

1 Conversion to Raw Vectors

base

`charToRaw()` — single string to a raw vector only

(none)

```
charToRaw("aA1")
```

```
## [1] 61 41 31
```

stringr

stringi

`stri_encode()` with argument `to_raw=TRUE` is vectorized over the first argument; it returns a list of raw vectors.

```
stri_encode("aA1", "", "", to_raw=TRUE)[[1]]
```

```
## [1] 61 41 31
```

```
stri_encode(c("aA1", " "), "", "", to_raw=TRUE)
```

```
## [[1]]
```

```
## [1] 61 41 31
```

```
##
```

```
## [[2]]
```

```
## [1] 20
```

TODO: add `stri_enc_toraw()` (?)

Performance comparison:

```
test1 <- "abcdefghijklmnopqrstuvwxy"
microbenchmark(charToRaw(test1), stri_encode(test1, "", "", to_raw=TRUE)[[1]])

## Unit: nanoseconds
##
##          expr   min      lq median      uq    max neval
## charToRaw(test1) 474  629.5  942.0 1081.5 26494   100
## stri_encode(test1, "", "", to_raw = TRUE)[[1]] 8462 8876.0 9205.5 9581.5 257945   100

test2 <- rep(c("abcdefghijklmnopqrstuvwxy", "ABCDEFGHJKLMNOPQRSTUVWXYZ", "0123456789"), 10)
microbenchmark(lapply(test2, charToRaw), stri_encode(test2, "", "", to_raw=TRUE))

## Unit: microseconds
##
##          expr   min      lq median      uq    max neval
## lapply(test2, charToRaw) 24.832 30.7825 34.1955 36.0160 249.735   100
## stri_encode(test2, "", "", to_raw = TRUE) 27.963 29.1155 29.9735 31.2885 114.889   100
```

2 Conversion from Raw Vectors

_____ base _____

`rawToChar()` — single raw vector to a single string only *(none)*

```
rawToChar(as.raw(c(97, 65, 49)))
## [1] "aA1"
```

_____ stringr _____

_____ stringi _____

`stri_encode()` also accepts a raw vector or a list of raw vectors as its first argument; by default, i.e. when `to_raw=FALSE`, the result is a character vector.

```
stri_encode(as.raw(c(97, 65, 49)), "")
## [1] "aA1"
stri_encode(list(as.raw(c(97, 65, 49)),
  as.raw(32)), "")
## [1] "aA1" " "
```

TODO: add `stri_enc_fromraw()` (?)

Performance comparison:

```
test1 <- as.raw(97:122)
microbenchmark(rawToChar(test1), stri_encode(test1, ""))

## Unit: nanoseconds
##          expr    min      lq median      uq     max neval
##  rawToChar(test1)  802   908.5   1232 1380.5  24935   100
##  stri_encode(test1, "") 4004 4206.5   4321 4603.5 159494   100

test2 <- rep(list(as.raw(97:122), as.raw(65:90), as.raw(48:57)), 10)
microbenchmark(lapply(test2, rawToChar), stri_encode(test2, ""))

## Unit: microseconds
##          expr      min       lq median      uq     max neval
##  lapply(test2, rawToChar) 33.956 40.0195 42.468 45.448 101.053   100
##  stri_encode(test2, "") 19.399 20.3780 20.888 21.477  60.833   100
```

3 Conversion to Integer Vectors (i.e. UTF-32)

base	stringr	stringi
<p><code>utf8ToInt()</code> — single string in UTF-8 to an integer vector only</p> <pre>utf8ToInt(enc2utf8("aA1")) ## [1] 97 65 49</pre>	<p>(none)</p>	<p><code>stri_enc_toutf32()</code> accepts a character vector on input and returns a list of integer vectors; like in all other functions from our package, native and UTF-8 encodings are handled properly</p> <pre>stri_enc_toutf32("aA1")[[1]] ## [1] 97 65 49 stri_enc_toutf32(c("aA1", " ")) ## [[1]] ## [1] 97 65 49 ## ## [[2]] ## [1] 32</pre>

Performance comparison:

```
test1 <- enc2utf8("abcdefghijklmnopqrstuvwxyz")
microbenchmark(utf8ToInt(test1), stri_enc_toutf32(test1)[[1]])

## Unit: nanoseconds
##           expr   min      lq median      uq   max neval
##   utf8ToInt(test1) 550  698.5  838.5  927.0 8137   100
## stri_enc_toutf32(test1)[[1]] 2141 2244.5 2314.5 2422.5 62685   100

test2 <- enc2utf8(rep(c("abcdefghijklmnopqrstuvwxyz", "ABCDEFGHIJKLMNOPQRSTUVWXYZ", "0123456789"), 10))
microbenchmark(lapply(test2, utf8ToInt), stri_enc_toutf32(test2))

## Unit: microseconds
##           expr    min      lq median      uq   max neval
## lapply(test2, utf8ToInt) 31.083 35.6960 38.3550 41.3805 116.385   100
## stri_enc_toutf32(test2)  7.899  9.0695  9.7565 12.8870  31.685   100
```

4 Conversion from Integer Vectors (i.e. UTF-32)

base	stringr	stringi
<p><code>intToUtf8()</code> — single integer vector to a single string only</p> <pre>intToUtf8(c(97L, 65L, 49L)) ## [1] "aA1"</pre>	<p>(none)</p>	<p><code>stri_enc_fromutf32()</code> a single integer vector or a list of integer vectors as its argument; the result is a UTF-8-encoded character vector.</p> <pre>stri_enc_fromutf32(c(97L, 65L, 49L)) ## [1] "aA1" stri_enc_fromutf32(list(c(97L, 65L, 49L), 32L)) ## [1] "aA1" " "</pre>

Performance comparison:

```
test1 <- 97:122
microbenchmark(intToUtf8(test1), stri_enc_fromutf32(test1))

## Unit: microseconds
##           expr    min      lq  median      uq    max neval
##   intToUtf8(test1) 1.293  1.408  1.4875  1.6225 30.481   100
##  stri_enc_fromutf32(test1) 1.711  1.764  1.8550  1.9560 89.277   100

test2 <- rep(list(97:122, 65:90, 48:57), 10)
microbenchmark(lapply(test2, intToUtf8), stri_enc_fromutf32(test2))

## Unit: microseconds
##           expr    min      lq  median      uq    max neval
##  lapply(test2, intToUtf8) 45.814 52.8155 55.6190 62.6930 106.162   100
##  stri_enc_fromutf32(test2)  6.255  6.7910  7.1695  7.9545  24.922   100
```