The POOLtype processor

(Version 3, September 1989)

The preparation of this report was supported in part by the National Science Foundation under grants IST-8201926 and MCS-8300984, and by the System Development Foundation. ‘TeX’ is a trademark of the American Mathematical Society.

March 17, 2021 at 13:06
1. **Introduction.** The POOLtype utility program converts string pool files output by TANGLE into a slightly more symbolic format that may be useful when TANGLEd programs are being debugged.

   It’s a pretty trivial routine, but people may want to try transporting this program before they get up enough courage to tackle T\TeX\ itself. The first 256 strings are treated as T\TeX\ treats them, using routines copied from T\TeX\82.

   define\ my\_name ≡ 'pooltype'

2. POOLtype is written entirely in standard Pascal, except that it has to do some slightly system-dependent character code conversion on input and output. The input is read from pool\_file, and the output is written on output. If the input is erroneous, the output file will describe the error.

   program POOLtype(pool\_file, output);
   type ⟨Types in the outer block 5⟩
   var ⟨Globals in the outer block 7⟩
   procedure initialize; \{ this procedure gets things started properly \}
   var ⟨Local variables for initialization 6⟩
   begin kpse\_set\_program\_name(argv[0], my\_name); parse\_arguments;
   ⟨Set initial values of key variables 8⟩
   end;
The original Pascal compiler was designed in the late 60s, when six-bit character sets were common, so it did not make provision for lowercase letters. Nowadays, of course, we need to deal with both capital and small letters in a convenient way, especially in a program for typesetting; so the present specification of TeX has been written under the assumption that the Pascal compiler and run-time system permit the use of text files with more than 64 distinguishable characters. More precisely, we assume that the character set contains at least the letters and symbols associated with ASCII codes 40 through 176; all of these characters are now available on most computer terminals.

Since we are dealing with more characters than were present in the first Pascal compilers, we have to decide what to call the associated data type. Some Pascals use the original name char for the characters in text files, even though there now are more than 64 such characters, while other Pascals consider char to be a 64-element subrange of a larger data type that has some other name.

In order to accommodate this difference, we shall use the name text_char to stand for the data type of the characters that are converted to and from ASCII_code when they are input and output. We shall also assume that text_char consists of the elements chr(first_text_char) through chr(last_text_char), inclusive. The following definitions should be adjusted if necessary.

\[
\begin{align*}
\text{define } & \text{ text_char } \equiv \text{ ASCII_code} \quad \{ \text{the data type of characters in text files} \} \\
\text{define } & \text{ first_text_char } = 0 \quad \{ \text{ordinal number of the smallest element of text_char} \} \\
\text{define } & \text{ last_text_char } = 255 \quad \{ \text{ordinal number of the largest element of text_char} \}
\end{align*}
\]

(Local variables for initialization \(6^*\))

\[
\begin{align*}
i: & \text{ integer} \\
\text{This code is used in section } & 2^*.
\end{align*}
\]

The ASCII code is “standard” only to a certain extent, since many computer installations have found it advantageous to have ready access to more than 94 printing characters. Appendix C of The TeXbook gives a complete specification of the intended correspondence between characters and TeX’s internal representation.

If TeX is being used on a garden-variety Pascal for which only standard ASCII codes will appear in the input and output files, it doesn’t really matter what codes are specified in xchr[0 . . 37], but the safest policy is to blank everything out by using the code shown below.

However, other settings of xchr will make TeX more friendly on computers that have an extended character set, so that users can type things like ‘≠’ instead of ‘\ne’. People with extended character sets can assign codes arbitrarily, giving an xchr equivalent to whatever characters the users of TeX are allowed to have in their input files. It is best to make the codes correspond to the intended interpretations as shown in Appendix C whenever possible; but this is not necessary. For example, in countries with an alphabet of more than 26 letters, it is usually best to map the additional letters into codes less than 40. To get the most “permissive” character set, change ‘_‘ on the right of these assignment statements to chr(i).

\[
\begin{align*}
\text{Set initial values of key variables } & 8 \quad +\equiv \\
& \text{for } i \leftarrow 0 \text{ to } 37 \text{ do } xchr[i] \leftarrow \text{chr}(i); \\
& \text{for } i \leftarrow 177 \text{ to } 377 \text{ do } xchr[i] \leftarrow \text{chr}(i);
\end{align*}
\]
15* This is the main program, where POOLtype starts and ends.

```pascal
define abort(#) ≡
begin writeLn(stderr, #); uexit(1);
end
```

begin initialize;
  ⟨Make the first 256 strings 16⟩;
  s ← 256;
  ⟨Read the other strings from the POOL file, or give an error message and abort 19*⟩;
writeLn(′(′, count : 1, ′|characters|in_all.)′); uexit(0);
end.

18* When the WEB system program called TANGLE processes a source file, it outputs a Pascal program and also a string pool file. The present program reads the latter file, where each string appears as a two-digit decimal length followed by the string itself, and the information is output with its associated index number. The strings are surrounded by double-quote marks; double-quotes in the string itself are repeated.

```pascal
(pool_file: packed file of text_char;  { the string-pool file output by TANGLE }
pool_name: const c_string;
xsum: boolean;  { has the check sum been found? }
)

19* ⟨Read the other strings from the POOL file, or give an error message and abort 19*⟩ ≡
xsum ← false;
if eof(pool_file) then abort(′!I can′t read the POOL file.′);
repeat ⟨Read one string, but abort if there are problems 20*⟩;
until xsum;
if ¬eof(pool_file) then abort(′!There′s junk after the check sum′);
```

This code is used in section 15*.

20* ⟨Read one string, but abort if there are problems 20*⟩ ≡
if eof(pool_file) then abort(′!POOL file contained no check sum′);
read(pool_file, m); read(pool_file, n);  { read two digits of string length }
if m ≠ ′*′ then
begin if (xord[m] < "0") ∨ (xord[m] > "9") ∨ (xord[n] < "0") ∨ (xord[n] > "9")
then abort(′!POOL line doesn′t begin with two digits′);
l ← xord[m] * 10 + xord[n] − "0" * 11;  { compute the length }
write(s : 3, ′|′); count ← count + l;
for k ← 1 to l do
begin if eoln(pool_file) then
begin writeLn(′|′); abort(′!POOL line was too short′);
end;
read(pool_file, m); write(xchr[xord[m]]); 
if xord[m] = "*" then write(xchr["*"]);
end:
writeLn(′|′); incr(s);
end
else xsum ← true;
readLn(pool_file)
```

This code is used in section 19*.
System-dependent changes. Parse a Unix-style command line.

```pascal
define argument_is(#) \equiv (strcmp(long_options[option_index].name, #) = 0)

(Define parse_arguments 21*) \equiv
procedure parse_arguments;
  const n_options = 2; { Pascal won’t count array lengths for us. }
  var long_options: array [0 .. n_options] of getopt_struct;
  getopt_return_val: integer; option_index: c_int_type; current_option: 0 .. n_options;
  begin (Define the option table 22*);
  repeat getopt_return_val \equiv getopt_long_only(argc, argv, ‘’, long_options, address_of(option_index));
    if getopt_return_val = -1 then
      begin do_nothing;
      end
    else if getopt_return_val = ‘?’ then
      begin usage(my_name);
      end
    else if argument_is(‘help’) then
      begin usage_help(POOLTYPE_HELP, nil);
      end
    else if argument_is(‘version’) then
      begin print_version_and_exit(‘This is POOLtype, Version 3.0’, nil, ‘D.E. Knuth’, nil);
      end;
  until getopt_return_val = -1; { Now optind is the index of first non-option on the command line. }
  if (optind + 1 \neq argc) then
    begin writeLn(stderr, my_name, ‘: Need exactly one file_argument.’); usage(my_name);
    end;
  pool_name \equiv extend_filename(cmdline(optind), ‘pool’);
  { Try opening the file here, to avoid printing the first 256 strings if they give a bad filename. }
  resetbin(pool_file, pool_name);
  end;
```

This code is used in section 2*.

Here are the options we allow. The first is one of the standard GNU options.

```pascal
(Define the option table 22*) \equiv
  current_option \equiv 0; long_options[current_option].name \equiv ‘help’;
  long_options[current_option].has_arg \equiv 0; long_options[current_option].flag \equiv 0;
  long_options[current_option].val \equiv 0; incr(current_option);
```

See also sections 23* and 24*.

This code is used in section 21*.

Another of the standard options.

```pascal
(Define the option table 22*) \equiv
  long_options[current_option].name \equiv ‘version’; long_options[current_option].has_arg \equiv 0;
  long_options[current_option].flag \equiv 0; long_options[current_option].val \equiv 0; incr(current_option);
```

An element with all zeros always ends the list.

```pascal
(Define the option table 22*) \equiv
  long_options[current_option].name \equiv 0; long_options[current_option].has_arg \equiv 0;
  long_options[current_option].flag \equiv 0; long_options[current_option].val \equiv 0;
```
**25** Index. Indications of system dependencies appear here together with the section numbers where each identifier is used.

The following sections were changed by the change file: 1, 2, 6, 10, 15, 18, 19, 20, 21, 22, 23, 24, 25.

- **help**: 22
- **version**: 23
- **abort**: 15, 19, 20
- **address_of**: 21
- **argv**: 21
- **argument_is**: 21
- **argv**: 2, 21
- **ASCII_code**: 4
- **ASCII_code**: 5, 6, 7
- **boolean**: 18
- **c_int_type**: 21
- **carriage_return**: 9, 17
- **char**: 6
- **character_set_dependencies**: 10, 17
- **chr**: 6, 7, 10, 11
- **cmdline**: 21
- **const_c_string**: 18
- **count**: 13, 14, 15, 16, 20
- **current_option**: 21, 22, 23, 24
- **deer**: 3
- **do_nothing**: 3, 21
- **eof**: 19, 20
- **eoln**: 20
- **extend_filename**: 21
- **false**: 19
- **first_text_char**: 6, 11
- **flag**: 22, 23, 24
- **getopt**: 21
- **getopt_long_only**: 21
- **getopt_return_val**: 21
- **getopt_struct**: 21
- **has_arg**: 22, 23, 24
- **i**: 6
- **incr**: 3, 16, 20, 22, 23
- **initialize**: 2, 15
- **integer**: 6, 12, 13, 21
- **invalid_code**: 9, 11
- **k**: 12
- **kpsie_set_program_name**: 2
- **l**: 12
- **last_text_char**: 6, 11
- **lc_hex**: 16
- **long_options**: 21, 22, 23, 24
- **m**: 12
- **my_name**: 1, 2, 21
- **n**: 12
- **n_options**: 21
- **name**: 21, 22, 23, 24
- **null_code**: 9

**optind**: 21
**option_index**: 21
**ord**: 7
**output**: 2
**parse_arguments**: 2, 21
**pool_file**: 2, 12, 18, 19, 20, 21
**pool_name**: 18, 21
**POOL_type**: 2
**POOLTYPE_HELP**: 21
**print_version_and_exit**: 21
**read**: 20
**readLn**: 20
**resetbin**: 21
**s**: 12
**stderr**: 15, 21
**strcmp**: 21

**system_dependencies**: 2, 6, 8, 10, 17
**The TeXbook**: 10, 17
**text_char**: 6, 7, 12, 18
**true**: 17, 20
**uexit**: 15
**usage**: 21
**usage_help**: 21
**val**: 22, 23, 24
**write**: 16, 20
**writeLn**: 15, 16, 20, 21
**xchr**: 7, 8, 10, 11, 16, 17, 20
**xord**: 7, 11, 20
**xsum**: 18, 19, 20
\( \langle \text{Character } k \text{ cannot be printed } 17 \rangle \) Used in section 16.
\( \langle \text{Define the option table } 22^*, 23^*, 24^* \rangle \) Used in section 21*.
\( \langle \text{Define } parse\_arguments\ 21^* \rangle \) Used in section 2*.
\( \langle \text{Globals in the outer block } 7, 12, 13, 18^* \rangle \) Used in section 2*.
\( \langle \text{Local variables for initialization } 6^* \rangle \) Used in section 2*.
\( \langle \text{Make the first 256 strings } 16 \rangle \) Used in section 15*.
\( \langle \text{Read one string, but abort if there are problems } 20^* \rangle \) Used in section 19*.
\( \langle \text{Read the other strings from the } POOL\text{ file, or give an error message and abort } 19^* \rangle \) Used in section 15*.
\( \langle \text{Set initial values of key variables } 8, 10^*, 11, 14 \rangle \) Used in section 2*.
\( \langle \text{Types in the outer block } 5 \rangle \) Used in section 2*. 