

6. Matlab - znaky a textové reťazce

Blaho Michal · MATLAB/Comsol

12.06.2009



V tretej časti nášho seriálu o Matlabe sme si popísali dátové typy, ktoré Matlab využíva. V krátkosti sme si povedali o znakových premenných a ako znakovú premennú vytvoríme. Dnes si znaky a textové reťazce popíšeme detailnejšie a ukážeme základnú prácu s nimi. Uvedieme aj zopár užitočných príkazov, ktoré pravdepodobne budete využívať pri manipulácii so znakovými a textovými premennými.

Znak je v Matlabe reprezentovaný ako celočíselná hodnota konvertovaná na jej Unicode znakový ekvivalent. Na deklarovanie znaku využijeme apostrofy, medzi ktoré napíšeme požadovaný znak. Základné znaky sú reprezentované číslami 32 - 127. Čísla menšie ako 32 sú netlačiteľné riadiace znaky. Ďalšie čísla reprezentujú znaky rôznych fontov na počítači. Číselnú hodnotu znaku môžeme potom získať napríklad príkazom **uint8**.

```
>>znak='P';
>>cislo=uint8(znak);
cislo=80
```

Matlab po zadaní týchto príkazov uloží do premennej znak písmeno P a do premennej cislo jeho číselný ekvivalent (v našom prípade číslo 80). Textové premenné, sa podobne ako znakové, deklarujú pomocou apostrofov. Text je uložený ako pole znakov. Číselné hodnoty jednotlivých znakov znova získame napríklad pomocou príkazu **uint8**. Čísla môžeme konvertovať na znaky alebo textové reťazce pomocou príkazu **char**.

```
>>retazec='Hello';
>>cisla=uint8(retazec);
cisla=[72 101 108 108 111]
```

```
>>portal=char([80 111 115 116 101 114 117 115]);
portal=Posterius
```

V premennej portal nájdete textový reťazec s menom nášho portálu. Spojenie reťazcov do jedného riadku môžeme urobiť pomocou hranatých zátvoriek a znaky, alebo reťazce oddeliť čiarkou, alebo medzerou. Spojenie v riadku môžeme urobiť aj pomocou príkazu

strcat.

```
>>spolu1=[retazec, ' ', portal,'!'];
spolu1=Hello Posterus!
```

Spojenie reťazcov v stĺpci sa robí podobne ako v predchádzajúcom príkaze, ale namiesto čiarky použijeme bodkočiarku. Pri vytváraní polí znakov treba dávať pozor nato, aby mali jednotlivé riadky rovnaký počet znakov. Konce krátkych znakov môžete dopĺňať medzerami. Príkaz **char** doplní prázdne znaky za vás. Na vertikálne spojenie textových reťazcov môžeme použiť príkaz **strvcat**.

```
>>spolu2=[retazec; portal]; %tento príkaz vyhlási chybu!
>>spolu2=['Hello  ';'Posterus'];
spolu2=
Hello      %za reťazcom sú ešte 3 medzery
Posterus
```

```
spolu3=char('Hello','Posterus');
spolu2=
Hello      %za reťazcom sú ešte 3 medzery
Posterus
```

Pomocou príkazu **blanks** môžeme vytvoriť prázdny reťazec s daným počtom medzier namiesto toho, aby sme museli stláčať niekoľko krát medzeru. Pre lepšiu názornosť ohraničíme prázdne znaky lomítkom

```
>>textblank=blanks(10);
>>text=['|',textblank,'|']
text=|          |
```

Znaky do reťazca môžeme vložiť aj pomocou maticových indexov. Prázdne miesta na začiatku a konci textového reťazca odstránime pomocou funkcie **strtrim**

```
>>textblank=blanks(10);
>>textblank(3:8)='Matlab'
>>text=['|',textblank,'|'];
text=|  Matlab  |
```

```
>>texttrim=strtrim(textblank);
>>text=['|',texttrim,'|']
text=|Matlab|
```

Zhrnutie príkazov na tvorbu textových reťazcov najdete v nasledujúcej tabuľke.

Funkcia	Popis
'str'	vytvorí znakové pole medzi apostrofmi
blanks	vytvorí pole medzier

strcat	spája znakové polia
strvcat	spája znakové polia vertikálne

Tab. Príkazy na vytvorenie polia znakov

Textový reťazec môžeme skontrolovať po znakoch, či obsahuje určitú sadu znakov. Väčšinou sa kontroluje, či textový reťazec obsahuje písmená, čísla, medzery alebo nejaké iné špeciálne znaky. Na nasledujúcich príkladoch si ukážeme hľadanie medzery a špeciálnych znakov.

```
>>text='Hello Posterus';
>>je=isspace(text);
je=[0 0 0 0 0 1 0 0 0 0 0 0 0 0]

>>je2=isstrprop(text,'upper');
je2=[1 0 0 0 0 0 1 0 0 0 0 0 0 0]

>>num2=[9 72 101 108 108 111 32 80 111 115 116 101 114 117 115 10];
>>str2=char(num2);
>>isstrprop(str2,'wspace');
```

Posledný príkaz vráti v našom prípade hodnotu 1 na pozíciach kde sú čísla 9, 32 a 10, čo sú znaky tabulátora, medzery a nového riadku. Znaky alebo reťazce môžeme identifikovať pomocou príkazov v nasledujúcej tabuľke.

Funkcia	Popis
ischar	vyhodnotí, či je argument pole znakov
isletter	testuje jednotlivé znaky na písmená
isspace	testuje jednotlivé znaky na medzery
isstrprop	testuje jednotlivé znaky na vlastnosti (zadávajú sa ako reťazec, viac manuál)

Tab. Príkazy na identifikáciu znakov alebo textových reťazcov

Často sa stáva, že v našom reťazci znakov potrebujeme nájsť nejaký kratší reťazec, prípadne porovnať či sa náš reťazec zhoduje s nejakým iným. Matlab má pripravených zopár príkazov na hľadanie a porovnávanie textových reťazcov. Z nich si ukážeme príkazy na vyhľadanie reťazca, porovnanie reťazcov a nájdenie tokena v reťazci.

```
>>text='Posterus';
>>kde=strfind(text,'s');
kde=[3 8] %pozícia písmena s v premennej text

>>rovne=strcmpi(text,'POSTERUS');
rovna=1 %reťazce sa zhodujú ignorujeme veľké a malé písmená

>>je=strncmp(text,'Post',4);
je=1 %na prvých 4 znakoch je textový reťazec Post
```

```
>>datum='12/06/2009';
>>[token,zvysok]=strtok(datum, '/');
token=12, zvysok=/06/2009
>>[token2,zvysok]=strtok(zvysok, '/')
token2=06, zvysok=/2009
>>[token3,zvysok]=strtok(zvysok, '/')
token3=2009, zvysok=Empty string: 1-by-0
```

V nasledujúcej tabuľke uvádzame ďalšie príkazy na hľadanie a porovnávanie reťazcov

Funkcia	Popis
findstr	hľadá pole znakov v druhom poli znakov
strcmp	porovnáva polia znakov
strcmpi	porovnáva polia znakov, ignoruje veľkosť znakov
strmatch	hľadá zadané pole znakov na začiatku iných polí znakov
strncmp	hľadá znakové pole v prvých N znakov iného znakového poľa
strncmpi	hľadá znakové pole v prvých N znakov iného poľa, ignoruje veľkosť
strtok	nájde token v poli znakov

Tab. Príkazy na hľadanie alebo porovnanie polí znakov

Ďalšia sada príkazov robí zmeny v textových reťazcoch. Vyúžívajú sa keď chceme náš textový reťazec upraviť do našej želanej podoby. V nasledujúcich príkazoch si ukážeme postupne odstránenie prázdnych miest, zmenu veľkosti písma, usporiadanie textu a nahradenie poľa znakov iným polom.

```
>>text='Posterus  '
>>text2=deblank(text);
text2=Posterus %odstráni zbytočné medzery na konci

>>text3=upper(text2);
text3=POSTERUS

>>text4=sort(text3);
text4=EOPRSSTU

>>text5='12/06/2009';
>>text6=strrep(text, '/', '.');
text6=12.06.2009;
```

Prehľad ďalších príkazov, ktorá sa dajú využiť na úpravu vašich reťazcov.

Funkcia	Popis
deblank	odstaňuje prázdne miesta
lower	zmení všetky písmená na malé
sort	usporiadava znaky zostupne alebo vzostupne

<i>strjust</i>	zarovná pole znakov
<i>strrep</i>	nahradí pole znakov iným poľom
<i>strtrim</i>	odstráni prázdne znaky na začiatku alebo na konci poľa znakov
<i>upper</i>	zmení všetky písmená na veľké

Tab. Príkazy na zmenu polí znakov

Matlab dokáže konvertovať číslo, ktoré je uložené ako textový reťazec priamo na požadovaný číselný dátový typ. Samozrejme, že to ide aj opačne, teda vieme konvertovať číslo, dokonca aj maticu na textový reťazec.

```
>>cislo=1.234;
>>text=num2str(cislo);
text=1.234 %premenná text je textový reťazec
```

```
>>matica=[1,2;3,4];
>> text=mat2str(matica);
text=[1 2;3 4] %premenná text je textový reťazec
```

```
>>str = '37.294e-1';
>> cislo=str2double(str);
cislo=3.7294 %číslo je dátový typ s pohyblivou čiarkou
```

V nasledujúcej tabuľke je súhrn príkazov na prevod medzi číslami a textovými reťazcami a naopak. Niektoré sme si už ukázali v úvode tohoto článku.

Funkcia	Popis
<i>char</i>	konvertuje na znakové pole
<i>double</i>	konvertuje pole znakov na číselné hodnoty
<i>int2str</i>	konvertuje celé číslo na znakové pole
<i>mat2str</i>	konvertuje maticu na znakové polia
<i>num2str</i>	konvertuje číslo na znakové pole
<i>str2num</i>	konvertuje znakové pole na číselnú hodnotu
<i>str2double</i>	konvertuje znakové pole na číslo s pohyblivou čiarkou

Tab. Príkazy na prevod medzi číselnými a znakovými typmi

V tejto časti sme si popísali základnú prácu so znakmi a textovými reťazcami. V nasledujúcej časti si ukážeme ako naformátovať výstup v Matlabe.

