command-line argument −F*c* may also be used to set **FS** to the character *c*.

If the record separator is empty, an empty input line is taken as the record separator, and blanks, tabs and newlines are treated as field separators.

The variable **FILENAME** contains the name of the current input file.

1.4. Printing

An action may have no pattern, in which case the action is executed for all lines. The simplest action is to print some or all of a record; this is accomplished by the *awk* command **print**. The *awk* program

```
{ print }
```

prints each record, thus copying the input to the output intact. More useful is to print a field or fields from each record. For instance,

```
print $2, $1
```

prints the first two fields in reverse order. Items separated by a comma in the print statement will be separated by the current output field separator when output. Items not separated by commas will be concatenated, so

```
print $1 $2
```

runs the first and second fields together.

The predefined variables **NF** and **NR** can be used; for example

```
{ print NR, NF, $0 }
```

prints each record preceded by the record number and the number of fields.

Output may be diverted to multiple files; the program

```
{ print $1 >"foo1"; print $2 >"foo2" }
```

writes the first field, **$1**, on the file **foo1**, and the second field on file **foo2**. The >> notation can also be used:

```
print $1 >>"foo"
```

appends the output to the file **foo**. (In each case, the output files are created if necessary.) The file name can be a variable or a field as well as a constant; for example,

```
print $1 >$2
```

uses the contents of field 2 as a file name.

Naturally there is a limit on the number of output files; currently it is 10.

Similarly, output can be piped into another process (on UNIX only); for instance,

```
print | "mail bwk"
```

mails the output to **bwk**.

The variables **OFS** and **ORS** may be used to change the current output field separator and output record separator. The output record separator is appended to the output of the **print** statement.

Awk also provides the **printf** statement for output formatting:

```
printf format expr, expr, ...
```

formats the expressions in the list according to the specification in **format** and prints them. For example,

```
printf "%8.2f  %10ld\n", $1, $2
```

prints **$1** as a floating point number 8 digits wide, with two after the decimal point, and **$2** as a 10-digit long decimal number, followed by a newline. No output separators are produced automatically; you must add them yourself, as in this example. The version of **printf** is identical to that used with C.[2]

2. Patterns

A pattern in front of an action acts as a selector that determines whether the action is to be executed. A variety of expressions may be used as patterns: regular expressions, arithmetic relational expressions, string-valued expressions, and arbitrary boolean combinations of these.

2.1. BEGIN and END

The special pattern **BEGIN** matches the beginning of the input, before the first record is read. The pattern **END** matches the end of the input, after the last record has been processed. **BEGIN** and **END** thus provide a way to gain control before and after processing, for initialization and wrapup.

As an example, the field separator can be set to a colon by

```
BEGIN   { FS = ":" }
... rest of program ...
```

Or the input lines may be counted by

```
END   { print NR }
```

If **BEGIN** is present, it must be the first pattern; **END** must be the last if used.

2.2. Regular Expressions

The simplest regular expression is a literal string of characters enclosed in slashes, like

```
/smith/
```

This is actually a complete *awk* program which will print all lines which contain any occurrence of the name "smith". If a line contains "smith" as part of a larger word, it will also be printed, as in

```
blacksmithing
```

Awk regular expressions include the regular expression forms found in the UNIX text editor *ed*[1] and *grep* (without back-referencing). In addition, *awk* allows parentheses for grouping, | for alternatives, + for "one or more", and ? for "zero or one", all as in *lex*. Character classes may be abbreviated: [a−zA−Z0−9] is the set of all letters and digits. As an example, the *awk* program

```
/[Aa]ho|[Ww]einberger|[Kk]ernighan/
```

will print all lines which contain any of the names "Aho," "Weinberger" or "Kernighan," whether capitalized or not.

Regular expressions (with the extensions listed above) must be enclosed in slashes, just as in *ed* and *sed*. Within a regular expression, blanks and the regular expression metacharacters are significant. To turn of the magic meaning of one of the regular expression characters, precede it with a backslash. An example is the pattern

```
/\/.*\//
```

which matches any string of characters enclosed in slashes.

One can also specify that any field or variable matches a regular expression (or does not match it) with the operators ~ and !~. The program

```
$1 ~ /[jJ]ohn/
```

prints all lines where the first field matches "john" or "John." Notice that this will also match "Johnson", "St. Johnsbury", and so on. To restrict it to exactly [jJ]ohn, use

```
$1 ~ /^[jJ]ohn$/
```

The caret ^ refers to the beginning of a line or field; the dollar sign $ refers to the end.

2.3. Relational Expressions

An *awk* pattern can be a relational expression involving the usual relational operators <, <=, ==, !=, >=, and >. An example is

```
$2 > $1 + 100
```

which selects lines where the second field is at least 100 greater than the first field. Similarly,

```
NF % 2 == 0
```

prints lines with an even number of fields.

In relational tests, if neither operand is numeric, a string comparison is made; otherwise it is numeric. Thus,

```
$1 >= "s"
```

selects lines that begin with an **s**, **t**, **u**, etc. In the absence of any other information, fields are treated as strings, so the program

```
$1 > $2
```

will perform a string comparison.

2.4. Combinations of Patterns

A pattern can be any boolean combination of patterns, using the operators || (or), && (and), and ! (not). For example,

```
$1 >= "s" && $1 < "t" && $1 != "smith"
```

selects lines where the first field begins with "s", but is not "smith". && and || guarantee that their operands will be evaluated from left to right; evaluation stops as soon as the truth or falsehood is determined.

2.5. Pattern Ranges

The "pattern" that selects an action may also consist of two patterns separated by a comma, as in

```
pat1, pat2     { ... }
```

In this case, the action is performed for each line between an occurrence of **pat1** and the next occurrence of **pat2** (inclusive). For example,

```
/start/, /stop/
```

prints all lines between **start** and **stop**, while

```
NR == 100, NR == 200 { ... }
```

does the action for lines 100 through 200 of the input.

3. Actions

An *awk* action is a sequence of action statements terminated by newlines or semicolons. These action statements can be used to do a variety of bookkeeping and string manipulating tasks.

3.1. Built-in Functions

Awk provides a "length" function to compute the length of a string of characters. This program prints each record, preceded by its length:

```
{print length, $0}
```

length by itself is a "pseudo-variable" which yields the length of the current record; **length(argument)** is a function which yields the length of its argument, as in the equivalent

```
{print length($0), $0}
```

The argument may be any expression.

Awk also provides the arithmetic functions **sqrt**, **log**, **exp**, and **int**, for square root, base *e* logarithm, exponential, and integer part of their respective arguments.

The name of one of these built-in functions, without argument or parentheses, stands for the value of the function on the whole record. The program

```
length < 10 || length > 20
```

prints lines whose length is less than 10 or greater than 20.

The function **substr(s, m, n)** produces the substring of **s** that begins at position **m** (origin 1) and is at most **n** characters long. If **n** is omitted, the substring goes to the end of **s**. The function **index(s1, s2)** returns the position where the string **s2** occurs in **s1**, or zero if it does not.

The function **sprintf(f, e1, e2, ...)** produces the value of the expressions **e1**, **e2**, etc., in the **printf** format specified by f. Thus, for example,

```
x = sprintf("%8.2f %10ld", $1, $2)
```

sets **x** to the string produced by formatting the values of **$1** and **$2**.

3.2. Variables, Expressions, and Assignments

Awk variables take on numeric (floating point) or string values according to context. For example, in

```
x = 1
```

x is clearly a number, while in

```
x = "smith"
```

it is clearly a string. Strings are converted to numbers and vice versa whenever context demands it. For instance,

```
x = "3" + "4"
```

assigns 7 to **x**. Strings which cannot be interpreted as numbers in a numerical context will generally have numeric value zero, but it is unwise to count on this behavior.

By default, variables (other than built-ins) are initialized to the null string, which has numerical value zero; this eliminates the need for most **BEGIN** sections. For example, the sums of the first two fields can be computed by

```
     { s1 += $1; s2 += $2 }
END { print s1, s2 }
```

Arithmetic is done internally in floating point. The arithmetic operators are +, −, *, /, and % (mod). The C increment ++ and decrement −− operators are also available, and so are the assignment operators +=, −=, *=, /=, and %=. These operators may all be used in expressions.

3.3. Field Variables

Fields in *awk* share essentially all of the properties of variables — they may be used in arithmetic or string operations, and may be assigned to. Thus one can replace the first field with a sequence number like this:

```
{ $1 = NR; print }
```

or accumulate two fields into a third, like this:

```
{ $1 = $2 + $3; print $0 }
```

or assign a string to a field:

```
{ if ($3 > 1000)
      $3 = "too big"
  print
}
```

which replaces the third field by "too big" when it is, and in any case prints the record.

Field references may be numerical expressions, as in

```
{ print $i, $(i+1), $(i+n) }
```

Whether a field is deemed numeric or string depends on context; in ambiguous cases like

```
if ($1 == $2) ...
```

fields are treated as strings.

Each input line is split into fields automatically as necessary. It is also possible to split any variable or string into fields:

```
n = split(s, array, sep)
```

splits the the string **s** into **array[1]**, ..., **array[n]**. The number of elements found is returned. If the **sep** argument is provided, it is used as the field separator; otherwise **FS** is used as the separator.

3.4. String Concatenation

Strings may be concatenated. For example

```
length($1 $2 $3)
```

returns the length of the first three fields. Or in a **print** statement,

```
print $1 " is " $2
```

prints the two fields separated by " is ". Variables and numeric expressions may also appear in concatenations.

3.5. Arrays

Array elements are not declared; they spring into existence by being mentioned. Subscripts may have *any* non-null value, including non-numeric strings. As an example of a conventional numeric subscript, the statement

```
x[NR] = $0
```

assigns the current input record to the NR-th element of the array x. In fact, it is possible in principle (though perhaps slow) to process the entire input in a random order with the *awk* program

```
     { x[NR] = $0 }
END { ... program ... }
```

The first action merely records each input line in the array x.

Array elements may be named by non-numeric values, which gives *awk* a capability rather like the associative memory of Snobol tables. Suppose the input contains fields with values like **apple**, **orange**, etc. Then the program

```
/apple/   { x["apple"]++ }
/orange/  { x["orange"]++ }
END       { print x["apple"], x["orange"] }
```

increments counts for the named array elements, and prints them at the end of the input.

3.6. Flow-of-Control Statements

Awk provides the basic flow-of-control statements **if-else**, **while**, **for**, and statement grouping with braces, as in C. We showed the if statement in section 3.3 without describing it. The condition in parentheses is evaluated; if it is true, the statement following the if is done. The **else** part is optional.

The **while** statement is exactly like that of C. For example, to print all input fields one per line,

```
i = 1
while (i <= NF) {
    print $i
    ++i
}
```

The for statement is also exactly that of C:

```
for (i = 1; i <= NF; i++)
    print $i
```

does the same job as the **while** statement above.

There is an alternate form of the **for** statement which is suited for accessing the elements of an associative array:

```
for (i in array)
    statement
```

does *statement* with i set in turn to each element of array. The elements are accessed in an apparently random order. Chaos will ensue if i is altered, or if any new elements are accessed during the loop.

The expression in the condition part of an if, **while** or **for** can include relational operators like <, <=, >, >=, == ("is equal to"), and != ("not equal to"); regular expression matches with the match operators ~ and !~; the logical operators ||, &&, and !; and of course parentheses for grouping.

The **break** statement causes an immediate exit from an enclosing **while** or **for**; the **continue** statement causes the next iteration to begin.

The statement **next** causes *awk* to skip immediately to the next record and begin scanning the patterns from the top. The statement **exit** causes the program to behave as if the end of the input had occurred.

Comments may be placed in *awk* programs: they begin with the character # and end with the end of the line, as in

```
print x, y # this is a comment
```

4. Design

The UNIX system already provides several programs that operate by passing input through a selection mechanism. *Grep*, the first and simplest, merely prints all lines which match a single specified pattern. *Egrep* provides more general patterns, i.e., regular expressions in full generality; *fgrep* searches for a set of keywords with a particularly fast algorithm. *Sed*[1] provides most of the editing facilities of the editor *ed*, applied to a stream of input. None of these programs provides numeric capabilities, logical relations, or variables.

Lex[3] provides general regular expression recognition capabilities, and, by serving as a C program generator, is essentially open-ended in its capabilities. The use of *lex*, however, requires a knowledge of C programming, and a *lex* program must be compiled and loaded before use, which discourages its use for one-shot applications.

Awk is an attempt to fill in another part of the matrix of possibilities. It provides general regular expression capabilities and an implicit input/output loop. But it also provides convenient numeric processing, variables, more general selection, and control flow in the actions. It does not require compilation or a knowledge of C. Finally, *awk* provides a convenient way to access fields within lines; it is unique in this respect.

Awk also tries to integrate strings and numbers completely, by treating all quantities as both string and numeric, deciding which representation is appropriate as late as possible. In most cases the user can simply ignore the differences.

Most of the effort in developing *awk* went into deciding what *awk* should or should not do (for instance, it doesn't do string substitution) and what the syntax should be (no explicit operator for concatenation) rather than on writing or debugging the code. We have tried to make the syntax powerful but easy to use and well adapted to scanning files. For example, the absence of declarations and implicit initializations, while probably a bad idea for a general-purpose programming language, is desirable in a language that is meant to be used for tiny programs that may even be composed on the command line.

In practice, *awk* usage seems to fall into two broad categories. One is what might be called "report generation" — processing an input to extract counts, sums, sub-totals, etc. This also includes the writing of trivial data validation programs, such as verifying that a field contains only numeric information or that certain delimiters are properly balanced. The combination of textual and numeric processing is invaluable here.

A second area of use is as a data transformer, converting data from the form produced by one program into that expected by another. The simplest examples merely select fields, perhaps with rearrangements.

5. Implementation

The actual implementation of *awk* uses the language development tools available on the UNIX operating system. The grammar is specified with *yacc*;[4] the lexical analysis is done by *lex*; the regular expression recognizers are deterministic finite automata constructed directly from the expressions. An *awk* program is translated into a parse tree which is then directly executed by a simple interpreter.

Awk was designed for ease of use rather than processing speed; the delayed evaluation of variable types and the necessity to break input into fields makes high speed difficult to achieve in any case. Nonetheless, the program has not proven to be unworkably slow.

Table I below shows the execution (user + system) time on a PDP-11/70 of the UNIX programs *wc*, *grep*, *egrep*, *fgrep*, *sed*, *lex*, and *awk* on the following simple tasks:

1. count the number of lines.
2. print all lines containing "doug".
3. print all lines containing "doug", "ken" or "dmr".
4. print the third field of each line.
5. print the third and second fields of each line, in that order.
6. append all lines containing "doug", "ken", and "dmr" to files "jdoug", "jken", and "jdmr", respectively.
7. print each line prefixed by "line-number : ".
8. sum the fourth column of a table.

The program *wc* merely counts words, lines and characters in its input; we have already mentioned the others. In all cases the input was a file containing 10,000 lines as created by the command *ls −l*; each line has the form

```
−rw−rw−rw−  1 ava 123 Oct 15 17:05 xxx
```

The total length of this input is 452,960 characters. Times for *lex* do not include compile or load.

As might be expected, *awk* is not as fast as the specialized tools *wc*, *sed*, or the programs in the *grep* family, but is faster than the more general tool *lex*. In all cases, the tasks were about as easy to express as *awk* programs as programs in these other languages; tasks involving fields were considerably easier to express as *awk* programs. Some of the test programs are shown in *awk*, *sed* and *lex*.

References

1. K. Thompson and D. M. Ritchie, *UNIX Programmer's Manual,* Bell Laboratories (May 1975). Sixth Edition
2. B. W. Kernighan and D. M. Ritchie, *The C Programming Language,* Prentice-Hall, Englewood Cliffs, New Jersey (1978).
3. M. E. Lesk, "Lex — A Lexical Analyzer Generator," Comp. Sci. Tech. Rep. No. 39, Bell Laboratories, Murray Hill, New Jersey (October 1975).
4. S. C. Johnson, "Yacc — Yet Another Compiler-Compiler," Comp. Sci. Tech. Rep. No. 32, Bell Laboratories, Murray Hill, New Jersey (July 1975).

	Task							
Program	1	2	3	4	5	6	7	8
wc	8.6							
grep	11.7	13.1						
egrep	6.2	11.5	11.6					
fgrep	7.7	13.8	16.1					
sed	10.2	11.6	15.8	29.0	30.5	16.1		
lex	65.1	150.1	144.2	67.7	70.3	104.0	81.7	92.8
awk	15.0	25.6	29.9	33.3	38.9	46.4	71.4	31.1

Table I. Execution Times of Programs. (Times are in sec.)

The programs for some of these jobs are shown below. The *lex* programs are generally too long to show.

AWK:

1.
```
END {print NR}
```
2.
```
/doug/
```
3.
```
/ken|doug|dmr/
```
4.
```
{print $3}
```
5.
```
{print $3, $2}
```
6.
```
/ken/     {print >"jken"}
/doug/    {print >"jdoug"}
/dmr/     {print >"jdmr"}
```
7.
```
{print NR ": " $0}
```
8.
```
    {sum = sum + $4}
END {print sum}
```

SED:

1.
```
$=
```
2.
```
/doug/p
```
3.
```
/doug/p
/doug/d
/ken/p
/ken/d
/dmr/p
/dmr/d
```
4.
```
/[^ ]* [ ]*[^ ]* [ ]*\([^ ]*\) .*/s//\1/p
```
5.
```
/[^ ]* [ ]*\([^ ]*\) [ ]*\([^ ]*\) .*/s//\2 \1/p
```
6.
```
/ken/w jken
/doug/w jdoug
/dmr/w jdmr
```

LEX:

1.
```
%{
int i;
%}
%%
\n   i++;
.    ;
%%
yywrap() {
        printf("%d\n", i);
}
```
2.
```
%%
^.*doug.*$      printf("%s\n", yytext);
.    ;
\n   ;
```

MANUAL CHANGE INFORMATION

At Tektronix, we continually strive to keep up with latest electronic developments by adding circuit and component improvements to our instruments as soon as they are developed and tested.

Sometimes, due to printing and shipping requirements, we can't get these changes immediately into printed manuals. Hence, your manual may contain new change information on following pages.

A single change may affect several sections. Since the change information sheets are carried in the manual until all changes are permanently entered, some duplication may occur. If no such change pages appear following this page, your manual is correct as printed.

Tektronix®
COMMITTED TO EXCELLENCE

MANUAL CHANGE INFORMATION

Date: 2-4-82 Change Reference: C1/282

Product: 8560 Auxiliary Utilities Package Users Manual Part No.: 070-4270-00

DESCRIPTION

TEXT CORRECTIONS

Page 2-1 Line up the heading "Installing the Auxiliary Utilities Package" with the left margin. Immediately under that heading, insert the following information:

<u>NOTE</u>

When you install this Auxiliary Utilities Package, you may receive the following message:

overwrite filename?

This message means that you are trying to install a file that already exists on your system.

If the Native Programming Package has previously been installed on your system, type <u>no</u> when you receive this message for the commands <u>join</u>, <u>sed</u>, and <u>tsort</u>.

If you receive this message for any other command, type <u>no</u>. Change the name of your previously existing file, and try again.