

Description

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

Syntax

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

Remarks and examples

`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, ustrregex)`:

<i>s</i> :	1×1
<i>ustrregex</i> :	1×1
<i>result</i> :	1×1 , $w = \text{number of parts in } s$

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**

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow and NetCourseNow are trademarks of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.

For suggested citations, see the FAQ on [citing Stata documentation](#).

