

7. Szövegek kezelése

A kódtábla

Karakter alatt értjük a betűket, számjegyeket, írásjeleket és a különleges jeleket (pl. tabulátor). Minden karakternek megfeleltetünk egy számot (a karakter kódját), és a számítógép ezt a számot tárolja, illetve továbbítja.

A karakterek kódolására sokféle kódkészlet alakítható ki.

A ma általánosan elterjedt kódrendszer az UNICODE, amely valamennyi nép valamennyi jelét tartalmazza, de szerepelnek benne a Braille írás jelei, a sakktabla bábui, de még a rovásírás karakterei is. Az UNICODE-ot folyamatosan bővítik.

Az első 128 karakter alkotja az ASCII (American Standard Code for Information Interchange) kódtáblát, kezdetben a kódtábla kezdetben csak ezeket a karaktereket tartalmazta. (Pl. a 13-as kódú karakter a bekezdés vége, a 32-es a szóköz.):

000 NUL	016 DLE	032 SP	048 0	064 @	080 P	096 `	112 p
001 SOH	017 DC1	033 !	049 1	065 A	081 Q	097 a	113 q
002 STX	018 DC2	034 "	050 2	066 B	082 R	098 b	114 r
003 ETX	019 DC3	035 #	051 3	067 C	083 S	099 c	115 s
004 EOT	020 DC4	036 \$	052 4	068 D	084 T	100 d	116 t
005 ENQ	021 NAK	037 %	053 5	069 E	085 U	101 e	117 u
006 ACK	022 SYN	038 &	054 6	070 F	086 V	102 f	118 v
007 BEL	023 ETB	039 '	055 7	071 G	087 W	103 g	119 w
008 BS	024 CAN	040 (056 8	072 H	088 X	104 h	120 x
009 HT	025 EM	041)	057 9	073 I	089 Y	105 i	121 y
010 LF	026 SUB	042 *	058 :	074 J	090 Z	106 j	122 z
011 VT	027 ESC	043 +	059 ;	075 K	091 [107 k	123 {
012 FF	028 FS	044 ,	060 <	076 L	092 \	108 l	124
013 CR	029 GS	045 -	061 =	077 M	093]	109 m	125 }
014 SO	030 RS	046 .	062 >	078 N	094 ^	110 n	126 ~
015 SI	031 US	047 /	063 ?	079 O	095 _	111 o	127 DEL

Az ASCII karaktertábla kiíratása (32-től 127-ig):

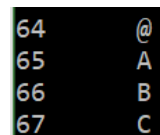
A karaktereket a **char** típus kezeli. Mivel minden karakternek (ami megjelenik a képernyőn) kölcsönösen egyértelmű módon megfeleltethető a karakter kódja (egy egész szám), a karakter és a karakter kódja könnyen „átváltható” egymásba.

Például, ha a karakter kódja az *i* egész szám, akkor a neki megfelelő karakter a (char)*i*, és fordítva, ha *ch* egy karakter, akkor a kódja (int)ch.

A (char)*i* típuskonverzió megadható így is: Convert.ToChar(*i*), illetve az (int)ch így is: Convert.ToInt32(*ch*).

A következő program kiírja a karaktereket és kódjaikat 32-től 127-ig:

```
Console.WriteLine ("ASCII kódok");
for (int i = 32; i < 128; i++)
    Console.WriteLine($"{i} \t {(char)i}");
```



```
64 @
65 A
66 B
67 C
```

A képernyőképen megjelenik a karakter kódja (*i*) majd egy tabulátorpozícióval (\t) elválasztva a karakter ((char) *i*)

Két különleges karakter jegyezzünk meg:

- \n - új sor karakter
- \t - tabulátor karakter

A string, mint karaktersorozat

A szöveg (string) egy karaktersorozat, a karakterek számozása 0-tól kezdődik. Egy string hosszát a *Length* tulajdonsága adja meg. A következő program beolvas egy szöveget, és kiírja annak karaktereit, illetve a karakterek kódját:

```
string szó;
Console.Write("Egy szó=");
szó = Console.ReadLine();
for (int i = 0; i < szó.Length; i++)
    Console.WriteLine($" {szó[i]} \t {(int)szó[i]} ");
```

Példánkban `szó[i]` a szó *i*-edik karaktere, ennek kódja az `(int) szó[i]`.

A karaktert aposztrófok közé kell tenni, ha hivatkozunk rá, a stringet macskakörmök közé. A következő program megszámolja a magánhangzókat egy nagybetűs angol szövegben:

```
Console.Write("Szó=");
string sz=Console.ReadLine();
int db=0;
for (int i=0; i<sz.Length; i++)
{
    char c=sz[i];
    if (c=='A' || c=='E' || c=='I' || c=='O' || c=='U')
        db++;
}
Console.Write("Szótagszám={0}", db);
```

Néhány stringművelet

Átalakítás nagybetűsre:

```
szoveg = szoveg.ToUpper()
```

kisbetűsre:

```
szoveg=szoveg.ToLower()
```

Szóközök eltávolítása:

```
szoveg=szoveg.Trim()
```

Karakterek cseréje. Például a `szoveg` stringben mindenhol lecseréli az `Á`-t `A`-ra:

```
szoveg = szoveg.Replace('Á', 'A')
```

Annak ellenőrzése, hogy egy karakter betűe:

```
if (Char.IsLetter(ch))
```