$\S 1$ GB_IO INTRODUCTION 1

1. Introduction. This is GB_IO, the input/output module used by all GraphBase routines to access data files. It doesn't actually do any output; but somehow 'input/output' sounds like a more useful title than just 'input'.

All files of GraphBase data are designed to produce identical results on almost all existing computers and operating systems. Each line of each file contains at most 79 characters. Each character is either a blank or a digit or an uppercase letter or a lowercase letter or a standard punctuation mark. Blank characters at the end of each line are "invisible"; that is, they have no perceivable effect. Hence identical results will be obtained on record-oriented systems that pad every line with blanks.

The data is carefully sum-checked so that defective input files have little chance of being accepted.

2. Changes might be needed when these routines are ported to different systems. Sections of the program that are most likely to require such changes are listed under 'system dependencies' in the index.

A validation program is provided so that installers can tell if GB_IO is working properly. To make the test, simply run test_io.

3. The external variable *io_errors* mentioned in the previous section will be set nonzero if any anomalies are detected. Errors won't occur in normal use of GraphBase programs, so no attempt has been made to provide a user-friendly way to decode the nonzero values that *io_errors* might assume. Information is simply gathered in binary form; system wizards who might need to do a bit of troubleshooting should be able to decode *io_errors* without great pain.

```
#define cant_open_file #1
                                /* bit set in io_errors if fopen fails */
#define cant_close_file #2
                               /* bit set if fclose fails */
                               /* bit set if the data file's first line isn't legit */
#define bad_first_line #4
#define bad_second_line #8
                                 /* bit set if the second line doesn't pass muster */
\#define\ \mathit{bad\_third\_line}\ \ \#10
                                 /* bit set if the third line is awry */
#define bad_fourth_line #20
                                  /* guess when this bit is set */
#define file_ended_prematurely
                                 #40 /* bit set if fgets fails */
                                   /* bit set if line is too long or '\n' is missing */
#define missing_newline #80
#define wrong_number_of_lines #100
                                         /* bit set if the line count is wrong */
#define wrong_checksum #200
                                    /* bit set if the checksum is wrong */
                                /* bit set if user tries to close an unopened file */
#define no_file_open #400
#define bad_last_line #800
                                 /* bit set if final line has incorrect form */
```

2 INTRODUCTION GB_IO §4

4. The C code for GB_IO doesn't have a main routine; it's just a bunch of subroutines to be incorporated into programs at a higher level via the system loading routine. Here is the general outline of gb_io.c:

```
\langle Header files to include 7 \rangle \langle Preprocessor definitions \rangle \langle External declarations 5 \rangle \langle Private declarations 8 \rangle \langle Internal functions 9 \rangle \langle External functions 12 \rangle
```

5. Every external variable is declared twice in this CWEB file: once for GB_IO itself (the "real" declaration for storage allocation purposes) and once in gb_io.h (for cross-references by GB_IO users).

```
\langle External declarations 5\rangle \equiv long io\_errors; /* record of anomalies noted by GB_IO routines */ This code is used in section 4.
```

```
6. ⟨gb_io.h 6⟩ ≡
⟨Header files to include 7⟩
extern long io_errors; /* record of anomalies noted by GB_IO routines */
See also sections 13, 16, 19, 21, 23, 25, 29, and 41.
```

7. We will stick to standard C-type input conventions. We'll also have occasion to use some of the standard string operations.

```
⟨Header files to include 7⟩ ≡
#include <stdio.h>
#ifdef SYSV
#include <string.h>
#else
#include <strings.h>
#endif
```

This code is used in sections 4 and 6.

 $\S 8$ GB_IO INPUTTING A LINE 3

8. Inputting a line. The GB_IO routines get their input from an array called *buffer*. This array is internal to GB_IO—its contents are hidden from user programs. We make it 81 characters long, since the data is supposed to have at most 79 characters per line, followed by newline and null.

```
\langle Private declarations 8\rangle \equiv static char buffer [81]; /* the current line of input */ static char *cur_pos = buffer; /* the current character of interest */ static FILE *cur_file; /* current file, or \Lambda if none is open */ See also sections 10, 11, and 33. This code is used in section 4.
```

9. Here's a basic subroutine to fill the buffer. The main feature of interest is the removal of trailing blanks. We assume that cur_file is open.

Notice that a line of 79 characters (followed by '\n') will just fit into the buffer, and will cause no errors. A line of 80 characters will be split into two lines and the $missing_newline$ message will occur, because of the way fgets is defined. A $missing_newline$ error will also occur if the file ends in the middle of a line, or if a null character ('\0') occurs within a line.

```
\langle \text{Internal functions } 9 \rangle \equiv
  static void fill_buf()
  { register char *p;
    if (\neg fgets(buffer, sizeof(buffer), cur_file)) {
      io\_errors = file\_ended\_prematurely;
      buffer[0] = more\_data = 0;
    if (p-- \equiv buffer \lor *p \neq `\n') {
      io\_errors \mid = missing\_newline;
    while (--p \ge buffer \land *p \equiv ' \sqcup '); /* move back over trailing blanks */
    *++p = '\n';
    *++p=0; /* newline and null are always present at end of line */
    cur\_pos = buffer; /* get ready to read buffer[0] */
  }
See also section 15.
This code is used in section 4.
```

4 CHECKSUMS GB_IO §10

10. Checksums. Each data file has a "magic number," which is defined to be

$$\left(\sum_{l} 2^{l} c_{l}\right) \bmod p.$$

Here p is a large prime number, and c_l denotes the internal code corresponding to the lth-from-last data character read (including newlines but not nulls).

The "internal codes" c_l are computed in a system-independent way: Each character c in the actual encoding scheme being used has a corresponding icode, which is the same on all systems. For example, the icode of '0' is zero, regardless of whether '0' is actually represented in ASCII or EBCDIC or some other scheme. (We assume that every modern computer system is capable of printing at least 95 different characters, including a blank space.)

We will accept a data file as error-free if it has the correct number of lines and ends with the proper magic number.

```
⟨Private declarations 8⟩ +≡

static char icode [256]; /* mapping of characters to internal codes */

static long checksum_prime = (1<sub>L</sub> ≪ 30) − 83;

/* large prime such that 2p + unexpected_char won't overflow */

static long magic; /* current checksum value */

static long line_no; /* current line number in file */

static long final_magic; /* desired final magic number */

static long tot_lines; /* total number of data lines */

static char more_data; /* is there data still waiting to be read? */
```

11. The *icode* mapping is defined by a single string, imap, such that character imap[k] has icode value k. There are 96 characters in imap, namely the 94 standard visible ASCII codes plus space and newline. If EBCDIC code is used instead of ASCII, the cents sign $\not e$ should take the place of single-left-quote \lq , and \lnot should take the place of \lq .

All characters that don't appear in *imap* are given the same *icode* value, called *unexpected_char*. Such characters should be avoided in GraphBase files whenever possible. (If they do appear, they can still get into a user's data, but we don't distinguish them from each other for checksumming purposes.)

The *icode* table actually plays a dual role, because we've rigged it so that codes 0–15 come from the characters "0123456789ABCDEF". This facilitates conversion of decimal and hexadecimal data. We can also use it for radices higher than 16.

§12 GB_IO CHECKSUMS 5

```
Users of GB_IO can look at the imap, but they can't change it.
\langle \text{External functions } 12 \rangle \equiv
  char imap\_chr(d)
       long d;
    return d < 0 \lor d > strlen(imap)? '\0': imap[d];
  long imap\_ord(c)
       char c;
    \langle Make sure that icode has been initialized 14\rangle;
    return (c < 0 \lor c > 255)? unexpected\_char : icode[c];
  }
See also sections 17, 18, 20, 22, 24, 26, 30, 32, 39, and 42.
This code is used in section 4.
13. \langle gb_io.h 6 \rangle + \equiv
#define unexpected_char
  extern char imap_chr();
                                  /* the character that maps to a given character */
  extern long imap_ord();
                                  /* the ordinal number of a given character */
     \langle Make sure that icode has been initialized 14\rangle \equiv
  if (\neg icode[',1']) icode\_setup();
This code is used in sections 12 and 30.
15. \langle \text{Internal functions } 9 \rangle + \equiv
  static void icode_setup()
  \{ \text{ register long } k; 
    register char *p;
    for (p = imap, k = 0; *p; p++, k++) icode[*p] = k;
  }
16. Now we're ready to specify some external subroutines that do input. Calling gb\_newline() will read
the next line of data into buffer and update the magic number accordingly.
\langle gb\_io.h 6 \rangle + \equiv
                                 /* advance to next line of the data file */
  extern void gb\_newline();
  extern long new_checksum(); /* compute change in magic number */
17. Users can compute checksums as gb\_newline does, but they can't change the (private) value of magic.
\langle \text{External functions } 12 \rangle + \equiv
  long new_checksum(s, old_checksum)
       char *s;
                    /* a string */
       \mathbf{long}\ \mathit{old\_checksum};
  { register long a = old\_checksum;
    register char *p;
    for (p = s; *p; p++) a = (a + a + imap\_ord(*p)) \% checksum\_prime;
    return a;
  }
```

6 CHECKSUMS GB_IO §18

```
18. The magic checksum is not affected by lines that begin with *.
\langle External functions 12\rangle + \equiv
  void gb_newline()
     if (++line\_no > tot\_lines) more\_data = 0;
     if (more_data) {
       fill_buf();
       if (buffer[0] \neq """) magic = new\_checksum(buffer, magic);
     }
  }
19. Another simple routine allows a user to read (but not write) the variable more_data.
\langle \texttt{gb\_io.h} \ 6 \rangle + \equiv
  extern long gb\_eof();
                                 /* has the data all been read? */
20. \langle External functions 12\rangle +\equiv
  \mathbf{long}\ gb\_eof()
     \textbf{return} \ \neg more\_data;
```

 $\S21$ GB.IO PARSING A LINE 7

21. Parsing a line. The user can input characters from the buffer in several ways. First, there's a basic $gb_char()$ routine, which returns a single character. The character is '\n' if the last character on the line has already been read (and it continues to be '\n' until the user calls $gb_newline$).

The current position in the line, cur_pos , always advances when gb_char is called, unless cur_pos was already at the end of the line. There's also a $gb_backup()$ routine, which moves cur_pos one place to the left unless it was already at the beginning.

```
\( \text{gb_io.h 6} \) +\equiv \( \text{extern char } gb_char(); \) /* get next character of current line, or '\n' */
\( \text{extern void } gb_backup(); \) /* move back ready to scan a character again */

22. \( \text{External functions } 12 \rangle +\equiv \)
\( \text{char } gb_char() \\
\{ \quad if (*cur_pos) return (*cur_pos++); \quad return '\n'; \\
\} \text{void } gb_backup() \\
\{ \quad if (cur_pos > buffer) cur_pos--; \\
\}
\)
\( \text{external functions } 12 \rangle +\equiv \)
\( \text{char } gb_char() \\
\{ \quad if (cur_pos > buffer) cur_pos--; \\
\}
\)
\( \text{external functions } 12 \rangle +\equiv \)
\( \text{char } gb_char() \\
\{ \quad if (cur_pos > buffer) cur_pos--; \\
\}
\]
\( \text{external functions } 12 \rangle +\equiv \)
\( \text{extern } \text{external functions } 12 \rangle +\equiv \)
\( \text{extern } \text{external functions } 12 \rangle +\equiv \)
\( \text{extern } \text{external functions } 12 \rangle +\equiv \)
\( \text{external functions } 12 \rangle +\equiv \)
\( \text{extern } \text{external functions } 12 \rangle +\equiv \)
\( \text{extern } \text{external functions } 12 \rangle +\equiv \)
\( \text{extern } \text{external functions } 12 \rangle +\equiv \)
\( \text{extern } \text{external functions } 12 \rangle +\equiv \)
\( \text{extern } \text{external functions } 12 \rangle +\equiv \)
\( \text{external fun
```

23. There are two ways to read numerical data. The first, $gb_digit(d)$, expects to read a single character in radix d, using icode values to specify digits greater than 9. (Thus, for example, 'A' represents the hexadecimal digit for decimal 10.) If the next character is a valid d-git, cur_pos moves to the next character and the numerical value is returned. Otherwise cur_pos stays in the same place and -1 is returned.

The second routine, $gb_number(d)$, reads characters and forms an unsigned radix-d number until the first non-digit is encountered. The resulting number is returned; it is zero if no digits were found. No errors are possible with this routine, because it uses **unsigned long** arithmetic.

```
\langle gb\_io.h 6 \rangle +\equiv extern long gb\_digit(); /* gb\_digit(d) reads a digit between 0 and d-1 */ extern unsigned long gb\_number(); /* gb\_number(d) reads a radix-d number */
```

24. The value of d should be at most 127, if users want their programs to be portable, because C does not treat larger **char** values in a well-defined manner. In most applications, d is of course either 10 or 16.

```
 \begin{array}{lll} \langle \text{ External functions } 12 \rangle + \equiv & & & & & & & & \\ & & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & \\ & & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & \\ & & & \\ & & \\ & & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\
```

8 Parsing a line gb_io $\S25$

25. The final subroutine for fetching data is $gb_string(p, c)$, which stores a null-terminated string into locations starting at p. The string starts at cur_pos and ends just before the first appearance of character c. If $c \equiv '\n'$, the string will stop at the end of the line. If c doesn't appear in the buffer at or after cur_pos , the last character of the string will be the '\n' that is always inserted at the end of a line, unless the entire line has already been read. (If the entire line has previously been read, the empty string is always returned.) After the string has been copied, cur_pos advances past it.

In order to use this routine safely, the user should first check that there is room to store up to 81 characters beginning at location p. A suitable place to put the result, called str_buf , is provided for the user's convenience.

The location following the stored string is returned. Thus, if the stored string has length l (not counting the null character that is stored at the end), the value returned will be p + l + 1.

```
\langle gb_io.h 6 \rangle + \equiv
#define STR_BUF_LENGTH 160
  extern char str_buf[];
                                /* safe place to receive output of gb_string */
                                /* gb\_string(p,c) reads a string delimited by c into bytes starting at p */
  extern char *gb\_string();
      \#define STR_BUF_LENGTH 160
\langle \text{External functions } 12 \rangle + \equiv
                                        /* users can put strings here if they wish */
  char str_buf [STR_BUF_LENGTH];
  char *qb\_string(p, c)
                    /* where to put the result */
       char *p;
                   /* character following the string */
    while (*cur\_pos \land *cur\_pos \neq c) *p++ = *cur\_pos++;
    *p++=0;
    return p;
```

 $\S27$ GB_IO PARSING A LINE 9

27. Here's how we test those routines in test_io: The first line of test data consists of 79 characters, beginning with 64 zeroes and ending with '123456789ABCDEF'. The second line is completely blank. The third and final line says 'Oops:(intentional mistake)'.

```
\langle Test the sample data lines; exit if there's trouble 27\rangle \equiv
  if (gb\_number(10) \neq 123456789) io\_errors |= 1_L \ll 20;
                                                                            /* decimal number not working */
  \textbf{if} \ (gb\_digit(16) \neq 10) \ \textit{io\_errors} \ |= 1_L \ll 21; \qquad /* \ \text{we missed the A following the decimal number} \ */
                                      /* get set to read '9A' again */
   gb\_backup(); gb\_backup();
  if (gb\_number(16) \neq \text{\#9ABCDEF}) io\_errors = 1_L \ll 22;
                                                                           /* hexadecimal number not working */
  gb_newline(); /* now we should be scanning a blank line */
  if (gb\_char() \neq \verb"'\n") io\_errors \mid = 1_L \ll 23; /* newline not inserted at end */
   \begin{array}{ll} \textbf{if} \ (gb\_char(\ ) \neq \ \verb'\n'') \ \textit{io\_errors} \mid = 1_L \ll 24; & /* \ \text{newline not implied after end } */ \\ \textbf{if} \ (gb\_number(60) \neq 0) \ \textit{io\_errors} \mid = 1_L \ll 25; & /* \ \text{number should stop at null character } */ \\ \end{array} 
  \{ \mathbf{char} \ temp[100]; 
     if (gb\_string(temp, `\n') \neq temp + 1) io\_errors |= 1_L \ll 26;
           /* string should be null after end of line */
     if (gb\_string(temp, ':') \neq temp + 5 \lor strcmp(temp, "Oops")) io\_errors = 1_L \ll 27;
           /* string not read properly */
  }
  if (io\_errors) exit\_test("Sorry, \_it_\_failed._\Look_\_at_\_the_\_error_\_code_\_for_\_clues");
  if (qb\_digit(10) \neq -1) exit\_test("Digit\_error\_not\_detected");
  if (gb\_char() \neq ":") io_errors |= 1_L \ll 28; /* lost synch after gb\_string and gb\_digit */
  if (gb\_eof()) io_errors |= 1_L \ll 29; /* premature end-of-file indication */
  qb\_newline();
  if (¬gb_eof()) io_errors |= 1<sub>L</sub> \ll 30; /* postmature end-of-file indication */
This code is used in section 2.
```

10 OPENING A FILE GB_IO $\S 28$

28. Opening a file. The call $gb_raw_open("foo")$ will open file "foo" and initialize the checksumming process. If the file cannot be opened, io_errors will be set to $cant_open_file$, otherwise io_errors will be initialized to zero.

The call $gb_open("foo")$ is a stronger version of gb_raw_open , which is used for standard GraphBase data files like "words.dat" to make doubly sure that they have not been corrupted. It returns the current value of io_errors , which will be nonzero if any problems were detected at the beginning of the file.

```
\langle Test the gb_open routine; exit if there's trouble 28 \rangle \equiv
  if (gb\_open("test.dat") \neq 0) exit\_test("Can't\_open\_test.dat");
This code is used in section 2.
      #define gb\_raw\_open gb\_r\_open
                                                  /* abbreviation for Procrustean external linkage */
\langle gb_io.h 6 \rangle + \equiv
#define gb_raw_open gb_r_open
                                         /* open a file for GraphBase input */
  extern void gb\_raw\_open();
  extern long gb\_open();
                                   /* open a GraphBase data file; return 0 if OK */
     \langle \text{External functions } 12 \rangle + \equiv
  void gb\_raw\_open(f)
       char *f;
     \langle Make sure that icode has been initialized 14\rangle;
     \langle \text{Try to open } f \text{ 31} \rangle;
     if (cur_file) {
       io\_errors = 0;
       more\_data = 1;
       line\_no = magic = 0;
       tot\_lines = #7fffffff;
                                       /* allow "infinitely many" lines */
       fill_buf();
     } else io\_errors = cant\_open\_file;
```

31. Here's a possibly system-dependent part of the code: We try first to open the data file by using the file name itself as the path name; failing that, we try to prefix the file name with the name of the standard directory for GraphBase data, if the program has been compiled with DATA_DIRECTORY defined.

 $\S32$ GB.IO OPENING A FILE 11

```
32. \langle External functions 12\rangle + \equiv
  long gb\_open(f)
       char *f;
     strncpy(file\_name, f, sizeof (file\_name) - 1);
                                                        /* save the name for use by gb\_close */
     qb\_raw\_open(f);
     if (cur_file) {
         Check the first line; return if unsuccessful 34);
        Check the second line; return if unsuccessful 35);
        Check the third line; return if unsuccessful 36);
       (Check the fourth line; return if unsuccessful 37);
       gb\_newline();
                         /* the first line of real data is now in the buffer */
     \mathbf{return}\ io\_errors;
  }
     \langle Private declarations 8 \rangle + \equiv
  static char file_name [20]; /* name of the data file, without a prefix */
```

- **34.** The first four lines of a typical data file should look something like this:
 - * File "words.dat" from the Stanford GraphBase (C) 1993 Stanford University
 - * A database of English five-letter words
 - * This file may be freely copied but please do not change it in any way!
 - * (Checksum parameters 5757,526296596)

We actually verify only that the first four lines of a data file named "foo" begin respectively with the characters

```
* File "foo"

*

*

* (Checksum parameters l,m)
```

where l and m are decimal numbers. The values of l and m are stored away as tot_lines and $final_magic$, to be matched at the end of the file.

```
⟨ Check the first line; return if unsuccessful 34⟩ ≡
    sprintf(str_buf, "*∟File_\\"%s\"", f);
    if (strncmp(buffer, str_buf, strlen(str_buf))) return (io_errors |= bad_first_line);
This code is used in section 32.

35. ⟨ Check the second line; return if unsuccessful 35⟩ ≡
    fill_buf();
    if (*buffer ≠ '*') return (io_errors |= bad_second_line);
This code is used in section 32.
```

```
36. ⟨Check the third line; return if unsuccessful 36⟩ ≡ fill_buf();
if (*buffer ≠ '*') return (io_errors |= bad_third_line);
This code is used in section 32.
```

12 OPENING A FILE GB_IO §37

```
37. 〈Check the fourth line; return if unsuccessful 37〉 \equiv fill\_buf(); if (strncmp(buffer, "*_{\square}(Checksum\_parameters_{\square}", 23)) return (io\_errors \mid = bad\_fourth\_line); cur\_pos += 23; tot\_lines = gb\_number(10); if (gb\_char() \neq `, `) return (io\_errors \mid = bad\_fourth\_line); final\_magic = gb\_number(10); if (gb\_char() \neq `) `) return (io\_errors \mid = bad\_fourth\_line); This code is used in section 32.
```

§38 GB_IO CLOSING A FILE 13

38. Closing a file. After all data has been input, or should have been input, we check that the file was open and that it had the correct number of lines, the correct magic number, and a correct final line. The subroutine gb_close , like gb_open , returns the value of io_errors , which will be nonzero if at least one problem was noticed.

```
\langle Test the qb\_close routine; exit if there's trouble 38\rangle \equiv
  if (gb\_close() \neq 0) exit\_test("Bad\_checksum,\_or\_difficulty\_closing\_the\_file");
This code is used in section 2.
39. \langle External functions 12\rangle + \equiv
  long gb\_close()
     if (\neg cur\_file) return (io\_errors |= no\_file\_open);
     fill_buf();
     sprintf(str\_buf, "*_{\square}End_{\square}of_{\square}file_{\square} \"%s"", file\_name);
     if (strncmp(buffer, str_buf, strlen(str_buf))) io\_errors |= bad\_last\_line;
                                       /* now the GB_IO routines are effectively shut down */
     more\_data = buffer[0] = 0;
        /* we have cur_pos = buffer */
     if (fclose(cur\_file) \neq 0) return (io\_errors |= cant\_close\_file);
     cur_{-}file = \Lambda;
     if (line\_no \neq tot\_lines + 1) return (io\_errors |= wrong\_number\_of\_lines);
     if (magic \neq final\_magic) return (io\_errors |= wrong\_checksum);
     return io_errors;
```

40. There is also a less paranoid routine, gb_raw_close , that closes user-generated files. It simply closes the current file, if any, and returns the value of the magic checksum.

Example: The $restore_graph$ subroutine in GB_SAVE uses gb_raw_open and gb_raw_close to provide system-independent input that is almost as foolproof as the reading of standard GraphBase data.

}

14 INDEX $GB_{-}IO$ §43

43. Here is a list that shows where the identifiers of this program are defined and used.

a: 17, 24. $bad_first_line: 3, 34.$ $bad_fourth_line: 3, 37.$ bad_last_line : 3, 39. $bad_second_line: 3, 35.$ $bad_third_line: \underline{3}, 36.$ buffer: 8, 9, 16, 18, 22, 34, 35, 36, 37, 39, 42. c: 12, 26. $cant_close_file: 3, 39.$ $cant_open_file: 3, 28, 30.$ checksum_prime: 10, 17. cur_file: 8, 9, 30, 31, 32, 39, 42. cur_pos: 8, 9, 21, 22, 23, 24, 25, 26, 37, 39, 42. d: 12, 24.DATA_DIRECTORY: 31. $exit_test\colon \ \ \underline{2},\ 27,\ 28,\ 38.$ $f: \ \underline{30}, \ \underline{32}.$ fclose: 3, 39, 42. fgets: 3, 9.file_ended_prematurely: 3, 9. file_name: 32, 33, 39. fill_buf: 9, 18, 30, 35, 36, 37, 39. final_magic: 10, 34, 37, 39. fopen: 3, 31. fprintf: 2. gb_backup : $\underline{21}$, $\underline{22}$, 27. gb_char : 21, 22, 27, 37. gb_close : 32, 38, <u>39</u>, <u>41</u>. gb_digit : $\underline{23}$, $\underline{24}$, 27. qb_eof : $\underline{19}$, $\underline{20}$, $\underline{27}$. $gb_newline \colon \ \ \underline{16}, \ 17, \ \underline{18}, \ 21, \ 27, \ 32.$ gb_number : 23, 24, 27, 37.

 qb_r_close : 41. qb_r_open : 29.

k: 15.

main: 2.

 $missing_newline: \underline{3}, 9.$

more_data: 9, 10, 18, 19, 20, 30, 39, 42.

 gb_open : 28, 29, 32, 38. qb_raw_close : 40, 41, 42. gb_raw_open : 28, <u>29</u>, <u>30</u>, 32, 40. gb_string : $\underline{25}$, $\underline{26}$, 27. $icode: \ \underline{10},\ 11,\ 12,\ 14,\ 15,\ 23,\ 24.$ $icode_setup: 14, \underline{15}.$ imap: 11, 12, 15. $imap_chr$: 12, 13. imap_ord: 12, 13, 17, 24. io_errors: 2, 3, 5, 6, 9, 27, 28, 30, 32, 34, 35, 36, 37, 38, 39. $line_no: 10, 18, 30, 39.$ magic: 10, 17, 18, 30, 39, 40, 42.

 $new_checksum: 16, 17, 18.$ $no_file_open: \underline{3}, 39.$ $old_checksum: 17.$ p: 9, 15, 17, 26. print f: 2. $restore_graph$: 40. s: 17.sprintf: 31, 34, 39. stderr: 2. $str_buf: 25, 26, 31, 34, 39.$ STR_BUF_LENGTH: 25, 26, 31. strcmp: 27.strlen: 12, 31, 34, 39. strncmp: 34, 37, 39. strncpy: 32. system dependencies: 31. SYSV: 7. temp: 27.tot_lines: 10, 18, 30, 34, 37, 39. unexpected_char: 10, 11, 12, 13, 15. $wrong_checksum: 3, 39.$ $wrong_number_of_lines: 3, 39.$

```
(Check the first line; return if unsuccessful 34) Used in section 32.
 Check the fourth line; return if unsuccessful 37 \rangle Used in section 32.
 Check the second line; return if unsuccessful 35 \rangle Used in section 32.
 Check the third line; return if unsuccessful 36 \rangle Used in section 32.
 External declarations 5 \ Used in section 4.
 External functions 12, 17, 18, 20, 22, 24, 26, 30, 32, 39, 42 \rightarrow Used in section 4.
 Header files to include 7 Used in sections 4 and 6.
 Internal functions 9, 15 \ Used in section 4.
 Make sure that icode has been initialized 14 \rangle Used in sections 12 and 30.
 Private declarations 8, 10, 11, 33 \ Used in section 4.
 Test the sample data lines; exit if there's trouble 27 \rangle Used in section 2.
 Test the gb\_close routine; exit if there's trouble 38 \rangle Used in section 2.
(Test the gb\_open routine; exit if there's trouble 28) Used in section 2.
\langle \text{Try to open } f \text{ 31} \rangle Used in section 30.
 gb_io.h 6, 13, 16, 19, 21, 23, 25, 29, 41 >
\langle \text{test\_io.c} 2 \rangle
```

$GB_{-}IO$

	Section	- 0
Introduction	1	1
Inputting a line	8	•
Checksums	10	4
Parsing a line	21	7
Opening a file	28	1(
Closing a file	38	13
Index	43	14

© 1993 Stanford University

This file may be freely copied and distributed, provided that no changes whatsoever are made. All users are asked to help keep the Stanford GraphBase files consistent and "uncorrupted," identical everywhere in the world. Changes are permissible only if the modified file is given a new name, different from the names of existing files in the Stanford GraphBase, and only if the modified file is clearly identified as not being part of that GraphBase. (The CWEB system has a "change file" facility by which users can easily make minor alterations without modifying the master source files in any way. Everybody is supposed to use change files instead of changing the files.) The author has tried his best to produce correct and useful programs, in order to help promote computer science research, but no warranty of any kind should be assumed.

Preliminary work on the Stanford Graph Base project was supported in part by National Science Foundation grant CCR-86-10181.