

ԱՇԽԱՏԱՆՔ ՏՈՂԵՐԻ ՀԵՏ

Մաս 1

Յունիկոդը և PHP –ն

Տողերը և դրանց մշակման ֆունկցիաները ծրագրավորման սկրիպտային լեզուների հիմնական գործիքներից են, հատկապես, եթե կիրառական գործունեությունը WEB- ի մշակումն է: Հետևաբար, մեծ ուշադրություն է դարձվում PHP-ում տող տիպի փոփոխականների հետ աշխատանքին:

PHP- ում տողերը դիտվում են որպես զանգված, որի տարրերը նիշեր են:

Օրինակ 1.

```
<?php
```

```
    $english = "ARM";
```

```
    echo $english[0]. '<br />';
```

```
    echo $english[1]. '<br />';
```

```
    echo $english[2]. '<br />';
```

```
?>
```

Արդյունքում կստանանք՝

A

R

M

Դիտարկենք մեկ այլ օրինակ, որտեղ փոփոխականին վերագրվում է հայերեն տեքստ:

Օրինակ 2.

```
<?php
```

```
    $armenian = "ՀԱՅ";
```

```
    echo $armenian[0]. '<br />';
```

```
echo $armenian[1]. '<br />';
```

```
echo $armenian[2]. '<br />';
```

```
?>
```

Արդյունքում կստանանք՝

❖

❖

❖

Յունիկոդ (UTF-8)

Mbstring ընդլայնումը

Այսօր բոլոր նոր ծրագրերը կենտրոնանում են UTF-8 կոդավորման վրա, որը դե ֆակտո ստանդարտ է աշխարհի բոլոր լեզուների կոդավորման համար: UTF-8- ին անցումը ոչ միայն հնարավորություն է տալիս մեկ փաստաթղթում տարբեր լեզուներով տեքստեր օգտագործել, այլ նաև թույլ է տալիս ընդլայնել կոդավորումը, երբ հայտնվում են նոր նիշեր:

PHP 7-ում UTF-8 կոդավորման նիշերը կարող են նշվել հատուկ շարահյուսության միջոցով՝ \u հաջորդականությանը հաջորդում է ձևավոր փակագծերում նիշի տասնվեցական ծածկագիրը.

```
<?php
```

```
echo "\u{0531}";
```

```
?>
```

Կարտաձվի հայերեն Ա մեծատառը:


Յունիքոդ ստանդարտում հայկական այբուբենին ամրագրված է U+0530—U+058F տիրույթը, որից այսօր օգտագործվում են.

U + 0531-U + 0556 -մեծատառեր;

U + 0561-U+0587 -փոքրատառեր;

U + 0559-U + 055F -կետադրական նշաններ;

U + 0589-U + 058A -կետադրական նշաններ:

Հայերենի տառեր և կետադրական նշաններ Յունիկոդում 																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+053x		Ա	Բ	Գ	Դ	Ե	Զ	Է	Ը	Թ	Ժ	Ի	Լ	Խ	Ճ	Կ
U+054x	Հ	Ձ	Ղ	Ճ	Մ	Յ	Ն	Ծ	Ո	Չ	Պ	Ջ	Ռ	Ս	Վ	Տ
U+055x	Ր	Յ	Ի	Փ	Ք	Օ	Ֆ			՝	՛	՛	՛	՛	՛	՛
U+056x	□	ա	բ	գ	դ	ե	զ	է	ը	թ	ժ	ի	լ	խ	ծ	կ
U+057x	հ	ձ	ղ	ճ	մ	յ	ն	շ	ո	չ	պ	ջ	ռ	ս	վ	տ
U+058x	ր	ց	ւ	փ	ք	օ	ֆ	և	□	:	-			ճ	ճ	ճ
U+201x	-					—										
U+202x													

UTF-8 -ում անգլերեն նիշերը զբաղեցնում են մեկ բայթ, իսկ հայերենը՝ երկու: Արդյունքում, եթե UTF-8 տողը դիտարկենք որպես նիշերի զանգված և արտածենք, անգլերեն լեզվի նիշերը նորմալ կարտածվեն, իսկ հայերեն նիշերի միայն կեսը կարտածվի (տես օրինակը 2-ը):

Գործնականում խնդիրը լուծելու համար սովորաբար միացնում են mbstring ընդլայնումը, որն ապահովում է բազմաբայթային կոդավորումը, կամ ուղղակիորեն օգտագործում են mbstring ֆունկցիաները, կամ կարգավորում են PHP - ն, որպեսզի PHP- ի տողերի հետ աշխատող ստանդարտ ֆունկցիաները փոխարինվեն mbstring անալոգներով:

Windows- ում mbstring ընդլայնումը միացնելու համար հարկավոր է php.ini կազմաձևման ֆայլում հեռացնել կետ-ստորակետները հետևյալ երկու տողերից.

```
;extension dir = "ext"
```

```
;extension=php_mbstring.dll
```

Դրանից հետո պետք է վերագործարկել սերվերը, որպեսզի այն կրկին կարդա php.ini կազմաձևման ֆայլը:

Օրինակ.

```
<?php
```

```
    $english = "Chess";
```

```
    $armenian = "Շախմատ";
```

```
    print "strlen() = " . strlen($english) . " $english -ի համար չ " .
```

```
    strlen($armenian) . " $armenian -ի համար <br/>";
```

```
    print "mb_strlen() = " . mb_strlen($english) . " $english -ի համար չ " .
```

```
    mb_strlen($armenian) . " $armenian -ի համար <br/>";
```

```
?>
```

Հայերեն այբուբենի տառերից յուրաքանչյուրը ներկայացնելու համար պահանջվում է երկու բայթ, ուստի, օրինակի կողմը իրականացնելիս արտածվում է հետևյալ արդյունքը.

```
strlen() = 5 Chess -ի համար և 12 Շախմատ -ի համար
```

```
mb_strlen() = 5 Chess -ի համար և 6 Շախմատ -ի համար
```

Այն գործողությունները, որոնք կախված են տողերում նիշերի տեղադրությունից (օրինակ, ենթատողերի դուրս բերումը), պետք է կատարվեն հաշվի առնելով սիմվոլիկ, այլ ոչ թե բայթային պատկերացումը:

Բերենք օրինակ, որում ցուցադրվում է թե ինչպես կարելի է օգտագործողի կողմից մուտքագրված հաղորդագրությունից առաջին 30 բայթը կորզել substr() և mb_substr() ֆունկցիաների միջոցով:

```
<?php
```

```
    $message = "Armenia - Հայաստան, Yerevan - Երևան";
```

```
    print "substr() - " . substr($message, 0, 30) . "<br />";
```

```
    print "mb_substr() - " . mb_substr($message, 0, 30) . "<br />";
```

```
?>
```

Արդյունքում կարտածվի`

`substr()` - Armenia - Հայաստան, Ye

`mb_substr()` - Armenia - Հայաստան, Yerevan –