usubstr() — Extract Unicode substring

Description

usubstr(s, n1, n2) returns the Unicode substring of s, starting at Unicode character n1, for a length of n2. If n1 < 0, n1 is interpreted as the distance from the last Unicode character of s; if n2 = . (missing), the remaining portion of the Unicode string is returned.

ustrleft(s, n) returns the first n Unicode characters of the Unicode string s.

ustrright(s, n) returns the last n Unicode characters of the Unicode string s.

When arguments are not scalar, the functions return element-by-element results.

Syntax

string matrix usubstr(string matrix s, real matrix n1, real matrix n2)
string matrix ustrleft(string matrix s, real matrix n)
string matrix ustrright(string matrix s, real matrix n)

Remarks and examples

n ≤ 0 is interpreted as the distance from the end of the Unicode string; n = −1 means the distance starting at the last Unicode character.

An invalid UTF-8 sequence is replaced with a Unicode replacement character \ufffd. Null terminator char(0) in a binary string is a valid UTF-8 character and will be counted and treated as such.

Use udsubstr() to extract a substring based on display columns. Use substr() to extract a substring based on bytes.

Conformability

usubstr(s, b, l):

s: r × c
b: r × c or 1 × 1
l: r × c or 1 × 1
result: r × c
Diagnostics

`usubstr(s, b, l), ustrleft(s, b, l), and ustrright(s, b, l)` return an empty string if an error occurs.

Also see

[M-5] `subinstr()` — Substitute text
[M-5] `substr()` — Extract substring
[M-5] `substr()` — Substitute into string
[M-5] `usubinstr()` — Replace Unicode substring
[M-5] `usubstr()` — Substitute into Unicode string
[M-4] `String` — String manipulation functions
[U] 12.4.2 Handling Unicode strings