



Ecole Polytechnique Sousse
Département d'Informatique
NIVEAU : 3^{eme} GI, TR, IG
AU : 2020-2021
Programmation C
Les chaînes de caractères

1 Introduction

- Il existe une convention de représentation des chaînes. Celle-ci est utilisée à la fois :
 - * Par le compilateur, pour représenter les chaînes constantes (notées entre double quotes).
 - * Par un certain nombre de fonctions qui permettent de réaliser :
 - Les lectures ou écritures de chaînes,
 - Les traitements classiques tels que concaténation, recopie, comparaison, conversions,...
- Une chaîne de caractères est traitée comme un tableau à une dimension de caractères.
- Une chaîne de caractères à une suite d'octets terminée par `'\0'`.
- Une chaîne de n caractères occupe n+1 octets.

2 Déclaration et Mémorisation

2.1 Déclaration

`char <nom variable>[<longueur>];`

Exemples :

1. `char NOM[20];`
2. `char PHRASE[30];`

2.2 Mémorisation

Le nom d'une chaîne est le représentant de l'adresse du premier caractère de la chaîne. Pour mémoriser une variable qui doit être capable de contenir un texte de N caractères nous avons besoin de N+1 octets en mémoire.

Exemple :

`char TXT[10] = "BONJOUR!";`

TXT

.....	'B'	'O'	'N'	'J'	'O'	'U'	'R'	' '	'!'	'\0'
Adresse :	1E04	1E05	1E06	1E07	1E08	1E09	1E0A	1E0B	1E0C	1E0D	1E0E

3 Initialisation de chaînes de caractères

1. `char CHAINE[] = 'B', 'O', 'N', 'J', 'O', 'U', 'R', '\0' ;`

Nous pouvons utiliser une initialisation plus confortable en indiquant simplement une chaîne de caractères constante.

```
char CHAINE[] = "BONJOUR";
CHAINE
```

'B'	'O'	'N'	'J'	'O'	'U'	'R'	'\0'
-----	-----	-----	-----	-----	-----	-----	------

2. `char TXT[8] = "BONJOUR" ;`
`TXT`

'B'	'O'	'N'	'J'	'O'	'U'	'R'	'\0'
-----	-----	-----	-----	-----	-----	-----	------

3. `char TXT[7] = "BONJOUR" ;` ERREUR pendant l'exécution.

4 Accès aux éléments d'une chaîne

L'accès à un élément d'une chaîne de caractères peut se faire de la même façon que l'accès à un élément d'un tableau.

Exemple :

```
char CH[8] = "BONJOUR" ;
```

CH

'B'	'O'	'N'	'J'	'O'	'U'	'R'	'\0'
CH[0]	CH[1]	CH[2]	CH[3]	CH[4]	CH[5]	CH[6]	CH[7]

5 Précédence alphabétique et lexicographique

5.1 Précédence alphabétique des caractères

La précédence des caractères dans l'alphabet d'une machine est dépendante du code de caractère utilisé.

Pour le code ASCII, nous pouvons constater l'ordre suivant :

..., 0, 1, 2 ..., A, B, C, ..., Z ..., a, b, c ...,

De la précédence alphabétique des caractères, on peut déduire une relation de précédence « est inférieur » sur l'ensemble des caractères.

Ainsi on peut dire : '0' est inférieur à 'Z' et noter '0' < 'Z' (code ASCII '0' = 48 et 'Z' = 90)

5.2 Précédence lexicographique des chaînes de caractères

En nous basant sur cette relation de précédence alphabétique des caractères nous pouvons définir une précédence lexicographique pour les chaînes de caractères.

Cette relation de précédence suit l'ordre du dictionnaire et est définie de façon récurrente :

1. La chaîne vide "" précède lexico graphiquement toutes les autres chaînes.
2. La chaîne A = "a1a2....ap" précède lexico graphiquement la chaîne B = "b1b2....bm" si l'une des deux conditions suivantes est remplie :
 - a) 'a1' < 'b1'
 - b) 'a1' = 'b1' et "a2a3....ap" précède lexico graphiquement "b2b3....bm"

Exemple :

1. "ABC" précède "BCD" car 'A' < 'B'
2. "ABC" précède "B" car 'A' < 'B'
3. "Abc" précède "abc" car 'A' < 'a'
4. "ab" précède "abcd" car "" < "cd"
5. " ab" précède "ab" car ' ' < 'a' code ASCII ' ' =32 et 'a' =97

6 Les fonctions de <stdio>

Le langage C offre plusieurs possibilités de lecture ou d'écriture de chaînes.

- L'utilisation du code de format %s dans les fonctions printf et scanf.
- Les fonctions spécifiques de lecture gets ou d'affichage puts d'une chaîne.

6.1 La fonction « puts »

Syntaxe :

`puts (<chaîne>);`

«puts» écrit la chaîne de caractères désignée par <chaîne> et provoque un retour à la ligne.
`puts (TXT);` \iff `printf("%s" , TXT);`

6.2 La fonction « gets »

Syntaxe :

`gets (<chaîne>);`

«gets» lit une ligne de caractères et la copie à l'adresse indiquée par <chaîne>. Le retour à la ligne final est remplacé par le symbole de fin de chaîne '\0'.

Exemple :

```
#include<stdio.h>
void main()
{ char nom[20], prenom[20], ville[25] ;
  printf ("Quelle est votre ville : " ) ;
  gets (ville) ;
  printf("Donner votre nom et votre prénom : " ) ;
  scanf("%s %s" , nom , prenom) ;
  printf("Bonjour cher %s %s qui habite à ", prenom, nom ) ;
  puts (ville) ;
}
```

Résultat :

Quelle est votre ville : Sousse
Donner votre nom et votre prénom : Makloul Sami
Bonjour cher Sami Makloul qui habite à Sousse

- Comme le nom d'une chaîne de caractères est le représentant de l'adresse du premier caractère de la chaîne, il ne doit pas être précédé de l'opérateur &.

- Dans les appels de fonctions «scanf» et «puts», les identificateurs de tableau nom et prénom ou ville n'ont pas besoin d'être précédés de l'opérateur & puisqu'ils représentent déjà des adresses (&nom = nom).
- La fonction «puts» réalise un changement de ligne à la fin de l'affichage de la chaîne, ce qui n'est pas le cas de «printf» avec le code de format %s.

ANNEXE

Les chaines de caractères

7 Les fonctions de <ctype>

Les fonctions de <ctype> servent à classifier et à convertir des caractères. Dans la suite, <c> représente une valeur du type int qui peut être représentée comme caractère. Les fonctions de classification suivantes fournissent un résultat du type int différent de zéro, si la condition respective est remplie, sinon zéro.

Fonction	Résultat
isupper(<c>)	si <c> est une majuscule ('A'...'Z')
islower(<c>)	si <c> est une minuscule ('a'...'z')
isdigit(<c>)	si <c> est un chiffre décimal ('0'...'9')
isspace(<c>)	si <c> est un signe d'espacement (' ', '\t', '\n', '\r', '\f')
isalnum(<c>)	retourne une valeur différente de 0 si c est une lettre ou un chiffre décimal.
isalpha(<c>)	retourne une valeur différente de 0 si c est une lettre (minuscule ou majuscule)

Les fonctions de conversion suivantes fournissent une valeur du type int qui peut être représentée comme caractère ; la valeur originale de <c> reste inchangée :

Fonction	Résultat
tolower(<c>)	retourne <c> converti en minuscule si <c> est une majuscule
toupper(<c>)	retourne <c> converti en majuscule si <c> est une minuscule

8 Les fonctions de <string>

La bibliothèque <string> fournit une multitude de fonctions pratiques pour le traitement de chaînes de caractères.

Dans les fonctions représentées si dessous :

- <n> représente un nombre de type int.
- <s> et <t> peuvent être remplacées par : Dans les fonctions représentées si dessous :
 - * Une chaîne de caractères constante.
 - * Le nom d'une variable déclarée comme tableau de char.
- <c> représente un caractère de type char.

Fonction	Résultat
strlen (<s>)	fournit la longueur de la chaîne <s> sans compter le '\0' final.
strcat(<s>,<t>)	ajoute <t> à la fin de <s>
strncat(<s>,<t>,<n>)	ajoute au plus <n> caractères de <t> à la fin de <s>
strcmp(<s>,<t>)	compare <s> et <t> lexico graphiquement et fournit un résultat <ul style="list-style-type: none"> — négatif si <s> précède <t> — zéro si <s> est égal à <t> — positif si <s> suit <t>

Fonction	Résultat
strcpy (<s>,<t>)	copie <t> vers <s>
strncpy (<s>,<t>,<n>)	copie au plus <n> caractère de <t> vers <s>
strchr (<s>,<c>)	Recherche, dans la chaîne <s> la première position où apparaît le caractère mentionné <c>
strrchr (<s>,<c>)	Recherche, dans la chaîne <s> (en commence à partir de la fin de <s>) la dernière position où apparaît le caractère mentionné <c>
strstr (<s>,<t>)	Recherche, dans la chaîne <s> la première occurrence complète de la sous chaîne <t> mentionné
strpbrk (<s>,<t>)	Recherche, dans la chaîne <s> la première occurrence d'un caractère quelconque de la chaîne <t> mentionné

9 Les fonctions de <stdlib>

La bibliothèque <stdlib> contient des déclarations de fonctions pour la conversion de nombres en chaînes de caractères et vice-versa.

9.1 Chaîne → Nombre

Dans les fonctions représentées si dessous :

<s> peut être remplacé par :

- une chaîne de caractères constante
- le nom d'une variable déclarée comme tableau de char

Fonction	Résultat
atoi (<s>)	retourne la valeur numérique représentée par <s> comme int
atol (<s>)	retourne la valeur numérique représentée par <s> comme long
atof (<s>)	retourne la valeur numérique représentée par <s> comme double

Règles générales pour la conversion :

- Les espaces au début d'une chaîne sont ignorés
- Il n'y a pas de contrôle du domaine de la cible
- La conversion s'arrête au premier caractère non convertible
- Pour une chaîne non convertible, les fonctions retournent zéro

9.2 Nombre → Chaîne

1. itoa (<n_int>, <s>,)
2. ltoa (<n_long>, <s>,)
3. ultoa (<n_uns_long>, <s>,)

Chacune de ces trois procédures convertit son premier argument en une chaîne de caractères qui sera ensuite attribuée à <s>. La conversion se fait dans la base .

Dans les fonctions représentées si dessus :

- <n_int> : un nombre du type int
- <n_long> : un nombre du type long
- <n_uns_long> : un nombre du type unsigned long
- <s> : une chaîne de caractères
- : est la base pour la conversion (2 ... 36)