# Şiruri de caractere\*)

**Capitolul** 

- Operații cu variabile de tip string
- Subprograme predefinite pentru string-uri
- **❖** Implementări sugerate
- **Probleme propuse**
- Soluțiile problemelor

10

Tipul *şir de caractere* se identifică în Pascal prin termenul *string*, deoarece în acest limbaj de programare există tipul predefinit **string** pentru tablouri unidimensionale care au elemente câte un caracter. Diferențele importante dintre un șir de caractere și un *string* se desprind din următoarele observații:

- Unui tablou i se specifică de la început numărul maxim de componente pe care poate să le conțină, pe când o variabilă de tip string are implicit cel mult 255 de elemente.
- Dacă la declarare se specifică pentru dimensiunea maximă a unei variabile de tip string un număr mai mic decât 255, va fi valabilă declarația făcută explicit.
- Un string se memorează pe atâția octeți câte caractere sunt specificate în declarare, la care se adaugă un octet în care se memorează lungimea string-ului. Acest lucru nu se întâmplă la tablourile de caractere care se memorează pe exact atâția octeți câte caractere apar în declarare.

#### Declararea variabilelor de tip string

Cu toate că tipul descrie o structură (șir de caractere), nu este obligatoriu să precizăm o anumită lungime, dar dacă dorim să declarăm un *string* mai scurt decât lungimea implicită (255), atunci acest lucru este posibil.

#### Exemplu

<sup>\*)</sup> În cele ce urmează, atunci când vom scrie *string*, în acest material, ne referim la un șir de caractere. Atunci când vom scrie **string**, este vorba de tipul predefinit în limbajul Pascal.

## 10.1. Operații cu variabile de tip string

#### 10.1.1. Citirea

O variabilă de tip string se poate citi în Pascal, specificându-i numele. Avantajul față de tablouri constă în faptul că nu trebuie să știm exact numărul caracterelor. În momentul în care am apăsat Enter, sau, în caz de citire dintr-un fișier de tip Text, atunci când urmează caracterul sfârșit de linie, *string*-ul "s-a terminat" și prima componentă a variabilei (cea cu indice 0) primește o valoare egală cu numărul caracterelor citite.

#### Exemplu

```
var s:string;
    ss:string[5];
...
    ReadLn(s);
    ReadLn(ss);
```

Dacă pentru s și pentru ss introducem șirul de caractere 'ABC', valoarea ambelor variabile va fi 'ABC'. Dacă pentru ss introducem 'ABCABC', valoarea lui ss va fi 'ABCAB', adică se vor păstra cel mult atâtea elemente câte permite declarația.

#### **10.1.2.** Afişarea

Afișarea se realizează specificând o expresie de tip string.

#### Exemplu

#### 10.1.3. Atribuirea

Unei variabile de tip **string** i se poate atribui valoarea unei expresii care are tipul **string**. Dacă variabila are un tip care precizează o lungime mai mică decât *string*-ul din expresie, caracterele care depășesc lungimea variabilei se pierd.

#### Exemplu

```
Fie declarația: var ss:string[5];
```

Dacă valoarea lui ss='123', atribuirea ss:=ss+'457890'; se va finaliza cu valoarea '12345' pentru ss.

#### 10.1.4. Concatenarea

Concatenarea se poate realiza utilizând operatorul '+', dar în Pascal există și o procedură predefinită în acest scop (Concat).

#### Exemplu

```
s3:=s1 + s2; { dac\check{a} s1 = 'a' \ si s2 = 'b', atunci valoarea lui s3 va fi 'ab' } 
 <math>s3:=s3 + '***' + s1; { valoarea \ lui s3 \ va fi 'ab***a' }
```

Să observăm că un *string* poate fi "construit" prin adăugare de caractere la un *string* inițial vid.

#### Exemplu

## 10.1.5. Operații relaționale

În cazul expresiilor de tip string se pot folosi operatorii relaționali cunoscuți:

```
<, <=, <>, >, >=, =.
```

Rezultatul a două comparații este o constantă logică (adevărat sau fals).

### Exemplu

```
var s1,s2,s3:string;
```

Având declarațiile de mai sus, în tabelul următor prezentăm câteva exemple de expresii relaționale după efectuarea următoarelor operații de atribuire:

```
s1:='A';
s2:='ABC';
s3:=s1 + s2;
```

Expresie	Rezultat	
s1 = 'A'	true (s-a comparat 'A' cu 'A')	
s2 <> 'ABC'	false (s-a comparat 'ABC' cu 'ABC')	
s3 < 'ABC'	true (s-a comparat 'AABC' cu 'ABC' și	
	'A' < 'B' pe cea de a doua poziție)	
s1 < 'AA'	true (s-a comparat 'A' cu 'AA')	
s3 >= s1	true (s-a comparat 'AABC' cu 'A')	

## 10.1.6. Accesul la o componentă

Componenta x[i] reprezintă al i-lea caracter din variabila x, de tip string.

## 10.2. Subprograme predefinite pentru string-uri

Unit-ul *System* din Borland Pascal 7.0 conține mai multe subprograme predefinite care pot fi utilizate în prelucrarea *string*-urilor. Fie *s* și *ss* variabile de tip **string**.

- Funcția Pos (ss, s) returnează valoarea poziției începând de la care se regăsește ss în s; dacă ss nu se găsește în s, atunci i ia valoarea 0.
- Funcția Copy (s, i, j) returnează *substring*-ul din *s* care începe la poziția *i* și are *j* caractere. Dacă în *s* nu există *j* caractere începând cu poziția *i*, se vor returna atâtea câte sunt.
- Funcția Length (s) returnează lungimea stringului (numărul caracterelor din s).
- Procedura Delete(s,i,j) are ca efect ștergerea din s, începând de la poziția i, a j caractere. Dacă în s nu există j caractere începând cu poziția i, se vor șterge atâtea câte sunt.
- Procedura Insert(ss,s,i) inserează caracterele din *string*-ul *ss* în *s*, începând de la poziția *i*.

## 10.3. Implementări sugerate

Pentru a vă familiariza cu prelucrările șirurilor de caractere, implementați programele aferente următoarelor exerciții:

- 1. citirea unui număr caracter cu caracter;
- 2. citirea unui cuvânt caracter cu caracter;
- 3. numărarea literelor dintr-un text dat:
- 4. numărarea cuvintelor dintr-un text dat;
- 5. numărarea propozițiilor dintr-un text dat;
- **6.** descompunerea în cuvinte a unui text dat;
- 7. ordonarea unui șir de cuvinte;
- **8.** eliminarea spațiilor duble dintre cuvintele unui text dat.

## 10.4. Probleme propuse

#### 10.4.1. Litere

Fie un text citit dintr-un fișier. Să se afișeze acele cuvinte din text în componența cărora numărul majusculelor este strict mai mare decât numărul literelor mici.

#### Date de intrare

Fișierul de intrare **LITERE**. **IN** conține textul scris pe mai multe linii.

#### Date de iesire

Cuvintele se vor scrie în fișierul de ieșire **LITERE.OUT** pe linii distincte, fiecare cuvânt fiind scris pe câte o linie nouă.

#### Restricții și precizări

- cuvintele sunt separate prin spații și/sau caractere de sfârșit de linie;
- fișierul nu conține linii vide;
- liniile conțin cuvinte nedespărțite;
- pe o linie sunt scrise cel mult 255 de caractere;
- un cuvânt poate avea cel mult 255 de caractere;
- cuvintele conțin numai litere mari și mici.

### Exemplu

LITERE.IN		LITERE.OUT
Ana AVEA MULte	Mere	AVEA
ACESTa		MULte
esTe un eXeMPLu		ACESTa
		eXeMPLu

#### Explicații

- Ana are două litere mici și una mare, deci nu se scrie.
- AVEA are patru litere mari și nici una mică, deci se scrie.
- MULte are trei litere mari și două mici, deci se scrie.
- Mere are o literă mare și trei mici, deci nu se scrie.
- ACESTa are patru litere mari și o literă mică, deci se scrie.
- esTe are o literă mare și trei litere mici, deci nu se scrie.
- un nu are nici o literă mare, deci nu se scrie.
- exemplu are 4 litere mari și trei litere mici, deci se scrie.

#### 10.4.2. Text

Se citesc mai multe propoziții dintr-un fișier. Se cere:

- să se afișeze numărul de cuvinte din fiecare propoziție;
- să se determine pentru fiecare propoziție numărul cuvintelor distincte;
- să se afișeze cuvintele care se repetă în cadrul aceleiași propoziții.

#### Date de intrare

Propozițiile sunt scrise în fișierul de intrare **TEXT.IN** pe mai multe linii. Două propoziții sunt separate prin caracterele '.', '!', '?'.

#### Date de ieșire

Rezultatele se vor scrie în fișierul de ieșire **TEXT.OUT**. Pe linia *i* din fișier se afișează datele referitoare la a *i*-a propoziție. Aceste date se vor separa prin spațiu și vor apărea în ordinea: *numărul cuvintelor* din propoziție, *numărul cuvintelor distincte*, *lista cuvintelor* care nu sunt distincte în cadrul propoziției.

#### Restricții și precizări

- din cuvinte poate să facă parte caracterul '-';
- două cuvinte se separă prin caracterele: ' ', ':', ', ';
- pot exista separatori inutili;
- numărul maxim de cuvinte pentru fiecare propoziție este 30;
- în text există litere mari și mici.

#### Exemplu

#### TEXT. IN

```
Sa revenim la coroana pe care am lasat-o in mana lui Arhimede, care se uita ganditor la ea nestiind ce sa-i faca.

O lasa deceptionat din mana si pleaca la alte treburi.

Dar problema, ca intrebare, staruie in mintea lui; s-o cantaresc, s-o cantaresc, aude in sine un murmur in surdina, da, asta stiu, s-o cantaresc, dar ce altceva sa-i mai fac?
```

#### TEXT.OUT

```
22 18 la care
10 10
30 21 in s-o cantaresc
```

#### 10.4.3. Traducere

Un elev are de scris un text în limba engleză, dar nu cunoaște destul de bine această limbă. Se hotărăște să lase cuvintele pe care nu le poate traduce în limba română, apoi

să le caute în dicționar. Din aceste cuvinte va alcătui o listă, urmând ca ulterior să le înlocuiască în text.

#### Date de intrare

Pe prima linie a fișierului **TRADUS. IN** se află textul elevului, iar pe fiecare dintre următoarele linii se află un cuvânt în limba română, urmat, după un spațiu, de corespondentul său în limba engleză.

#### Date de ieșire

Textul tradus, împreună cu semnele de punctuație existente în textul dat în fișierul de intrare, se va afișa în fisierul de ieșire **TRADUS.OUT**.

#### Restricții și precizări

- textul în fișierul de intrare trebuie să conțină cel mult 255 de caractere;
- separatorii folosiți în text pot fi '', '.', ', ';
- textul din fişierul de ieşire va avea exact aceeaşi separatori ca şi cel de intrare.

## Exemplu TRADUS.IN

Ana has many mere, dar she is going to the piata to vanda them. mere apples dar but piata marketplace vanda sell

#### TRADUS.OUT

Ana has many apples, but she is going to the marketplace to sell them.

#### 10.4.4. Culori

La o grădiniță se va realiza un spectacol, unde copiii trebuie să fie îmbrăcați în costume de culori diferite. Fiecare copil își spune culoarea preferată, iar învățătoarea formează grupuri de copii îmbrăcați în costume de aceeași culoare. Pe scenă copiii se vor așeza în grupuri, astfel încât cel mai numeros grup va sta în fundal, în fața acestora va sta grupul din care fac parte mai puțini copii etc., iar în față grupul cel mai puțin numeros.

#### Date de intrare

Datele de intrare se citesc din fisierul **CULORI.IN** care are următoarea structură:

- pe prima linie se află numărul natural n, reprezentând numărul de copii;
- pe următoarele *n* linii sunt scrise numele și prenumele copiilor; numele este separat de prenume printr-un spațiu;

- pe următoarea linie este scris numărul natural m, reprezentând numărul de culori;
- pe următoarele *n* linii sunt trecute culorile preferate ale celor *n* copii sub forma unor numere naturale din intervalul [1, *m*] cu care s-au codificat culorile; fiecare copil alege o singură culoare.

#### Date de iesire

Datele de ieșire se scriu în fișierul **CULORI.OUT** pe *m* linii. Fiecare linie conține codul culorii, urmat de numele copiilor care preferă culoarea respectivă. Numele vor fi despărțite printr-un spațiu. Numărul numelor va fi cel mai mare pe prima linie, mai mic pe linia a doua și cel mai mic pe ultima linie.

#### Restricții și precizări

- $2 \le n \le 20$ ;
- $2 \le m \le 7$ ;
- fiecare culoare dintre cele m va fi preferată de cel puțin un copil;
- în cazul în care există două sau mai multe culori preferate de un număr egal de copii, atunci ordinea în care grupurile corespunzătoare se vor așeza pe scenă nu contează.

## Exemplu CULORI.IN

#### CULORI.OUT

7 3 Popescu Marius Rus Andrei Rusu Teodor
Pop Mihai 1 Pop Mihai Popan Ana
Popan Ana 2 Popa Ion Trif Ionut
Popa Ion
Popescu Marius
Rus Andrei
Rusu Teodor
Trif Ionut
3
1

## 10.5. Soluțiile problemelor propuse

#### 10.5.1. Litere

Vom citi câte un *string* de pe o linie din fișier și îl vom parcurge în felul următor: vom ignora spațiile până la apariția primului caracter diferit de spațiu. Apoi, vom construi din caractere consecutive litere un cuvânt, numărând în paralel și majusculele. În momentul în care apare primul spațiu, sau am ajuns la sfârșitul *string*-ului, prelucrăm cuvântul, respectiv decidem dacă trebuie să-l scriem în fișier sau nu. Reluăm algoritmul pentru restul caracterelor din *string*-ul citit, apoi pentru restul liniilor din fișier.

```
Subalgoritm Litere:
```

```
cât timp nu urmează marca de sfârșit de fișier execută:
    citește t
                                                        { citim o linie din fișier }
                                                   { vom parcurge stringul citit }
    i \leftarrow 1
    cât timp i \le lungimea textului t execută:
       cât timp (i \le lungimea textului t) şi (t[i]=' ') execută:
         i \leftarrow i + 1
                                                          { spațiile le ignorăm }
       cuv ← ''
                                                 { vom construi cuvântul curent }
       maj \leftarrow 0
       cât timp (i ≤ lungimea textului t) și (t[i] ≠ ' ') execută:
          dacă (t[i] \ge 'A') și (t[i] \le 'Z') atunci
                                                                  { majusculă }
            maj \leftarrow maj + 1
         sfârșit dacă
         cuv ← cuv + t[i] { concatenăm caracterul curent în cuvântul curent }
         i \leftarrow i + 1
       sfârșit cât timp
       dacă maj > [(lungimea cuvântului cuv)/2] atunci
         scrie cuv
       sfârșit dacă
    sfârșit cât timp
  sfârșit cât timp
sfârsit subalgoritm
```

#### 10.5.2. Text

Vom reține fiecare propoziție într-un șir de caractere și vom construi șirul cuvintelor care fac parte din ea pentru a facilita căutarea cuvintelor care nu sunt unice în cadrul propoziției.

Din cauza că o propoziție se poate continua în fișierul de intrare pe mai multe linii, vom forma propozițiile, citind din fișier caractere. Notăm cu pr șirul de caractere care reține propoziția curentă și cu c caracterul citit din fișier. Inițializăm propoziția cu

stringul vid și caracterele citite le concatenăm în pr, până la apariția unui separator sau a marcajului de sfârșit de linie.

- În cazul în care propoziția se încheie (a apărut un separator) înlocuim separatorul, adăugând la sfârșitul variabilei *pr* un spațiu, pentru a simplifica împărțirea pe cuvinte. Trecem la analiza propoziției.
- Dacă pe parcursul citirii a apărut marcaj de sfârşit de linie, fără ca propoziția să se termine, acesta se citeşte din fişier, continuând citirea caracterelor şi concatenarea lor în aceeaşi propoziție de pe linia următoare.

Algoritmul care realizează formarea unei propoziții este:

```
Algoritm Text:
  pr<mark>←''</mark>
                                                  { inițializarea primei propoziției }
  cât timp nu urmează marca de sfârșit de fișier execută:
     citește c
       dacă c nu este separator de propoziții și
                                       nu urmează marca de sfârșit de linie atunci
                                             { alipim caracterul curent propoziției }
          pr \leftarrow pr + c
       altfel
          dacă c este separator de propoziții atunci
            pr ← pr + ' '
                                              { suprascriem separatorul cu spațiu }
                                                             { analiza propoziției }
            Analiza(pr)
            scrie marca de sfârșit de linie
            pr ← ''
                                              { inițializarea următoarei propoziții }
          altfel
            citește marca de sfârșit de linie
         sfârșit dacă
      sfârșit dacă
  sfârșit cât timp
sfârșit algoritm
```

Analiza propoziției reținute în șirul de caractere pr începe cu înlocuirea separatorilor cu spații, apoi se elimină spațiile inutile și se formează tabloul de cuvinte. Se scrie numărul n al cuvintelor depistate în fișierul de ieșire. Algoritmul Analiza (pr) apelează subalgoritmul distincte (cuvinte, dist) care numără cuvintele care apar o singură dată în propoziție și scrie acest număr în fișierul de ieșire.

De asemenea, se ține și evidența cuvintelor care se repetă și începând de la a doua apariție se înlocuiesc cu spațiu pentru a nu le mai lua în calcul.

După scrierea celor două numere în fișierul de ieșire se trece la căutarea cuvintelor care se repetă. Se numără fiecare apariție și se înlocuiește dublura lui cu spațiu. Dacă se găsesc cuvinte care se repetă, acestea se vor scrie în fișierul de ieșire.

```
{ analizăm o propoziție }
Subalgoritm Analiza(pr):
  suprascriem separatorii cu spații
  eliminăm spațiile în plus
  căutăm cuvinte, le reținem în șirul cuvinte și le ștergem din propoziție
                                            { numărul cuvintelor din propoziție }
  scrie n,' '
  distincte(cuvinte,dist)
                                           { apelăm subalgoritmul distincte }
  scrie dist,' '
  pentru i=1, n-1 execută:
                                      { numărul cuvintelor care nu sunt distincte }
    apare \leftarrow 1
    dacă cuvinte[i] # ' ' atunci
       pentru j=i+1,n execută:
         dacă cuvinte[i] = cuvinte[j] atunci
                             { un cuvânt este înlocuit cu spațiu la a doua apariție }
            cuvinte[j] \leftarrow ' '
            apare \leftarrow apare + 1
         sfârșit dacă
       sfârşit pentru
       dacă apare > 1 atunci
                                                     { cuvântul nu este distinct }
         scrie a[i],' '
       sfârșit dacă
    sfârșit dacă
  sfârşit pentru
sfârșit subalgoritm
```

În algoritmul care calculează numărul de cuvinte care apar o singură dată în propoziția pr se numără aparițiile fiecărui cuvânt în șirul cuvintelor. Dacă acesta apare o singură dată se incrementează contorul cuvintelor distincte.

#### 10.5.3. Traducere

Dificultatea acestei probleme constă în faptul că din textul inițial numai anumite cuvinte se înlocuiesc. Cele care trebuie înlocuite și traducerile lor apar în fișierul de intrare începând de pe linia a doua.

Se parcurge textul inițial *t* și se rețin pe rând cuvintele în variabila *cuvânt*. Pentru fiecare cuvânt *cuvânt* se verifică dacă acesta apare printre cuvintele care sunt scrise în dicționar. Dacă acesta există în dicționar, se reține traducerea lui, urmând ca aceasta să substituie cuvântul în textul inițial.

Căutăm cuvântul *cuvânt* în fișierul care conține dicționarul cu următorul subalgoritm:

```
Subalgoritm Caută (cuvânt):

citește marcajul de sfârșit de linie din fișierul de intrare
cât timp nu urmează marca de sfârșit de fișier execută:

citim în dictio o linie din fișier (o pereche de cuvinte)

căutăm cuvântul cuvânt în dictio

dacă l-am găsit începând cu poziția 1 atunci

ștergem din perechea de cuvinte originalul

returnăm traducerea

ieșim forțat din cât timp

altfel

returnăm stringul vid

sfârșit dacă

sfârșit cât timp

sfârșit subalgoritm
```

În fișierul de intrare prima linie este ocupată de textul care trebuie tradus. De fiecare dată când se caută traducerea unui cuvânt în fișier trebuie să se realizeze saltul peste prima linie, deci se citește doar marcajul de sfârșit de linie.

Selectarea cuvintelor care se caută în dictionar se face în felul următor:

```
Subalgoritm Traducere(t):
    cuvânt ← ''
    i ← 1
    cât timp i ≤ lungimea textului t execută:
        cât timp t[i] este literă execută:
        cuvânt ← cuvânt + t[i] { în cuvânt construim cuvinte din text }
        i ← i + 1
        sfârșit cât timp
```

```
trad ← traducerea lui cuvânt, returnat de subalgoritmul Caută (cuvânt)

dacă trad ≠ '' atunci { dacă există traducere }

inserăm traducerea cuvântului în t, începând cu poziția i

ștergem cuvântul de tradus din t

i ← i + lungimea cuvântului trad - lungimea cuvântului cuvânt + 1

{ recalculăm poziția din care continuăm prelucrarea lui t }

cuvânt ← '' { reinițializăm cuvânt pentru a fi util pentru un cuvânt nou }

altfel

i ← i+1 { avansăm, cuvântul nu trebuie tradus }

cuvânt ← ''

sfârșit dacă

sfârșit cât timp

sfârșit subalgoritm
```

#### 10.5.4. Culori

În rezolvare variabila num reprezintă tabloul care conține numele celor n copii, în tabloul pref păstrăm tabloul care conține numărul de ordine a culorii preferate de copii, iar cul este tabloul în care se reține numărul de apariții a fiecărei culori.

#### Exemplu

Fie tabloul num = (Pop Mihai, Popan Ana, Popa Ion, Popescu Marius, Rus Andrei, Rusu Teodor, Trif Ionut). Presupunem că există 3 culori (m = 3), iar șirul culorilor preferate de copii, în ordinea numelor din tabloul num este pref = (1, 1, 1, 1, 3, 3, 2).

Elementele tabloului cul se determină în paralel cu citirea. Acesta va conține pe poziția i numărul copiilor care preferă culoarea i, deci cul = (4, 1, 2) cu semnificația: culoarea 1 apare de patru ori, culoarea 2 apare o dată, culoarea 3 apare de două ori.

În algoritm vom căuta valoarea maximă în acest șir, apoi pe acei copii care au ales culoarea i, unde i este poziția pe care s-a găsit valoarea maximă. În exemplu, valoarea maximă în tabloul cul este egal cu 4 și se găsește pe poziția poz = 1. Această variabilă astfel reține codul unei culori. Se caută în tabloul pref indicii elementelor egale cu 1 (selectare). Aceștia sunt: Pop Mihai, Popan Ana, Popa Ion, Popescu Marius.

Acum trebuie să găsim grupul în care sunt mai puțini copii, dar care, după ce am pus primul grup deoparte, sunt cei mai numeroși. Înseamnă că ne interesează "al doi-lea maxim" din tabloul *cul*. Pentru a ușura găsirea acestuia, vom înlocui în tabloul *cul* valoarea 4 cu 0. La pasul următor maximul din tabloul *cul* = (0, 1, 2) este 2 pe poziția 3 (adică culoarea 3). În tabloul *pref* se caută toți indicii (copiii) elementelor de valoare *poz*. Aceștia sunt: Rus Andrei, Rusu Teodor.

Se înlocuiește valoarea 2 cu 0, deci cul = (0, 1, 0). Acum max se găsește pe poziția 2 și este egal cu 1. Se caută, pe baza tabloului pref toți copiii care preferă culoarea 2. Aceștia sunt: Trif Ionuț.

Se înlocuiește valoarea 1 cu 0 în tabloul cul, acesta devenind cul = (0, 0, 0) și procesul se încheie.

```
Subalgoritm Afişare:
                                             { copiii se grupează în m grupuri }
  pentru k=1, m execută:
                                   { maximul din șirul cul se află pe poziția poz }
    Maxim(m,cul,poz)
                                                              { codul culorii }
    scrie k,' '
    pentru i=1,n execută:
       dacă pref[i] = poz atunci { afișăm copiii care preferă culoarea poz }
         scrie num[i],' '
       sfârșit dacă
    sfârşit pentru
                                                  { eliminăm "ultimul maxim" }
    cul[poz] \leftarrow 0
  sfârşit pentru
sfârșit subalgoritm
```