

Chapter II

Overview

2.1 Components of the AWK-to-C Translation System

The AWK-to-C translation system consists of three major components:

1. The AWK-to-C Translator is the program invoked by the user to translate the AWK source program into an equivalent C program. The filename for the translator is `awc`.

2. The Skeleton File is the file used by the translator as the skeleton prototype for building the C output file. It contains the C codes that are common in all generated C programs. The filename for the skeleton file is `awcskel.c`.

3. The Run-time Library comprises all the compiled supporting routines to be called by the generated C programs. The library routines being called are linked to the generated program at link time. The filename for the run-time library is `libawc.a`.

In addition, the generated C program also needs two header files, `awclib.h` and `regex.h`, to be included at preprocessing time.

2.2 Making an Executable Program from an AWK Program

By using the AWK-to-C translation system and an existing C compiler, we can obtain an executable program from an AWK program by the procedure shown in figure 2.1. In this figure, an AWK source program named `foo.awk` is fed into the translator `awc`, which reads `awcskel.c` as the skeleton and produces `foo.c`, a functionally equivalent C program. Then, `foo.c`, along with the included header

files `awclib.h` and `regex.h`, will be compiled by the system's C compiler `cc` to obtain the final product `foo`, which is readily executable.

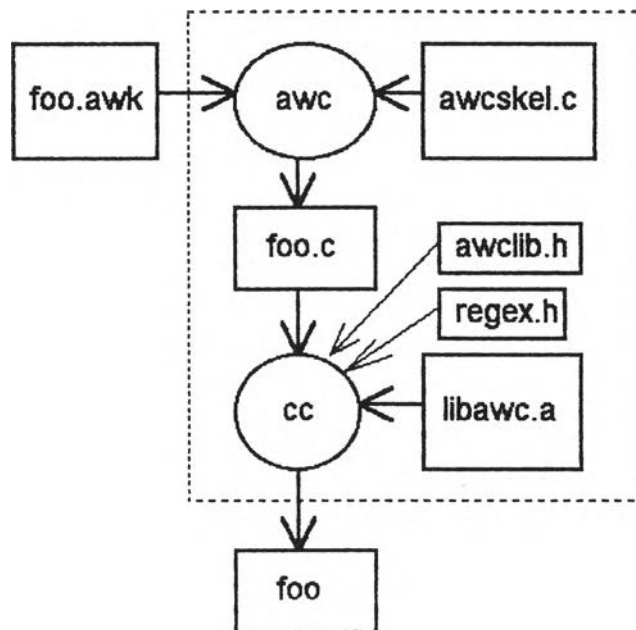


Figure 2.1 Making an executable program from an AWK source program

2.3 Overview of the System Development

2.3.1 Development Environment

The AWK-to-C translation system was developed and tested on the following hardware and system software platform:

Machine Type	IBM PC AT compatible
CPU	Intel 80386
RAM size	8 megabytes
Video Display	EGA monitor
Mass Storage	150 megabyte fixed disk
Operating System	AT&T Unix System V Release 4.0 Version 2.0

The programming language used for coding the entire translation system is ANSI C.

2.3.2 Software Development Tools Used

Constructing the translation system made use of several Unix software development tools. The major ones are:

1. The ANSI C Compiler. The Free Software Foundation's GNU C compiler `gcc` was used to compile all of the source programs. This compiler is preferable to the system's bundled C compiler `cc` because it generally produces faster, better optimized code.

2. The Scanner Generator. The Unix scanner generating program `lex` was used to develop the scanner module of the translator.

3. The Parser Generator. The Unix parser generating program `yacc` was used to develop the parser/code-generator module of the translator.

4. The Library Archiver. The Unix `ar` program was used to bundle all the object files of the `libawc` routines together to form a single run-time library file `libawc.a`.

5. The Program Maintainer. The Unix file updating program `make` was used to automate nearly all stages of software development, which includes generating and compiling the source programs, building

the translator and the library, testing the system, measuring the performance, packaging the source code, and installing the whole software package into the system.

6. GNU Regular Expression Library. The translator-generated programs manipulate AWK regular expression by calling regular expression handling routines in the libawc library. These regular expression handling routines are actually taken from the Free Software Foundation's GNU Regular Expression Library Release 0.12, which is freely available in C source-code form. The code was compiled and the object modules were archived to become part of the libawc library.