

Συμβολοσειρές (Strings)

Συμβολοσειρά (string) είναι μια σειρά αλφαριθμητικών χαρακτήρων (γενικά εκτυπώσιμων συμβόλων ASCII). Όταν λέμε σειρά εννοούμε διαδοχικές θέσεις μνήμης που μπορούν να αντιμετωπισθούν σαν ένα σύνολο διατεταγμένων στοιχειωδών δεδομένων. Η αναγκαιότητα των συμβολοσειρών πρέπει να έχει γίνει ήδη κατανοητή αφού πρόκειται για τα μηνύματα που συνοδεύουν τις εκτυπώσεις με τη συνάρτηση printf.

Στη C η συμβολοσειρά ορίζεται σαν μια σειρά δεδομένων του τύπου char τα οποία τερματίζονται με ένα μηδενικό χαρακτήρα (NULL). Όταν η C χειρίζεται τη συμβολοσειρά με κάποιο τρόπο, π.χ. να τη συγκρίνει με κάποια άλλη συμβολοσειρά, να την εκτυπώσει, να την αντιγράψει σε μια άλλη συμβολοσειρά, κ.λ.π., οι συναρτήσεις που καλούνται είναι σχεδιασμένες να εκτελούν τη λειτουργία τους μέχρι να συναντήσουν τον μηδενικό χαρακτήρα.

Πίνακας είναι μια σειρά από δεδομένα του αυτού τύπου, ο οποίος όμως τύπος δεν είναι απαραίτητα στοιχειώδης (π.χ. ακέραιος, πραγματικός, κ.λ.π). Η συμβολοσειρά είναι ένας ειδικός τύπος πίνακα με δεδομένα του τύπου char και με τερματισμό μέσω του μηδενικού χαρακτήρα.

```
char name[5];    /* define a string of characters */
```

Το πρώτο που πρέπει να προσέξουμε είναι ότι ο τρόπος δήλωσης μιας συμβολοσειράς γίνεται ως πίνακας χαρακτήρων. Η δήλωση μοιάζει με τη δήλωση μεταβλητής χαρακτήρα αλλά μετά το όνομα της μεταβλητής, και χωρίς κενό, ακολουθεί ένα ζεύγος τετράγωνων παρενθέσεων που περιέχει έναν αριθμό ο οποίος καθορίζει το μέγεθος του πίνακα, δηλαδή τον αριθμό των στοιχειωδών δεδομένων που περιέχει ο πίνακας. Στην πιο πάνω δήλωση η συμβολοσειρά μπορεί να έχει το πολύ 5 χαρακτήρες. Η δήλωση μεγέθους είναι απαραίτητη για να δεσμεύσει ο μεταφραστής 5 συνεχόμενες θέσεις μνήμης, όπου η κάθε μια έχει χωρητικότητα ενός χαρακτήρα. Οι θέσεις αυτές έχουν η κάθε μια το δικό της όνομα που καθορίζεται ως συνδυασμός δύο τμημάτων: το όνομα του πίνακα, δηλαδή name, και ο αριθμός σειράς του στοιχείου. Στην C ο αριθμός σειράς ξεκινά από το 0 και φθάνει μέχρι το μέγεθος του πίνακα μείον μια θέση. Έτσι έχουμε τις μεταβλητές name[0], name[1], name[2], name[3], και name[4].

Για το μεταφραστή και το σύστημα εκτέλεσης του προγράμματος τα ονόματα αυτά σημαίνουν ακριβώς:

Βασικός χειρισμός συμβολοσειρών

```
char name[5];    /* define a string of characters */
```

Σύμφωνα με την πιο πάνω δήλωση, ο πίνακας χαρακτήρων `name` έχει 5 θέσεις. Για να τον χρησιμοποιήσουμε όμως ως συμβολοσειρά πρέπει ο τελευταίος χαρακτήρας να είναι ο μηδενικός, άρα μένουν τέσσερις ωφέλιμες θέσεις χαρακτήρων. Όταν λέμε μηδενικός χαρακτήρας εννοούμε το αριθμητικό μηδέν, όχι το χαρακτήρα μηδέν. Μερικές φορές θα το συνατήσετε και ως `'\0'`.

Η εκτύπωση μιας σταθερής συμβολοσειράς γίνεται με τη βοήθεια της συνάρτησης `printf`. Αλλά η συμβολοσειρά - μήνυμα περικλείεται από διπλά εισαγωγικά, π.χ.

```
printf("This is a constant string");
```

Στη περίπτωση του πίνακα `name` έχουμε μια μεταβλητή συμβολοσειρά και η εμφάνισή της γίνεται με τον χαρακτήρα ελέγχου `'%s'`.

```
printf("%s",name);
```

Σημειώστε ακόμη ότι μόνο το όνομα του πίνακα δίνεται, χωρίς την ένδειξη των τετράγωνων παρενθέσεων. Αυτό σημαίνει ότι στην εκτύπωση της συμβολοσειράς χρησιμοποιείται μόνο η διεύθυνση της βάσης, δηλαδή του ονόματος `name`. Ο μηδενικός χαρακτήρας δεν εκτυπώνεται αλλά χρησιμοποιείται από την `printf` για τον έλεγχο τερματισμού της συμβολοσειράς.

Ο κάθε χαρακτήρας που είναι στοιχείο της συμβολοσειράς μπορεί να εκτυπωθεί αυτόνομα με τη χρήση του χαρακτήρα ελέγχου `'%c'` και του ονόματος του συγκεκριμένου χαρακτήρα, π.χ. `name[2]`.

```
printf("%c",name[2]);
```

Μπορούμε επίσης να εκφράσουμε τη λειτουργία εκτύπωσης ολόκληρης της συμβολοσειράς με τη βοήθεια μιας επανάληψης `for` ως

```
for (index = 0; name[index]; index++)  
    printf("%c",name[index]);
```

Σημειώστε ότι ο μηδενικός χαρακτήρας ισοδυναμεί με το αριθμητικό 0, δηλαδή τη λογική τιμή του ψευδούς άρα καθιστά τη συνθήκη ελέγχου ψευδή και τερματίζει την επανάληψη.

Τέλος υπάρχει η δυνατότητα να εκτυπώσουμε μια συμβολοσειρά όχι από την αρχή αλλά από κάποιο σημείο και μετά. Όπως είδαμε η εκτύπωση της συμβολοσειράς απαιτεί μόνο μια διεύθυνση, την διεύθυνση εκκίνησης της εκτύπωσης. Όταν εκτυπώνουμε το σύνολο

της συμβολοσειράς, η διεύθυνση εκκίνησης συμπίπτει με τη διεύθυνση της βάσης, δηλαδή το όνομα `name`. Όταν όμως θέλουμε να ξεκινήσουμε από τον πρώτο, ας πούμε, χαρακτήρα, πρέπει να ορίσουμε τη δική του διεύθυνση ως διεύθυνση εκκίνησης.

Βλέπουμε ότι η διεύθυνση `name` ισοδυναμεί με τη διεύθυνση `&name[0]` που διαβάζεται ως 'η διεύθυνση του μηδενικού στοιχείου του πίνακα `name`'. Η χρησιμότητα του συμβόλου της διεύθυνσης `&` θα γίνει πιο κατανοητή στα επόμενα κεφάλαια.

Μερικές συναρτήσεις συμβολοσειρών

Οι συμβολοσειρές είναι ιδιαίτερα χρήσιμες στην επεξεργασία κειμένου κάθε είδους. Μη ξεχνάτε ότι και τα προγράμματά μας είναι κείμενα μιας ειδικής γλώσσας, άρα και η μετάφραση ενός προγράμματος είναι μια εξειδικευμένη επεξεργασία κειμένου. Τελικά στη μνήμη του υπολογιστή όλες οι πληροφορίες (εντολές, αριθμοί, χαρακτήρες) δεν είναι παρά σύμβολα. Επίσης πολλά προγράμματα συστημάτων (όπως για παράδειγμα οι επικοινωνίες) στηρίζονται σε επεξεργασία συμβολοσειρών. Γι' αυτούς τους λόγους η C παρέχει μια βιβλιοθήκη συναρτήσεων για την επεξεργασία συμβολοσειρών.

#include <string.h>

Με την πιο πάνω εντολή το πρόγραμμα ειδοποιεί το μεταφραστή να ενσωματώσει το αρχείο κεφαλής **string.h**. Το αρχείο αυτό περιέχει τη βασική βιβλιοθήκη συναρτήσεων για τη διαχείριση συμβολοσειρών.

Μια πολύ χρήσιμη συνάρτηση που επιτρέπει την αντιγραφή μιας συμβολοσειράς σε μια άλλη είναι η **strcpy**. Η αντιγραφή διακόπτεται όταν στη συμβολοσειρά - παραγωγό βρεθεί ο μηδενικός χαρακτήρας, δηλαδή

for (i = 0; (consum[i] = produc[i]); i++);

Προσέξτε τη συνθήκη ελέγχου που κάνει ουσιαστικά όλη τη δουλειά. Εκχωρεί έναν χαρακτήρα από τον παραγωγό στον καταναλωτή αλλά επειδή η εκχώρηση είναι και έκφραση έχει ως εξαγόμενο τη τιμή του χαρακτήρα `produc[i]`. Όταν βρεθεί ο μηδενικός χαρακτήρας η συνθήκη γίνεται ψευδής και η επανάληψη τερματίζεται. Σημειώστε το ερωτηματικό μετά το κλείσιμο της παρένθεσης (σημαίνει ότι το `for` έχει άδειο σώμα, δηλαδή δεν εκτελεί καμία πρόταση).

```
strcpy(name1, "Christos");  
strcpy(name, name1);
```

Όπως φαίνεται στις πιο πάνω εντολές, η `strcpy` δουλεύει με παραγωγό είτε σταθερά είτε μεταβλητή συμβολοσειρά. Για τις σταθερές συμβολοσειρές χρησιμοποιούμε το διπλό εισαγωγικό. Η `strcpy` μπορεί να παρομοιαστεί με μια εκχώρηση μεταξύ συμβολοσειρών. Με την εκτέλεση των πιο πάνω εντολών οι συμβολοσειρές `name1` και `name2` θα πάρουν την τιμή 'Christos'.

Μια άλλη συνάρτηση συμβολοσειρών είναι η **`strcmp`**, η οποία συγκρίνει δύο συμβολοσειρές `s1` και `s2`. Επιστρέφει 1 αν η `s1` είναι μεγαλύτερη, 0 αν είναι ίσες, -1 αν η `s2` είναι μεγαλύτερη. Η σύγκριση γίνεται σύμφωνα με τον κώδικα ASCII. Δύο συμβολοσειρές είναι ίσες αν έχουν το ίδιο μήκος και ίδιους χαρακτήρες σε κάθε θέση. Μια συμβολοσειρά είναι μεγαλύτερη από μίαν άλλη εάν (α) έχει μεγαλύτερο μήκος ή (β) σε κάποια θέση υπάρχει χαρακτήρας που έχει μεγαλύτερο κωδικό ASCII από τον αντίστοιχο χαρακτήρα της άλλης συμβολοσειράς. Για παράδειγμα η συμβολοσειρά 'Zeke' είναι μεγαλύτερη από τη 'Rosalina' γιατί το Z έχει μεγαλύτερο κωδικό από το R. Παρόμοια η συμβολοσειρά 'Rosalina' είναι μεγαλύτερη από τη συμβολοσειρά 'Rosa' γιατί έχει περισσότερα γράμματα ενώ όσα έχουν είναι ίδια. Υπάρχουν συναρτήσεις βιβλιοθήκης που λαμβάνουν υπόψη και τα αν τα γράμματα είναι κεφαλαία ή πεζά κ.ο.κ.

Ακόμη μια συνάρτηση συμβολοσειρών είναι η **`strcat`**, η οποία συνενώνει δύο συμβολοσειρές `s1` και `s2` θέτοντας την `s2` τέλος της `s1` και το όλο αποτέλεσμα τοποθετείται στην `s1`. Η λειτουργία αυτή μπορεί να εκφραστεί ως:

```
for (length1 = 0; (s1[length1]); length1++);  
for (i = 0; (s1[length1+i] = s2[i]); i++);
```

όπου `length1` είναι το μήκος της συμβολοσειράς `s1`.

Για παράδειγμα οι πιο κάτω εντολές:

```
strcpy (s1,"Christos");  
strcpy (s2,"Ttofis");  
printf("My name is %s.\n",s1);  
printf("My surname is %s.\n",s2);  
strcat(s1, " ");  
strcat(s1, s2);  
printf("Both together are %s.\n",s1);
```

Θα έχουν το ακόλουθο αποτέλεσμα:

```
My name is Christos.  
My surname is Ttofis.  
Both together are Christos Ttofis.
```

Η βιβλιοθήκη συναρτήσεων της C για διαχείριση συμβολοσειρών περιέχει πολλές χρήσιμες συναρτήσεις οι οποίες μπορούν να χρησιμοποιηθούν σαν βάση για πιο σύνθετες συναρτήσεις. Ο αναγνώστης προτρέπεται να μελετήσει προσεκτικά αυτή τη βιβλιοθήκη.

Πανεπιστήμιο Κύπρου - Τμήμα Πληροφορικής
ΕΠΑ 032 – Προγραμματισμός Μεθόδων Επίλυσης Προβλημάτων
Σημειώσεις Εργαστηρίων
Μαρία Σταυρινού Ιωάννου