

Package: otter (via r-universe)

October 11, 2026

Title Ott's Utility Functions

Version 0.1.0

Description A growing collection of personal utility functions.
Currently provides tools to parse Rich Text Format (RTF) files and extract their tables into data frames, automatically detecting header rows, merging multi-page tables, and resolving merged cells. Particularly useful for tables produced by SAS or by the 'r2rtf' package, which are commonly used for clinical trial and regulatory reporting.

License MIT + file LICENSE

URL <https://github.com/ottvahtrik/otter>,
<https://ottvahtrik.github.io/otter/>

BugReports <https://github.com/ottvahtrik/otter/issues>

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Suggests lintr, testthat (>= 3.0.0)

Config/testthat/edition 3

Repository <https://ottvahtrik.r-universe.dev>

Date/Publication 2026-09-10 11:51:02 UTC

RemoteUrl <https://github.com/ottvahtrik/otter>

RemoteRef HEAD

RemoteSha 0ecfea9a43a10f0128545089692407bc5548acc2

Contents

rtf_to_df	2
Index	4

rtf_to_df

Parse an RTF file and extract table body cells into a data frame

Description

Reads an RTF file and returns a data.frame of the body cell values.

Usage

```
rtf_to_df(
  path,
  include_header = FALSE,
  header_row = NULL,
  encoding = "UTF-8",
  table_index = NULL
)
```

Arguments

path	Character scalar. Path to the .rtf file.
include_header	Logical scalar. If TRUE, column names are taken from the detected header row. When no header row is auto-detected, the first body row is used as column names and dropped from the data. Default: FALSE
header_row	Integer scalar. Which header row to use as column names when include_header = TRUE and there are multiple detected header rows. Defaults to the last header row. Use 1L for the first header row. Default: NULL
encoding	Character scalar. Encoding passed to <code>readLines()</code> when reading the RTF file. Use "latin1" for older SAS RTF that is not UTF-8. Default: "UTF-8"
table_index	Integer scalar. Index of the table to extract when the RTF contains multiple distinct tables. 1L selects the first table, 2L the second, etc. NULL merges all tables. Default: NULL

Details

Header rows are detected automatically via \trhdr (explicit; Word/SAS) or via cell vertical-alignment hints (\clvertalb / \clvertalt) used by `{r2rtf}`.

When include_header = TRUE and no header rows are auto-detected, the first body row is promoted to column names and removed from the data (analogous to `read.csv(header = TRUE)`). Footnote and source rows are excluded by their cell count (they span the full table width as a single cell).

When an RTF file contains multiple logically distinct tables (separated by paragraph text), use table_index to select a specific one. The default (NULL) merges all tables into a single data frame, which is the correct behaviour for paginated single-table outputs.

Value

A `data.frame` with one row per body row and one column per cell. All values are character strings. Column names are V1, V2, ... unless `include_header = TRUE`.

Examples

```
path <- system.file("extdata", "example.rtf", package = "otter")
rtf_to_df(path)
rtf_to_df(path, include_header = TRUE)

multi_path <- system.file("extdata", "multi_table_example.rtf", package = "otter")
rtf_to_df(multi_path, include_header = TRUE, table_index = 2L)
```

Index

`readLines()`, [2](#)
`rtf_to_df`, [2](#)