

1. ARRAYS, STRINGS AND POINTERS

1.1 Introduction to Strings

A string is a sequence of characters (array), with the last character being the NUL character (\0).

For example, the array of 6 characters: I n d i a \0 is a string of length 5 (the NUL character is not counted as part of the length). Had the same array been without the terminating \0, it would not be called a string, but simply a character array.

Because of this extra \0 character (whose ASCII value is 0) that signals the end of strings, they can be treated in special ways that normal arrays cannot be treated.

1.2 String variables, string constants and declarations

`char word[40];` declares an array of forty characters and can be treated as a string if a \0 is one of the elements.

Consider these three array declarations:

```
char arr[] = "Hello";
char brr[7] = {'H', 'e', 'l', 'l', 'o'};
char crr[] = {'H', 'e', 'l', 'l', 'o'};
```

The entity `arr` is a string, because the variable `arr` is being initialized with the string constant “Hello” at the time of declaration. (This is similar to saying `int i = 4;` insofar initializing an integer variable with an integer constant.)

Similarly, `brr` is also a string, for if even one element of an array is initialized (at the time of declaring the array), the remainder of the uninitialized elements are automatically set to 0.

The third array, `crr` is not a string, because there is no \0 assignment in the declaration, and only 5 bytes are set aside in memory. Hence, `crr` is just a character array.

The following three entities have different meanings: 4, '4', "4"
 4 is an integer constant, '4' is a character constant (with an ASCII value of 52), while "4" is a string constant.

1.3 Getting input into string variables and to printing out strings

The conversion specifier "%s" is used by both `scanf()` and `printf()` functions to get string input, and give string output, respectively. Consider this piece of code:

```
char word1[100];
printf("Enter the word: ");
scanf("%s", word1);
printf("The word you entered is: %s\n", word1);
```

Note that `scanf()` expects an argument that is an address (pointer). Since the name of the array is its base address, you don't need to have its name preceded with the address of operator (&).

The %s specifier in the `scanf()` function will take input until the first whitespace character, [i.e., a space, \n (newline) or \t (tab) character] is encountered. So, how to take a line of input?

```
scanf("%[^\\n]", word1);
```

By specifying the character class as the format specifier, where you state that all characters till \n (the caret symbol denotes negation of the character class) should be accepted, you can now take a line of input.

The character class can be used in `scanf()` in other usual ways such as %[A-Za-z0-9] to accept a string of only alphanumeric characters, %[aeiouAEIOU,] to accept a string of vowels and a comma, and so on. In such cases, the `scanf()` stops accepting input at the first instance of the condition being violated. While no further input is taken into the string, the remaining input continues to stay in the input buffer! This may cause problems with subsequent calls to an input function like `scanf()` and `getchar()`. It is important to keep this in mind when writing code.

For instance, when receiving input for a line of text using the statement `scanf("%[^\\n]", word1);` or even when using `scanf()` to get a numeric input, e.g., `scanf("%d", &N)`, the \n character entered in the keyboard continues to stay in the input buffer, available for the subsequent call to `scanf()` potentially causing problems. For example, an empty string gets stored in the array `word1` when a subsequent `scanf("%[^\\n]", word1);` statement is encountered. One way to remedy this issue is by including an input statement to eat away the residual \n still present in the input buffer stream:

```
scanf("%[^\\n]", word1);
getchar(); /* an empty getchar() to dissipate the \\n */
```

LABORATORY SESSION #8

Exercise 8-1: Write a program that accepts a line of text as a string from the user and prints out only the non-whitespace characters of it. You may assume a line contains at most 99 characters.

Next, modify the program so it works for N lines of text, where N is a number between 1 and 10 input by the user. You can take one line, process it and print the output before accepting the next line of input into the same array. **Do not** use a 2D array, but just a nested loop.

Omit the empty `getchar()` statement after the `scanf()` and observe what happens. You should understand how to handle the input buffer stream properly, otherwise your program is likely to show unexpected behaviour.

1.4 String Functions

The essence of treating a character array as a string is best put to use by string functions (available in the standard C library). Use of these functions requires the programmer to include the `<string.h>` header file in the program. The prototypes and use of some of the commonly used string library functions are given below:

`size_t strlen(const char *s); // size_t is long unsigned int on gcc`

Returns the length (number of bytes, excluding `\0`) of the string `s`.

`char *strcpy(char *dest, const char *src);`

Copies the string `src` to `dest`, returning a pointer to the start of `dest`.

`int strcmp(const char *s1, const char *s2);`

Compares the strings `s1` with `s2`. Returns 0 if `s1` is identical to `s2`, a negative value if `s1 < s2`, and a positive value if `s1 > s2`.

`char *strcat(char *dest, const char *src);`

Appends the string `src` to the string `dest`, returning a pointer to `dest`.

`char *strchr(const char *s, int c);`

Returns a pointer to the first occurrence of the character `c` in the string `s`.

```
char *strstr(const char *haystack, const char *needle);
```

Finds the first occurrence of the substring **needle** in the string **haystack**, and returns a pointer to the found substring, or NULL if the substring is not found.

Exercise 8-2: Type `man string` on the shell prompt to see the manual page entry of the collection of string functions available. Look up the individual manual page entries of these four functions specifically: `strlen()`, `strcmp()`, `strcpy()`, `strcat()`. Write few lines of code to understand how these functions work. For instance, you may try these as part of your program:

```
char s1[50] = "BITS", s2[] = "Pilani", s3[50];
printf("%lu %lu\n", strlen(s1), sizeof(s1));
/* It is easy to get confused between strlen() and sizeof() */
strcat(s1, " ");
strcat(s1, s2);
strcat(s1, ", ");
strcpy(s3,"Pilani Campus");
strcat(s1, s3);
puts(s1);      /* Look up the man page of puts() to learn more */
/* comparing two strings... */
if (!strcmp(s2,s3)) /* strcmp() has returned 0 */
    puts("Both strings are the same!\n");
else
    if (strcmp(s2,s3) > 1)
        printf("%s is greater than %s\n", s2, s3);
    else
        printf("%s is greater than %s\n", s3, s2);
```

Exercise 8-3: In the mid-semester test you took last week, a pointer-based implementation of `strlen()` function was provided:

```
int my_strlen(char p[])
{
    char *q = p;
    while (*q)
        q++;
    return q-p;
}
```

Now, you implement the same function without using pointers. You may name this function as `my_strlen1()`.

Exercise 8-4: Implement a function `convertLower()` (whose prototype is given below) to convert a given string into one that contains entirely lower case letters.

```
void convertLower(char *);
```

The function performs the conversion on the string received as its argument. First implement your function using ASCII codes only. Next, use the library function `tolower()` to do the job. Remember to include the header `<ctype.h>`. You can learn more about this function by using the online manual page.

[*Note:* There are a suite of useful library functions like `tolower()` and `toupper()` available in the ctype library. Get to know them by typing `man isalpha` at the shell prompt. You should be familiar with the first 13 functions – starting from `isalnum()` and ending with `isblank()`, whose descriptions are also provided in the manual page, so you can use them in your programs when needed.]

Exercise 8-5: Write a function that checks to see if the string it takes as an argument is a palindrome or not.

- (a) Write the first version of the function `isPalind_1()` using array notation, and the second version `isPalind_2()` using pointers.
- (b) Now, modify one of the above two functions so that `isPalind_3()` that you write now uses a function `strrev()` whose definition is given below:

```
void strrev(char *a) /* function to reverse the given string */
{
    char *b, temp;
    b = a + strlen(a) - 1; /* b is made to the last character preceding
                           the terminating \0 */
    for ( ; a < b ; a++, b--)
        { /* swapping the contents pointed to by a and b */
            temp = *a;
            *a = *b;
            *b = temp;
        }
    return;
}
```

1.5 Arrays and Pointer Arithmetic

Arrays and pointers have a strong connection. The name of an array represents the base address of the array, and can be assigned to a pointer. Consider the following code segment:

```
char arr[13];      char *parr;      parr = arr;
```

Here, `arr` represents an array of 13 integer elements, and `arr` is the base address of the array, i.e., address of the 0th element. In other words, `arr` is equivalent to `&arr[0]`.

Delving on this further,

`arr[i]` is the same as writing `*(arr + i)`. Both refer to the value stored at the address given by the expression `arr + i * sizeof (char)`.

`&arr[i]` is the same as writing `arr + i`. Both refer to the address of the i^{th} element of the array.

Because the name of an array is its base address, when an array is passed to a function, what is being passed is only the array's base address, and not the entire array.

<pre>void foo(char []); int main() { char arr[100]; strcpy(arr, "BITS Pilani"); printf("%s\n", arr); foo(arr); }</pre>	<pre>void foo(char ptr[]) { printf("Inside foo..."); printf("%p\n", ptr); printf("%s\n", ptr); return; }</pre>
--	--

In the above example, the array `arr` (storing the string) is allocated 100 bytes of memory only in the `main()` function, and not in `foo()`. What is being passed is the base address of `arr` (i.e., the starting location of the array), and *the value of this address* is received by `foo()` as an argument and is allotted storage under the variable name `ptr`. Since what is being received is an address, `ptr` is treated as a pointer, and so 8 bytes are allotted for it as a local variable in `foo()`. Now, when you print the value of it in the function, the base address gets printed. And using this address, you can also retrieve the contents of the array, and hence print the string “BITS Pilani” from the function as well (just as you are able to do the same within the `main()` function).

Because what is being received is an address, the formal argument can also be received a pointer. So, `void foo(char ptr[])` could have also been written as `void foo(char *ptr)`. In fact, the compiler treats the array notation of a formal argument to a function only as a pointer. How do we know this is true?

`sizeof(ptr)` inside `foo()` will yield 8, which is the number of bytes allocated to a pointer, whereas `sizeof(arr)` inside `main()` will yield 100, which is the size of the array allocated.

However, a pointer is pointer, and an array is an array!

Consider the following declaration and initialization:

```
char arr[108];      char *p = arr;
```

While `p` can be used synonymously with `arr` e.g., `p[i]`, `strlen(p)`, `strcmp(p, name)` and the likes, `arr` cannot be used as a replacement for `p`. The best illustration is you can legally do `p++` to increment the address by one byte (since `p` is declared as a `char *`), but it is illegal (a syntax error) to increment the base address by saying `arr++`.

To understand pointers and strings better, the following implementations of the library function `strcpy()` may be useful:

```
char * my_strcpy1(char *dest, char *src) /* copies src into destination array */
{
    int i, len;
    len = strlen (src);
    for ( i = 0; i <= len; ++i)
        dest[i] = src[i]; /* copies everything till '\0', including it */
    return dest;
}

char * my_strcpy2(char *dest, char *src)
{
    int i = 0;
    while ((dest[i] = src[i]) != '\0') i++;
    /* first copy to dest, then check for NUL character */
    return dest;
}

char my_strcpy2a(char *dest, char *src)
{
    int i = 0;
```

```
while (dest[i] = src[i]) i++; /* using ASCII value of '\0' is zero */
return dest;
}

char my_strcpy3(char *dest, char *src) /* pointer version */
{
    char *old_dest = dest; /* making a copy of dest */
    while (*dest = *src)
        { dest++; src++; }
    return old_dest; /* return the base address of dest array */
}
char my_strcpy3a(char *dest, char *src) /* a compact, pointer version */
{
    char *old_dest = dest; /* making a copy of dest */
    while (*dest++ = *src++) /* dereference, use the dereferenced values and
then increment pointers */
        ; /* null statement */
    return old_dest;
}
```

2. TWO-DIMENSIONAL ARRAYS AND ARRAYS OF POINTERS

2.1 Introduction

Since you are all familiar with matrices in mathematics, using 2D arrays will be very intuitive. *A 2D array is just an array of elements, each element being a one-dimensional array.*

A 1D array of 10 elements, each element of which is an integer is declared as:

```
int arr[10];
```

A 2D array of 10 elements, each element being an array of 4 integers is declared as:

```
int brr[10][4];
```

Memory allocated for brr is $10 * (4 * \text{sizeof (int)})$ bytes. On the GCC you use in this course, since the size of an integer is 4 bytes, brr will be allocated 160 bytes of memory.

Though brr can now be treated as 10 x 4 matrix (10 rows and 4 columns), memory allocation is done linearly, as done for any other array. So, brr[0] would be first (16 bytes), followed by brr[1] and so on till brr[9].

2.2 Getting data into a 2D array

The first way to get data is by initializing the elements when declaring the array, as shown here:

```
int mat1[2][4] = {{1,11,-2,6}, {2,4,-10,16}};
```

The above statement declares a 2D array of dimension 2x4, and makes the following initializations:

mat1[0][0] = 1	mat1[1][0] = 2
mat1[0][1] = 11	mat1[1][1] = 4
mat1[0][2] = -2	mat1[1][2] = -10
mat1[0][3] = 6	mat1[1][3] = 16

The following declaration initializes some elements (row 1 entirely, first two elements of row 2, first element of row 3) while the remaining are set to zero:

```
int mat2[4][4] = {{1,11,-2,6}, {-2,4}, {-5}};
```

The next declaration does not mention the number of rows, and lets the compiler calculate it automatically as 3, based on the initialization done:

```
int mat3[][4] = {{1,11,-2,6}, {-2,4}, {-5}};
```

The second way by which input can be obtained into a 2D array is by using `scanf()`. That is being done here by a user-defined function `matrixScan()` which takes the matrix as its argument:

```
void matrixScan(int s[ROWS][COLS])
    /* ROWS and COLS are symbolic constants #defined by user */
{ int i,j;
    for (i = 0; i < ROWS; i++)
        for (j = 0; j < COLS; j++)
            scanf("%d",&s[i][j]);
}
```

And here is a function that prints out the values of a 2D array in a matrix notation:

```
void matrixPrint(int s[][COLS])
{ int i,j;
    for (i = 0; i < ROWS; i++)
    {
        for (j = 0; j < COLS; j++)
            printf("%7d ",s[i][j]);
        putchar('\n'); /* to move to the next line after printing a row */
    }
}
```

Notice how the first dimension is not mentioned when the 2D array is being received as an argument to the function, because it is optional. The second dimension, however, is required to be mentioned (else, it is a compile-time error).

The function `matrixAdd()` adds matrices **a** and **b**, and stores the result in **c**:

```
void matrixAdd(int a[][COLS], int b[][COLS], int c[][COLS])
{
    int i, j;
    for (i = 0; i < ROWS; i++)
        for (j = 0; j < COLS; j++)
            c[i][j] = a[i][j] + b[i][j];
    return;
}
```

We can also declare 2D arrays of characters like this:

```
char objects[4][3] = {{'m', 'a', 't'}, {'b', 'a', 't'}, {'n', 'e', 't'},
                      {'r', 'o', 'd'}, {'p', 'i', 'n'}};
```

Or like this:

```
char objects[][3] = {{'m', 'a', 't'}, {'b', 'a', 't'}, {'n', 'e', 't'},
                      {'r', 'o', 'd'}, {'p', 'i', 'n'}};
```

Not only is this method of declaring arrays cumbersome (writing each character), but is also of limited utility. Rather, 2D arrays can be used to store and work with strings, with each row storing a string each. The same array objects we declared earlier can now be made an array of strings:

```
char objects[][][4] = {"mat", "bat", "net", "rod", "pin"};
```

The compiler automatically makes this array a 5 x 4 matrix of characters.

A few quick practice problems with integer matrices: writing functions for doing each of these tasks:

1. To check if a given matrix received as the argument is a diagonal matrix or not
2. To check if a given matrix is a lower triangular matrix or not
3. To check if a given matrix is an upper triangular matrix or not

```
int isDiagonal(int m[][COLS]) {
    int i, j;
    for (i = 0; i < ROWS; ++i)
        for (j = 0; j < COLS; ++j)
            if (i != j && m[i][j])
                return 0; /* found a non-diagonal element that is non-zero */
    return 1;
}

int isLowerTriangular(int m[][COLS]) {
    int i,j;
    for (i = 0; i < ROWS; ++i)
        for (j = 0; j < COLS; ++j)
            if (j > i && m[i][j])
                return 0; /* a non-zero element found in the upper triangle */
    return 1;
}

int isUpperTriangular(int m[][COLS]){
    int i,j;
    for (i = 0; i < ROWS; ++i)
        for (j = 0; j < COLS; ++j)
            if (i > j && m[i][j])
                return 0; /* a non-zero element found in the lower triangle */
    return 1;
}
```

LABORATORY SESSION #9

Exercise 9-1: For the piece of code shown below that finds out the amount of memory allocated (in bytes) for two arrays, can you guess what gets printed?

```
int arr[] = {10,20,45,67,68};
int mat[][2] = {{10,20}, {30,40}, {50,60}};
printf("Sizeof(arr) = %lu; sizeof (mat) = %lu\n",
       sizeof(arr), sizeof(mat));
```

Now when the array is passed to a function in which the size is calculated (as shown below), what do you think gets printed?

```
void foo(int arr[], int mat[][2])
{
    printf("Sizeof(arr) = %lu; sizeof (mat) = %lu\n",
           sizeof(arr), sizeof(mat));
}
```

A copy of the program is stored at /home/share/9.1.c. Copy it to your working directory, compile and check the outputs. Were they what you expected? Record them in your lab notebook.

Exercise 9-2: Copy the program at /home/share/9.2.c into your working directory.

```
1 #include <stdio.h>
2 int main()
3 {
4     int i = 7;
5     int arr[5] = {i, 2*i, 3*i}, *pa, *pb;
6     pa = &arr[0];
7     pb = arr+4;
8     for (i = -4; i <= 0; ++i)
9         printf("%d, ", pb[i]); /* negative array
index! */
10    putchar('\n');
11
12    printf("pb - pa = %ld\n", pb-pa); /* pointer
subtraction */
13    return 0;
14 }
```

Several interesting concepts are illustrated in this program. Understand these:

- a. What aspects of array initialization are illustrated (lines 5, 9)? Write in your lab notebook.
- b. What aspects of pointer arithmetic are illustrated (lines 9, 12)? Write them down.
- c. Try subtracting two pointers of different data types (e.g., `char *` and `int *`). Try adding two pointers of the same type; multiplying; adding a floating point number to a pointer. Write down which of these the compiler permits.

Exercise 9-3: ABC bank is trying to establish ATM branches in different regions of a city. There are `m` branches and each branch does `n` transactions per day. The transaction could be a positive value indicating deposit, or negative indicating withdrawal, or zero indicating it was neither a deposit nor a withdrawal. Write a complete C program that accomplishes these tasks:

- a. Reads values for `m` and `n`, and then reads and stores the transaction values in an `m x n` matrix. Data for the program is stored at `/home/share/9.3.txt`. Take a look at the file – the first row has values to be stored in `m` and `n`, and the rest of the matrix follows. (Tip: Use input redirection to read the data from file.)
- b. Prints the branch number (same as row number) of the branch that had the maximum total deposits. Use the following function to achieve this task:

```
int computeSumOfDeposits(int row[], int size);
```

- c. Prints the branch number of the branch that had the maximum transaction value (Hint: Use the `abs()` function from the `<stdlib.h>` library, or implement the equivalent functionality yourself.)

```
int computeMaxTransac(int row[], int size);
```

This function computes and returns the maximum transaction amount for the given ATM branch. You can view sample output of the program by executing `/home/share/9.3` to see how your program is expected to work.