#### A First Book of ANSI C Fourth Edition

#### Chapter 9 Character Strings

## Objectives

- String Fundamentals
- Library Functions
- Input Data Validation
- Formatting Strings (Optional)
- Case Study: Character and Word Counting
- Common Programming and Compiler Errors

- On a fundamental level, <u>strings are simply</u> <u>arrays of characters</u> that can be manipulated using standard element-by-element arrayprocessing techniques.
- On a higher level, string library functions are available for treating strings as complete entities.
- This chapter explores the input, manipulation, and output of strings using both approaches.
- We will also examine the particularly close connection between string-handling functions and pointers.

# 9.1 String Fundamentals

- A string literal is any sequence of characters enclosed in double quotes.
- A string literal is also referred to as a string constant and string value, and more conventionally as a string.
- For example, "This is a string", "HelloWord!", and "xyz123\*!#@&" are all strings.

- Because a string is stored as an array of characters, the individual characters in the array can be input, manipulated, or output using standard array-handling techniques utilizing either subscript or pointer notations.
- <u>The end-of-string null character is useful</u> <u>for detecting the end of the string</u> when handing strings in this fashion

# String Input and Output

 Table 9.1 lists the commonly available library functions for both character-by-character and complete string input and output.

#### Table 9.1

Input gets() scanf() getchar() Output puts() printf() putchar()

# Example of String Input and Output

Program 9.1

illustrates the use of gets() and puts() to input and output a string entered at the user's terminal.

```
#include <stdio.h>
   int main()
 2
 3
   {
     #define MSIZE 81
 4
     char message[MSIZE]; /* enough storage for 80 characters plus '\0' */
 5
 б
     printf("Enter a string:\n");
 7
     gets(message);
 8
 9
     printf("The string just entered is:\n");
10
     puts(message);
11
12
     return 0;
13 }
               Sample run:
               Enter a string:
               This is a test input of a string of characters.
               The string just entered is:
               This is a test input of a string of characters.
```

- The <u>gets() function</u> used in Program 9.1 <u>continuously accepts and stores the characters</u> <u>typed at the terminal into the character array</u> <u>named message</u>.
- Pressing the Enter key at the terminal generates a newline character, <u>\n</u>, which is <u>interpreted by</u> <u>gets() as the end-of-character entry</u>.
- <u>All the characters encountered by gets(), except</u> the newline character, are stored in the message array.

- Before returning, the <u>gets() function</u>
   <u>appends the null character to the stored set</u>
   <u>of characters</u>, as illustrated in Figure 9.2a.
- The puts()function is then used to display the string.
- The <u>scanf()</u> function <u>reads a set of</u> <u>characters up to either a blank space or a</u> <u>newline character</u>, whereas <u>gets() stops</u> <u>accepting characters only when a newline is</u> <u>detected</u>.

- Trying to enter the characters This is a string using the statement scanf("%s", message); results in the word This being assigned to the message array.
- Entering the complete line using a scanf() function call would require a statement such as
- scanf("%s %s %s %s", message1, message2, message3, message4);

 This allows us to understand how the standard library functions are constructed and to create our own library functions.

 For a specific example, consider the function strcopy(), which copies the contents of string2 to string1.

```
void strcopy(char string1[],
              char string2[])
   // i will be used as a subscript
   int i=0;
   while (string2[i]!='\0')
        string1[i]=string2[i];
        i++;
   // terminate the first string
   string1[i]='\0';
```

```
Program 9.2: includes the strcopy() function in a complete program
```

```
#include<stdio.h>
 /*expects two arrays of chars */
void strcopy(char[],char[]);
int main()
```

```
// enough storage for a complete line
char message[81];
// enough storage for a copy of message
char newMessage[81];
int i;
printf(" Enter a sentence: ");
gets(message);
strcopy(newMessage, message);
/*pass two array addresses*/
```

```
puts(newMessage);
```

```
return 0;
```

```
/* copy string2 to string1 */
/*two arrays are passed */
void strcopy(char string1[],
               char string2[])
    int i=0; // i will be used as a subscript
    /* check for the end-of-string */
    while( string2[i]!='0^{\circ} )
           /* copy the element to string 1*/
          string1[i]=string2[i];
          i++;
    /* teminate the first string */
    string1[i]='\0';
```

Example Output Enter a sentence: I am a CS CMU student. I am a CS CMU student.

# **Character-by-Character Input**



#### Program 9.3

```
#include <stdio.h>
                                รับค่า string ทีละ character ผ่าน function ชื่อ
    int main()
 2
                                getchar() จนกระทั่ง user เคาะ enter.
 3
    {
      #define LSIZE 81
 4
                              /* enough storage for 80 characters plus '\0' */
 5
      char message[LSIZE];
      char c;
 6
 7
      int i;
                                           ข้อควรระวัง: ถ้าลืมใส่วงเล็บ จะมีค่าเท่ากับ
 8
      printf("Enter a string:\n");
 9
                                           c = (getchar() != ! \n')
10
      i = 0;
11
      while(i < (LSIZE-1) & (c = qetchar()) != '\n')
12
       {
13
        message[i] = c; /* store the character entered */
14
        i++;
15
       }
16
      message[i] = '\0'; /* terminate the string */
17
      printf("The string just entered is: \n");
18
      puts(message);
19
20
      return 0;
21
    }
```

#### String Processing (continued)

Program 9.4

```
1
    #include <stdio.h>
   void getline(char []); /* function prototype */
 2
   #define LSIZE 81
 3
 4
 5
    int main()
 6
    {
 7
      char message[LSIZE]; /* enough storage for 80 characters plus '\0' */
 8
 9
      printf("Enter a string: \n");
      getline(message);
10
      printf("The string just entered is:\n");
11
12
     puts(message);
13
                                   สร้าง function ชื่อ getline() เพื่อใช้รับค่า string ทีละ
14
      return 0;
15 }
                                   character ผ่าน function ชื่อ getchar() จนกระทั่ง user
16
                                   เคาะ enter.
   void getline(char strng[])
17
18
19
      int i = 0;
20
      char c;
21
      while(i < (LSIZE-1) & (c = getchar()) != '\n')
22
23
      {
        strng[i] = c; /* store the character entered */
24
25
        i++;
26
      }
      strng[i] = '\0'; /* terminate the string */
27
28
   }
```

# Library Functions

Table 9.2 String Library Routines (Required Header File is string.h)

Name	Description	Example
strcpy(str1, str2)	Copies str2 to str1, including the '\0'	<pre>strcpy(test, "efgh")</pre>
strcat(str1, str2)	Appends str2 to the end of str1	strcat(test, "there")
strlen(string)	Returns the length of string. Does not include the '\0' in the length count.	strlen("Hello World!")
strcmp(str1, str2)	Compares str1 to str2. Returns a negative integer if str1 < str2, 0 if str1 == str2, and a positive integer if str1 > str2.	strcmp("Beb", "Bee")

Note: Attempting to copy a larger string into a smaller string causes the copy to overflow the destination array beginning with the memory area immediately following the last array element.

- When comparing strings, their individual characters are evaluated in pairs; if a difference is found, the string with the first lower character is the smaller one
  - "Good Bye" is less than "Hello" because the first
     'G' in Good Bye is less than the first 'H' in Hello
  - "Hello" is less than "Hello " because the '\0' terminating the first string is less than the ' ' in the second string
  - "123" is greater than "122" because '3' in 123 is greater than '2' in 122
  - "1237" is greater than "123" because '7' in 1237 is greater than '\0' in 123



#### Program 9.5

```
#include <stdio.h>
 1
    #include <string.h> /* required for the string function library */
 2
 3
    int main()
 4
 5
    {
      #define MAXELS 50
 6
 7
      char string1[MAXELS] = "Hello";
      char string2[MAXELS] = "Hello there";
 8
 9
      int n;
10
11
      n = strcmp(string1, string2);
12
      if (n < 0)
13
14
        printf("%s is less than %s\n\n", string1, string2);
15
      else if (n == 0)
        printf("%s is equal to %s\n\n", string1, string2);
16
17
      else
18
        printf("%s is greater than %s\n\n", string1, string2);
19
20
      printf("The length of string1 is %d characters\n", strlen(string1));
A First Book of ANSI C, Fourth Edition
```

```
printf("The length of string2 is %d characters\n\n", strlen(string2));
21
22
23
      strcat(string1," there World!");
24
25
     printf("After concatenation, string1 contains the string value\n");
26
     printf("%s\n", string1);
27
     printf("The length of this string is %d characters\n\n",
28
                                                        strlen(string1));
29
      printf("Type in a sequence of characters for string2:\n");
30
      gets(string2);
31
32
      strcpy(string1, string2);
33
34
     printf("After copying string2 to string1");
35
     printf(" the string value in string1 is:\n");
36
     printf("%s\n", string1);
37
      printf("The length of this string is %d characters\n\n",
38
                                                      strlen(string1));
39
     printf("\nThe starting address of the string1 string is: %d\n",
40
                                                      (void *) string1);
41
      return 0;
42
   }
```

• Sample output:

Hello is less than Hello there

The length of string1 is 5 characters The length of string2 is 11 characters

```
After concatenation, string1 contains the string value
Hello there World!
The length of this string is 18 characters
```

Type in a sequence of characters for string2: It's a wonderful day After copying string2 to string1, the string value in string1 is: It's a wonderful day The length of this string is 20 characters

The starting address of the string1 string is: 1244836

#### **Character Routines**

Table 9.3 Character Library Routines (Required Header File is ctype.h)

Required Prototype	Description	Example
int isalpha(char)	Returns a non-0 number if the character is a letter; otherwise, it returns 0.	isalpha('a')
int isupper(char)	Returns a non-0 number if the character is uppercase; otherwise, it returns 0.	isupper('a')
int islower(char)	Returns a non-0 number if the character is lowercase; otherwise, it returns 0.	islower('a')
int isdigit(char)	Returns a non-0 number if the character is a digit (0 through 9); otherwise, it returns 0.	isdigit('a')
int isascii(char)	Returns a non-0 number if the character is an ASCII character; otherwise, it returns 0.	isascii('a')
int isspace(char)	Returns a non-0 number if the character is a space; otherwise, it returns 0.	isspace(' ')
int isprint(char)	Returns a non-0 number if the character is a printable character; otherwise, it returns 0.	isprint('a')
int iscntrl(char)	Returns a non-0 number if the character is a control character; otherwise, it returns 0.	iscntrl('a')
int ispunct(char)	Returns a non-0 number if the character is a punctuation character; otherwise, it returns 0.	ispunct('!')
int toupper(char)	Returns the uppercase equivalent if the character is lowercase; otherwise, it returns the character unchanged.	toupper('a')
int tolower(char)	Returns the lowercase equivalent if the character is uppercase; otherwise, it returns the character unchanged.	tolower('A')

### Character Routines (continued)



#### Program 9.6

```
#include <stdio.h>
 1
    #include <ctype.h> /* required for the character function library */
 2
 3
    int main()
 4
 5
    {
      #define MAXCHARS 100
 б
 7
      char message [MAXCHARS];
 8
      void convertToUpper(char []); /* function prototype */
 9
10
      printf("\nType in any sequence of characters:\n");
11
      gets(message);
12
13
      convertToUpper(message);
14
15
      printf("The characters just entered, in uppercase are:\n%s\n", message);
16
      return 0;
17
18
   }
    // this function converts all lowercase characters to uppercase
19
    void convertToUpper(char message[])
20
21
    {
      int i;
22
      for(i = 0; message[i] != ' \setminus 0'; i++)
23
24
        message[i] = toupper(message[i]);
25 }
```

#### **Conversion Routines**

 Table 9.4
 Conversion Routines (Required Header File is stdlib.h)

Prototype	Description	Example
int atoi(string)	Converts an ASCII string to an integer. Conversion stops at the first noninteger character.	atoi("1234")
double atof(string)	Converts an ASCII string to a double- precision number. Conversion stops at the first character that cannot be interpreted as a double.	atof("12.34")
char[] itoa(string)	Converts an integer to an ASCII string. The space allocated for the returned string must be large enough for the converted value.	itoa(1234)

#### **Conversion Routines (continued)**



Program 9.7

```
#include <stdio.h>
 1
    #include <string.h>
 2
    #include <stdlib.h> // required for test conversion function library
 3
 4
 5
    int main()
 б
    {
    #define MAXELS 20
 7
      char test[MAXELS] = "1234";
 8
      int num;
 9
      double dnum;
10
11
12
      num = atoi(test);
13
      printf("The string %s as an integer number is %d\n", test,num);
      printf("This number divided by 3 is: %d\n", num/3);
14
15
16
      strcat(test, ".96");
17
      dnum = atof(test);
18
      printf("\nThe string %s as a double number is: %f\n", test,dnum);
19
      printf("This number divided by 3 is: %f\n", dnum/3);
20
21
22
      return 0;
23 }
```

### Input Data Validation

- Successful programs always try to anticipate invalid data and isolate such data from being accepted and processed
  - First validate that the data is of the correct type; if not, request the user to re-enter the data
  - Explain why the entered data was invalid
- One of the most common methods of validating input data is to accept all numbers as strings
  - Each character can then be checked to ensure that it complies with the data type being requested



#### Program 9.8

```
#include <stdio.h>
 1
  #include <stdlib.h> /* needed to convert a string to an integer */
 2
 3 #define MAXCHARS 40
 4 #define TRUE 1
 5 #define FALSE 0
 б
 7
   int isvalidInt(char []); /* function prototype */
 8
   int main()
 9
10
   {
11
12
      char value[MAXCHARS];
13
      int number;
14
15
      printf("Enter an integer: ");
16
      gets(value);
17
18
      if (isvalidInt(value) == TRUE)
19
      {
        number = atoi(value);
20
        printf("The number you entered is %d\n", number);
21
22
      }
23
      else
24
        printf("The number you entered is not a valid integer.\n");
25
      return 0;
26
27 }
28
```

```
29 int isvalidInt(char val[])
30 {
31
      int start = 0;
32
      int i;
     int valid = TRUE;
33
      int sign = FALSE;
34
35
36
     /* check for an empty string */
      if (val[0] == ' \setminus 0') valid = FALSE;
37
38
      /* check for a leading sign */
39
      if (val[0] == '-' || val[0] == '+')
40
41
      {
42
        sign = TRUE;
        start = 1; /* start checking for digits after the sign */
43
44
      }
45
46
      /* check that there is at least one character after the sign */
47
      if (sign == TRUE && val [1] == ' \setminus 0') valid = FALSE;
48
49
      /*now check the string, which we know has at least one non-sign char */
      i = start;
50
51
      while(valid == TRUE && val[i] != ' \setminus 0')
52
      {
        if (val[i] < '0' || val[i] > '9') /* check for a non-digit */
53
           valid = FALSE;
54
55
        i++;
56
      }
57
58
      return valid;
59 }
```

• We can use isvalidInt() in a loop that continually requests an integer until a valid integer value is entered

Set an integer variable named isanInt to 0

do

Accept a string value

If the string value does not correspond to an integer

Display the error message "Invalid integer - Please re-enter: "

Send control back to expression being tested by the do-while statement

Set isanInt to 1 (this causes the loop to terminate)

while(isanInt is 0)

Return the integer corresponding to the entered string



Program 9.9

```
17
    #define TRUE 1
    #define FALSE 0
18
    #define MAXCHARS 40
19
20
    int getanInt()
21
    {
22
      int isvalidInt(char []); /* function prototype */
23
24
      int isanInt = FALSE;
25
      char value [MAXCHARS];
26
27
      do
28
      {
        gets(value);
29
30
        if (isvalidInt(value) == FALSE)
31
        {
32
          printf("Invalid integer - Please re-enter: ");
33
          continue; /* send control to the do-while expression test */
34
35
        isanInt = TRUE;
36
      }while (isanInt == FALSE);
37
38
      return (atoi(value)); /* convert to an integer */
39
```

# Creating a Personal Library

- Programmers create their own libraries of functions
  - This permits the functions to be incorporated in any program without further expenditure of coding time
- Each file in a library contains related functions
  - #include <C:\\mylibrary\\dataChecks.h>
  - #include "C:\\mylibrary\\dataChecks.h"
    - The #include statement for dataChecks.h must be placed after the #include statements for the stdio.h and stdlib.h header files (the functions in dataChecks.h require stdio.h and stdlib.h functions to correctly compile)

### **Formatting Strings**

- Examples:
  - printf("|%25s|","Have a Happy Day"); • | Have a Happy Day|
  - -printf("|%-25s|","Have a Happy Day");
    - |Have a Happy Day
  - -printf("|%25.12s|","Have a Happy
    Day");
    - | Have a Happy|
  - -printf("|%.12s|","Have a Happy Day");
    - |Have a Happy|

### In-Memory String Conversions

- The sprintf() and sscanf() functions provide capabilities for writing and scanning strings to and from memory variables
  - sprintf(disStrn,"%d %d", num1, num2);
  - sscanf(data,"%c%lf %d",&dol,&price,&units);
    - "\$23.45 10"
  - sscanf(date,"%d/%d/%d", &month, &day, &year);
    - "07/01/94"

# Format Strings

- The control string containing the conversion control sequences need not be explicitly contained within the function
  - printf("\$%5.2f %d",num1,num2);
  - Or, char fmat[] = "\$%5.2f %d"; printf(fmat,num1,num2);
- Useful for listing format strings with other variable declarations at the beginning of a function
  - If you need to change a format, it is easy to find the desired control string without searching to locate the appropriate printf() or scanf() function calls

# Case Study: Character and Word Counting

- We construct two string-processing functions
  - Count the number of characters in a string
  - Count words in a sting
    - What constitutes a word?

# Program Requirement: Character Counting

- Pass a string to a function and have the function return the number of characters in the string
- Any character in the string (blank, printable, or nonprintable character) is to be counted
- The end-of-string NULL character is not to be included in the final count

#### Analyze the Problem

- Determine the input data
- Determine the required outputs
- List the algorithm(s) relating the inputs to the outputs

#### Analyze the Problem (continued)



Figure 9.5 Counting characters in a string

#### Code the Function

```
int countchar(char list[])
{
    int i, count = 0;
    for(i = 0; list[i] != '\0'; i++)
        count++;
    return(count);
}
```

# Test and Debug the Function



#### Program 9.10

```
#include <stdio.h>
 1
    #define MAXNUM 1000
 2
 3
    int countchar(char []); /* function prototype */
 4
 5
    int main()
 6
 7
    {
      char message [MAXNUM];
 8
 9
      int numchar;
10
      printf("\nType in any number of characters: ");
11
12
      gets(message);
13
      numchar = countchar(message);
14
     printf("The number of characters just entered is %d\n", numchar);
15
      return 0;
16
17
    }
18
   . . .
```

# Requirement Specification: Word Counting

- The last word does not have a trailing blank
- More than one blank may be used between words
- Leading blanks may be used before the first word

#### Analyze the Problem

- Determine the input data
- Determine the required outputs
- Algorithm:

Set an integer variable named inaword to the symbolic constant NO Set the word count to 0 For all the characters in the array If the current character is a blank set inaword to NO Else if (inaword equals NO) set inaword to the symbolic constant YES increment the word count EndIf EndFor Return the count

#### Code the Function

```
int countword(char list[])
#define YES 1
#define NO 0
{
  int i, inaword, count = 0;
  inaword = NO;
  for (i = 0; list[i] != '\0'; i++)
  {
    if (list[i] == ' ')
      inaword = NO;
    else if (inaword == NO)
      inaword = YES;
      count++;
    }
  return (count);
}
```

## Test and Debug the Function



#### Program 9.11

```
#include <stdio.h>
 1
    #define MAXNUM 1000
 2
 3
    int countword(char []); /* function prototype */
 4
 5
    int main()
 б
 7
   {
 8
      char message[MAXNUM];
      int numchar;
 9
10
11
      printf("\nType in any number of words: ");
12
      gets(message);
13
      numchar = countword(message);
14
      printf("The number of words just entered is %d\n", numchar);
15
16
      return 0;
17
   }
18
    . . .
```

# Test and Debug the Function (continued)

• A sample run using Program 9.11 follows:

Type in any number of words: This is a test line with a bunch of words The number of words just entered is 10

- Further tests that should be performed are
  - Enter words with multiple spaces between them
  - Enter words with leading spaces before the first word
  - Enter words with trailing spaces after the last word
  - Enter a sentence that ends in a period or question mark

# **Common Programming Errors**

- Forgetting the terminating NULL character, '\0', when processing existing strings in a character-bycharacter manner
- Forgetting to terminate a newly created character string with the NULL character
- Forgetting that the newline character, '\n', is a valid data input character
- Forgetting to include the string.h, ctype.h, and stdlib.h header files when using the string library, character library, and conversion library functions, respectively

## **Common Compiler Errors**

Error	Typical Unix-based Compiler Error Message	Typical Windows-based Compiler Error Message
Attempting to assign a single character into an ele- ment of the array using double, rather than single quotes. For example, message [5] = "A";	(W) Operation between types "unsigned char" and "unsigned char*" is not allowed.	error : cannot convert from 'const char [2]' to 'char'
Not using a system predefined constant in all capital letters. For example, message[10] = NULL;	(S) Undeclared identifier NULL.	error: 'NULL' : undeclared identifier
Forgetting to insert a length in the Size of the Array without initializers. For example, char message[];	(S) Explicit dimension specification or initializer required for an auto or static array.	error: 'message' : unknown size

# Common Compiler Errors (continued)

Error	Typical Unix-based Compiler Error Message	Typical Windows-based Compiler Error Message
Comparing against an escape sequence that is inside double quotes. For example, while((c = getchar()) != "\n")	(W) Operation between types "int" and "unsigned char*" is not allowed.	error: '!=' : no conversion from 'const char *' to 'int'
Providing an incorrect path for including header files. For example, #include "c:\\ stdio.h"	(S) #include file "c:\\stdio.h" not found.	fatal error: Cannot open include file: 'c:\\stdio.h': No such file or directory

# String & Pointer

```
#include <stdio.h>
int main()
{
    char *message2 = "this is a string";
```

```
printf("\nThe string is: %s", message2);
printf("\n The first address of this string is %p", message2);
```

```
message2 = "A new message";
printf("\nThe string is now: %s", message2);
printf("\n The first address of this string is %p", message2);
```



# String & Pointer (Cont.)



A First Book of ANSI C, Fourth Edition

# Pointer Arrays (Program)

```
#include <stdio.h>
```

int main()

•

int n;

```
char *seasons[] = { "Winter",
```

```
"Spring",
"Summer",
"Fall"};
```



for(n = 0; n < 4; ++n)

printf("\nThe season is %s.",seasons[n]);

return 0;

# Pointer Arrays



seasons[0] = "Winter";

seasons[1] = "Spring";

seasons[2] = "Summer";

seasons[3] = "Fall";

# Scaling a set of numbers into a more useful set

```
#include <stdio.h>
int main()
{
    int n;
    char *seasons[] = { "Winter",
        "Spring",
        "Summer",
        "Fall"};
```

printf("\nEnter a month (use 1 for Jan., 2 for Feb., etc.): "); scanf("%d", &n); n = (n % 12) / 3; /\* create the correct subscript \*/ printf("The month entered is a %s month.",seasons[n]);

```
return 0;
```

}

# Scaling a set of numbers into a more useful set (Cont.)

Months	Season
December, January, February	Winter
March, April, May	Spring
June, July, August	Summer
September, October, November	Fall



# Summary

- A string is an array of characters terminated by the NULL ('  $\0$  ') character
- Character arrays can be initialized using a string assignment of the form char arrayName[] = "text";
- Strings can always be processed using standard array-processing techniques
- The gets(), scanf(), and getchar() library functions can be used to input a string
- The puts(), printf(), and putchar() functions can be used to display strings

# Summary (continued)

- Many standard library functions exist for processing strings as a complete unit
- The standard C library also includes individual character-handling functions (ctype.h)
- One of the major uses of strings is validating user input, which is an essential part of any program
- The conversion routines atoi () and atof () are provided in the stdlib.h header file for converting strings to integer and double-precision numeric values