

文字処理の練習

山本昌志*

2004年12月24日

1 本日の学習内容

本日と次回の授業では、プログラム作成を通して、文字処理の実際を学習する。

2 文字処理の復習

2.1 コンピューターの文字の取り扱い

まずは、基本的なことである。

- コンピューターで文字を扱うときは、それはすべて、整数に置き換えて取り扱う。文字と整数との対応を決めたものをコード表と言う。
- 1文字を表すためには、英数字では1バイト、日本語では2バイト必要である。
- 文字を扱うときには文字型の変数を使う。変数宣言の型は、「char」である。これで、宣言された変数は1バイトの情報を記憶できる。すなわち、英数字の1文字分である。
- 複数の文字が連なったものを文字列と言う。文字列を扱うためには、文字型の配列を使う。
- 文字型の配列に文字列を記憶させた場合、文字列のすぐあとに「\0」が格納される。

2.2 文字型変数と代入

つぎに、文字型の変数宣言とそれへの値の代入である。

- 英数字が1文字の場合

– 変数宣言

```
char hoge;          /* 文字型の変数 */
```

*国立秋田工業高等専門学校 電気情報工学科

- 代入。シングルクォーテーションで囲めば、代入演算子「=」が使える。

```
hoge='A';          /* 代入 */
```

- 英数字の文字列の場合

- 変数宣言

```
char hoge[10];    /* 文字型の配列 */
```

- 文字型の配列に文字を格納する方法はいくつかある。代表的な方法は、「strcpy」と「sprintf」を使う方法である。ただし、前者を使う場合、「string.h」をインクルードする必要がある。

```
strcpy(hoge, "Akita");    /* 文字列の格納 */
sprintf(hoge, "Akita");
```

- 日本語の場合は、文字型の配列を使わなくてはならない。必要な配列のサイズは、2×文字数+1である。なぜならば、日本語の1文字は2バイトで、最後に「\0」を付加するためである。

- 変数宣言

```
char hoge[10];    /* 文字型の配列 */
```

- 文字型の配列に日本語の格納は、英数字と同じで、ダブルクォーテーションで囲む。

```
strcpy(hoge, "秋田");    /* 文字列の格納 */
sprintf(hoge, "秋田");
```

2.3 入出力

文字を取り扱う場合の入出力関数を表1に示す。これらの関数を使う場合、以下の注意が必要である。

- 「gets」は改行を除去し、「puts」は改行を付加する。
- 「fgets」は改行を残し、「fputs」は改行を付加しない。
- 「fscanf」では、空白文字(英数字のスペース)は読み込めない。それは、文字列の区切りを表すからである。

表 1: 入出力に用いられる関数

	標準			ファイル		
	1文字	1行	書式付き	1文字	1行	書式付き
入力	getchar	gets	scanf	fgetc	fgets	fscanf
出力	putchar	puts	printf	fputc	fputs	fprintf

2.4 標準ライブラリー関数

文字処理のための表 2、文字列処理のための表 3 の標準ライブラリー関数について学習した。

表 2: 1 文字処理関数。#include <ctype.h>が必要。変数は、int c;。

関数名	動作
isalnum(c)	英数字なら真
isalpha(c)	英文字なら真
iscntrl(c)	制御文字なら真
isdigit(c)	数字なら真
isgraph(c)	印字可能文字なら真
islower(c)	小文字なら真
isprint(c)	空白以外の印字可能文字なら真
ispunct(c)	区切り文字なら真
isspace(c)	空白類文字なら真
isupper(c)	大文字なら真
isxdigit(c)	16 進表示文字なら真
tolower(c)	文字 c を小文字に変換
toupper(c)	文字 c を大文字に変換

表 3: 文字列処理関数。#include <string.h>が必要。変数は、char s1[256],s2[256]; のように文字型の配列。配列のサイズは、処理に必要なサイズよりも大きいこと (256 とは限らない)。c は文字型の変数、char c; である。

関数名	動作
strlen(s1)	文字列 s1 の長さ、すなわち文字数を整数値返す。
strcpy(s1,s2)	s1 に、文字列 s2 をコピーする。
strcat(s1,s2)	文字列 s1 の後に、文字列 s2 をコピーする。
strcmp(s1,s2)	文字列 s1 と s2 を比較する。 s1 > s2 の場合、戻り値は正 s1 == s2 の場合、戻り値は 0 s1 < s2 の場合、戻り値は負
strncpy(s1,s2,n)	s1 に文字列 s2 の先頭から n 文字をコピーする。
strncat(s1,s2,n)	文字列 s1 の後に文字列 s2 の先頭から n 文字を連結する。
strncmp(s1,s2,n)	文字列 s1 と文字列 s2 の先頭から n 文字を比較する。比較の結果は、strcmp と同じ。
strchr(s1,c)	文字列 s1 中の文字 c の位置を整数で返す。文字がないときは、NULL を返す。
strstr(s1,s2)	文字列 s1 中にある文字列 s2 の位置を整数で返す。もし、文字列がない場合、NULL を返す。

3 練習問題 (プログラム作成)

教科書の練習問題 p.236 のプログラムを作成せよ。どうしてもわからない場合は、答えをみて、そのプログラムの内容を理解せよ。

3.1 文字で答えるクイズ

教科書の練習問題 Lesson 6-1 のプログラムを作成せよ。

Lesson 6-1

「アルファベットの 10 番目の文字は?」という問題を表示し、解答を入力させる。入力した解答が「j」または「J」であれば「正解」そうでないとき「不正解」と表示するプログラムを作成しなさい。

3.2 ファイルからの文字の読みとり字数を数えるプログラム

教科書の練習問題 Lesson 6-2 のプログラムを作成せよ。

Lesson 6-2

ある英文原稿を 400 文字以内で書かなくてはならないとする。ただし、「.」や「,」などの記号は文字数に含めず、英字と数字のみの字数で数えるものとする。

このとき、下書き原稿のテキストファイルが何文字であるかを調べるプログラムを作成しなさい。

ただし、テキストファイルの 1 行の最大文字数は 79 文字とし、空白は文字数として数え、改行は文字数に入れないものとする。なお、テキストファイルは、/tmp/1e/read.txt とする。

以下のヒントを与えるので、それを参考にすること。

- ファイルの読み込みに関しては、教科書の例題 4(p.229-231) を参考にせよ。
- 文字型の 2 次元配列を上手に使うこと。

3.3 総合応用問題

教科書の練習問題 Lesson 6-3 のプログラムを作成せよ。

Lesson 6-3

10人の電話番号(最大11桁)と名前(氏名の順のローマ字で最大30文字、氏名の間には空白1文字)のデータを管理する電話帳を作りたい。

STEP 1 まずはデータを10人分入力し、「address.txt」ファイルに保存するプログラムを作成したい。そこで名前、電話番号を入力し、図に示すテキストファイルに書き出すプログラムを作成しなさい。

STEP 2 STEP 1で作成したデータファイルを読み込み、氏名を入力したときに該当する電話番号をすべて表示するプログラムを作成したい。ただし、氏名のうち氏だけを入力したときには、名が違っているものも含めてすべての該当する電話番号を表示するようにしたい。

STEP 3 STEP 1, 2で作成したプログラムをまとめてひとつのプログラムにしたい。そこで、はじめにメニューを表示し「登録」「検索」「終了」ができるようにしたひとつの電話帳システムを作成しなさい。

```
Koizumi Junichiro
0180000000
Yamamoto Masashi
0181234567
Hamasaki Ayumi
0181112222
.
.
.
Kimura Takuya
0189876543
```

図 1: ファイル「address.txt」の例

このプログラムは、難しく感じるかもしれない。以下のヒントを与えるので、それを参考にすること。

- 教科書の例題 3(p.226-228) を参考にせよ。
- 文字型の 2次元配列を上手に使うこと。

4 レポート

4.1 内容

練習問題のプログラムのうち、「文字で答えるクイズ」と「ファイルからの文字の読みとり字数を数えるプログラム」を完成させること。実行の確認ができたならば、それをプリントアウトして、レポートとして

提出する。

将来、情報処理関係の仕事に就きたいと思っている者は、頑張って、最後の「総合応用問題」を完成させよ。最後の問題は、できた者のみレポートとして提出のこと。

自宅のプログラムを作っても良い。そして、実行の確認がとれたならば、それをプリントアウトし、レポートとして提出しても良い。

4.2 レポート提出要領

提出方法は、次の通りとする。

期限	1月21日(金)PM5:00まで
用紙	A4
提出場所	山本研究室の入口のポスト
表紙	表紙を1枚つけて、以下の項目を分かりやすく記述すること。 授業科目名「情報処理Ⅰ」 課題名「課題 文字処理の練習」 1E 学籍番号 氏名 提出日
内容	ソースプログラム(プリントアウトのみ、手書きは不可)