

ustrsplit() — Split string into parts based on a Unicode regular expression

Description
Also see

Syntax

Remarks and examples

Conformability

Description

`ustrsplit(s, ustrregexp)` returns the contents of *s* split into parts based on *ustrregexp*.

Syntax

string rowvector `ustrsplit(string scalar s, string scalar ustrregexp)`

Remarks and examples

stata.com

`ustrsplit()` splits a string into parts according to a Unicode regular expression.

For example,

```
ustrsplit("$12.31 €6.75", "[$€]") = ("", "12.31 ", "6.75")
```

Note that the first element of the result is an empty string. This is because `ustrsplit()` encountered the first “\$” in the string being split and an empty string (that is, nothing) was to the left of that “\$”. Assuming you put the above result into a string scalar named “`result`”, you can type the following to remove that leading empty string and any other all-whitespace parts:

```
select(result, strlen(ustrtrim(result))!=0)
```

The example above splits on any dollar sign or Euro symbol. To split on any Unicode character that is considered a currency symbol, we can use the regular expression `[\p{Sc}]`,

```
ustrsplit("$12.31 €6.75 ¥100.50 ₩25.45", "[\p{Sc}]")
```

which returns

```
("", "12.31 ", "6.75 ", "100.50 ", "25.45")
```

Conformability

`ustrsplit(s, ustrregexp)`:

<i>s</i> :	1×1
<i>ustrregexp</i> :	1×1
<i>result</i> :	1×1 , $w =$ number of parts in <i>s</i>

Also see

[M-5] [ustrword\(\)](#) — Obtain Unicode word from Unicode string

[M-4] [String](#) — String manipulation functions

[P] [tokenize](#) — Divide strings into tokens

[U] [12.4.2 Handling Unicode strings](#)