

## String — String manipulation functions

[Contents](#)[Description](#)[Remarks and examples](#)[Also see](#)

## Contents

[M-5] Manual entry	Function	Purpose
<b>Parsing</b>		
<b>tokens()</b>	<code>tokens()</code>	obtain tokens (words) from string
<b>invtokens()</b>	<code>invtokens()</code>	concatenate string vector into string scalar
<b>ustrword()</b>	<code>ustrword()</code>	return <i>n</i> th Unicode word
	<code>ustrwordcount()</code>	return the number of Unicode words
<b>strmatch()</b>	<code>strmatch()</code>	pattern matching
<b>tokenget()</b>	...	advanced parsing
<b>ustrsplit()</b>	<code>ustrsplit()</code>	split string into parts based on a Unicode regular expression
<b>Length &amp; position</b>		
<b>strlen()</b>	<code>strlen()</code>	length of string in bytes
<b>ustrlen()</b>	<code>ustrlen()</code>	length of string in Unicode characters
<b>udstrlen()</b>	<code>udstrlen()</code>	length of string in display columns
<b>fmtwidth()</b>	<code>fmtwidth()</code>	width of <i>%fmt</i>
<b>strpos()</b>	<code>strpos()</code>	find substring within string from left
	<code>strrpos()</code>	find substring within string from right
<b>ustrpos()</b>	<code>ustrpos()</code>	find Unicode substring within string, first occurrence
	<code>ustrrpos()</code>	find Unicode substring within string, last occurrence
<b>indexnot()</b>	<code>indexnot()</code>	find character not in list
<b>Editing</b>		
<b>substr()</b>	<code>substr()</code>	extract substring
<b>usubstr()</b>	<code>usubstr()</code>	extract Unicode substring
<b>udsubstr()</b>	<code>udsubstr()</code>	extract Unicode substring based on display columns

Editing, <i>continued</i>
---------------------------

---

<b>strupper()</b>	strupper() strlower() strproper()	convert to uppercase convert to lowercase convert to proper case
<b>ustrupper()</b>	ustrupper() ustrlower() ustrtitle()	convert Unicode characters to uppercase convert Unicode characters to lowercase convert Unicode characters to titlecase
<b>strtrim()</b>	stritrim()  strltrim() strrtrim() strtrim()	replace multiple, consecutive internal blanks with one blank  remove leading blanks remove trailing blanks remove leading and trailing blanks
<b>ustrtrim()</b>	ustrtrim()  ustrltrim()  ustrrtrim()	remove leading and trailing Unicode whitespace characters and blanks  remove leading Unicode whitespace characters and blanks  remove trailing Unicode whitespace characters and blanks
<b>subinstr()</b>	subinstr() subinword()	substitute text substitute word
<b>usubinstr()</b>	usubinstr()	replace Unicode substring
<b>_substr()</b>	_substr()	substitute into string
<b>_usubstr()</b>	_usubstr()	substitute into Unicode string
<b>strdup()</b>	*	duplicate string
<b>strreverse()</b>	strreverse()	reverse string in bytes
<b>ustrreverse()</b>	ustrreverse()	reverse string in Unicode characters
<b>soundex()</b>	soundex() soundex_nara()	convert to soundex code convert to U.S. Census soundex code

Stata
-------

---

<b>abbrev()</b>	abbrev()	abbreviate Unicode strings to display columns
<b>strtoname()</b>	strtoname()	translate strings to Stata 13 compatible names
<b>ustrtoname()</b>	ustrtoname()	translate Unicode strings to Stata names

## Text translation

---

<b>strotoreal()</b>	strotoreal()	convert real to string
<b>strtooreal()</b>	strtooreal()	convert string to real
<b>ustrto()</b>	ustrto()	convert a Unicode string to a string in another encoding
	ustrfrom()	convert a string in one encoding to a Unicode string
<b>ustrunescape()</b>	ustrunescape()	convert the escaped hex sequences to Unicode
	ustrtohex()	convert a Unicode sequence to hex sequences
<b>urlencode()</b>	urlencode()	convert a string to a valid ASCII format for web transmission
	urldecode()	decode the string obtained from urlencode()
<b>ascii()</b>	ascii()	obtain ASCII or byte codes of string
	char()	make string from ASCII or byte codes
<b>uchar()</b>	uchar()	make Unicode character from Unicode code-point value
<b>isascii()</b>	isascii()	whether string scalar contains only ASCII codes

## Unicode utilities

---

<b>ustrcompare()</b>	ustrcompare()	compare or sort Unicode strings
	ustrsortkey()	obtain sort key of Unicode string
<b>ustrfix()</b>	ustrfix()	replace invalid sequences in Unicode string
<b>ustrnormalize()</b>	ustrnormalize()	normalize Unicode string

## Base64 utilities

---

<b>base64encode()</b>	base64encode()	encode a string into Base64 format
	base64decode()	decode a Base64-encoded string
	base64encodefile()	encode text in a file into Base64 format
	base64decodefile()	decode Base64-encoded text in a file

---

## Description

The above functions are for manipulating strings. Strings in Mata are strings of Unicode characters in [UTF-8 encoding](#), usually the printable characters, but Mata enforces no such restriction. In particular, strings may contain binary 0.

## Remarks and examples

In addition to the above functions, two operators are especially useful for dealing with strings.

The first is `+`. Addition is how you concatenate strings:

```
: "abc" + "def"
abcdef
: "Café " + "de Flore"
Café de Flore
: command = "list"
: args = "mpg weight"
: result = command + " " + args
: result
list mpg weight
```

The second is `*`. Multiplication is how you duplicate strings:

```
: 5*"a"
aaaaa
: "Allô"*2
AllôAllô
: indent = 20
: title = indent*" " + "My Title"
: title
                My Title
```

## Also see

[M-4] [Intro](#) — Categorical guide to Mata functions

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 © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.



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