Package 'rtfqc'

November 29, 2026

Title Read Clinical RTF Outputs and Check Them Against QC Data

Version 0.1.15

Description Reads RTF (Rich Text Format) tables, listings and figures
produced for clinical study reports (for example by 'SAS' ODS RTF or
'r2rtf') directly, without 'Word', 'LibreOffice' or 'Pandoc'. Splits them into pages with titles,
headers, table body and footnotes, rebuilds the body table as a data
frame, and compares it cell by cell with an independently programmed
QC dataset, classifying each difference as a value difference,
truncation, rounding, number format or spacing. Also checks
``Page x of y'' numbering, blank pages, and titles and footnotes
against a table of contents. Works on single files or whole folders.

License GPL (>= 3)

Encoding UTF-8

Language en-US

Depends R (>= 4.1.0)

Imports utils

Suggests haven, openxlsx2, readxl, testthat (>= 3.0.0)

Config/testthat/edition 3

URL https://github.com/XianhuaZeng/rtfqc

BugReports https://github.com/XianhuaZeng/rtfqc/issues

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Xianhua Zeng [aut, cre]

Maintainer Xianhua Zeng <huazizeng@gmail.com>

Repository CRAN

Date/Publication 2026-11-29 10:40:02 UTC

Contents

rtf_cell_diff 2
rtf_check 3
rtf_check_pages 3
rtf_check_titles 4
rtf_compare 5
rtf_compare_report 6
rtf_qc_dir 6
rtf_read 7
rtf_table 8

Index **9**

`rtf_cell_diff` *Classify the difference between two table cells*

Description

Classify the difference between two table cells

Usage

```
rtf_cell_diff(a, b, min_trunc = 8L)
```

Arguments

a, b	Cell texts (RTF and QC). Vectorised.
min_trunc	Shortest common prefix (characters) for a difference to be reported as truncation rather than a value difference.

Value

Character vector: ‘""‘ (same; trailing white space and line breaks inside the cell are ignored), ‘"spacing"‘ (only the spaces differ: inside the text, e.g. ‘9(31.0)‘ vs ‘9 (31.0)‘, or leading spaces / indentation, e.g. ‘" n"‘ vs ‘"n"‘), ‘"truncated"‘ (one text is a prefix of the other), ‘"format"‘ (same numbers written differently, e.g. ‘5‘ vs ‘5.0‘), ‘"rounding"‘ (the numbers differ by at most one unit in the last decimal shown; whole numbers such as counts never count as rounding) or ‘"diff"‘.

Examples

```
rtf_cell_diff(c("12 (40.0)", "12 (40.0)", "3.46", "Headache and migr"),
              c("12 (40.0)", "12 (40.00)", "3.45", "Headache and migraine"))
```

rtf_check *Run all checks on one RTF output*

Description

Run all checks on one RTF output

Usage

```
rtf_check(x, qc = NULL, titles = NULL, footnotes = NULL, toc = NULL)
```

Arguments

x	Path to an RTF file or an ‘rtf_doc‘.
qc	Optional QC data (see [rtf_compare()]); when given, value differences are added to the issues.
titles	Character vector of expected title lines (for example from your TOC / shell). ‘NULL‘ skips the title check.
footnotes	Character vector of expected footnote lines. ‘NULL‘ skips it.
toc	Optional data frame (or path to an ‘.xlsx‘ / ‘.csv‘ file) with one row per output, used instead of ‘titles‘ / ‘footnotes‘. It needs a column naming the output (‘output‘, ‘name‘, ‘display‘ or ‘table‘, matched to the RTF file name without extension) and title / footnote columns (names starting with ‘title‘ or ‘footnote‘).

Value

A data frame of issues with columns ‘check‘, ‘page‘, ‘message‘.

Examples

```
f <- system.file("extdata", "demo.rtf", package = "rtfqc")
qc <- data.frame(COL1 = c("Age (years)", "  n"), COL2 = c("", "30"), COL3 = c("", "31"))
rtf_check(f, qc = qc, titles = "Table 14.1.1 Demographic Characteristics")
```

rtf_check_pages *Check page numbering and blank pages*

Description

Looks for the page number in the header, footer and body of every page. It can be written as text ("Page x of y", "Page x/y" or the Chinese "di x ye gong y ye" form): then page k must say "Page k of n", where n is the real number of pages. Or it can be Word PAGE / NUMPAGES fields, the usual SAS ODS RTF and r2rtf output: Word fills the numbers in when it opens the file, so they only need to be present on every page. Also reports pages that have a header but no body content.

Usage

```
rtf_check_pages(x, require = TRUE)
```

Arguments

x	Path to an RTF file or an 'rtf_doc'.
require	If 'TRUE' (default), a file with no page number at all is reported too; set 'FALSE' for outputs that are meant to have none.

Value

A data frame of issues ('check', 'page', 'message'); zero rows when everything is fine.

Examples

```
f <- system.file("extdata", "demo.rtf", package = "rtfqc")
rtf_check_pages(f)
```

---

rtf_check_titles *Check titles and footnotes against the table of contents*

Description

Every expected title and footnote line must appear on every page. The comparison ignores case, spaces and punctuation, so line breaks and different dash characters do not count as differences. Footnotes are also checked to be the same on all pages.

Usage

```
rtf_check_titles(x, titles = NULL, footnotes = NULL, toc = NULL)
```

Arguments

x	Path to an RTF file or an 'rtf_doc'.
titles	Character vector of expected title lines (for example from your TOC / shell). 'NULL' skips the title check.
footnotes	Character vector of expected footnote lines. 'NULL' skips it.
toc	Optional data frame (or path to an '.xlsx' / '.csv' file) with one row per output, used instead of 'titles' / 'footnotes'. It needs a column naming the output ('output', 'name', 'display' or 'table', matched to the RTF file name without extension) and title / footnote columns (names starting with 'title' or 'footnote').

Value

A data frame of issues ('check', 'page', 'message').

Examples

```
f <- system.file("extdata", "demo.rtf", package = "rtfqc")
rtf_check_titles(f,
  titles = c("Table 14.1.1 Demographic Characteristics", "Safety Population"),
  footnotes = c("Note: Percentages are based on N.", "Source: ADSL"))
```

`rtf_compare` *Compare an RTF output with its QC dataset*

Description

Leading empty cells (values a listing leaves out because they repeat the row above) are filled from the row above on both sides. Rows are then aligned: identical rows first, then on the first column (the row labels), then the remaining rows in order, so an extra or missing row does not shift every row after it. Every cell is classified with [rtf_cell_diff()].

Usage

```
rtf_compare(rtf, qc, qc_cols = NULL)
```

Arguments

`rtf`	Path to the RTF file, an ‘rtf_doc‘ from [rtf_read()], or a data frame such as the result of [rtf_table()] or ‘artful::rtf_to_df()‘.
`qc`	QC data: a data frame or a path to ‘.sas7bdat‘, ‘.xpt‘, ‘.rds‘ or ‘.csv‘. Columns ‘COL1‘, ‘COL2‘, ... are used in that order when present, otherwise all character columns.
`qc_cols`	Optional character vector of QC columns to use, in display order.

Value

An object of class ‘rtf_compare‘: a list with ‘rows‘ (one entry per aligned row: ‘status‘, RTF cells, QC cells, per-cell classes), ‘count‘ (number of rows per worst status) and some metadata. Use ‘as.data.frame()‘ for the cell-level differences.

Examples

```
f <- system.file("extdata", "demo.rtf", package = "rtfqc")
qc <- data.frame(COL1 = c("Age (years)", "  n", "  Mean (SD)"),
                 COL2 = c("", "30", "45.2 (10.11)"),
                 COL3 = c("", "31", "47.0 (9.80)"))
res <- rtf_compare(f, qc)
res
as.data.frame(res)
```

rtf_compare_report *Write a comparison report*

Description

Write a comparison report

Usage

```
rtf_compare_report(x, file)
```

Arguments

x	An ‘rtf_compare‘ from [rtf_compare()] or the data frame from [rtf_qc_dir()].
file	Output path ending in ‘.xlsx‘ (needs openxlsx2) or ‘.csv‘.

Value

‘file‘, invisibly.

Examples

```
f <- system.file("extdata", "demo.rtf", package = "rtfqc")
qc <- data.frame(COL1 = c("Age (years)", "  n", "  Mean (SD)"),
                 COL2 = c("", "29", "45.2 (10.1)"),
                 COL3 = c("", "31", "47.0 (9.80)"))
out <- tempfile(fileext = ".csv")
rtf_compare_report(rtf_compare(f, qc), out)
read.csv(out)
unlink(out)
```

rtf_qc_dir *Compare every RTF in a folder with its QC dataset*

Description

Pairs ‘<main_dir>/<name>.rtf‘ with ‘<qc_dir>/<prefix><name>.sas7bdat‘ (by default the common ‘v_‘ naming, e.g. ‘t_14_1_1.rtf‘ with ‘v_t_14_1_1.sas7bdat‘), compares each pair and runs the page checks.

Usage

```
rtf_qc_dir(
  main_dir,
  qc_dir,
  pattern = "\\.rtf$",
  prefix = "v_",
  ext = "sas7bdat",
  toc = NULL
)
```

Arguments

main_dir	Folder with the RTF outputs.
qc_dir	Folder with the QC datasets.
pattern	Regular expression for the RTF files to include.
prefix	Prefix of the QC dataset names.
ext	QC dataset extension (‘"sas7bdat"‘, ‘"xpt"‘, ‘"rds"‘ or ‘"csv"‘).
toc	Optional TOC (see [rtf_check_titles()]) for title / footnote checks.

Value

A data frame with one row per RTF: counts of each difference type, number of page / title issues, and an ‘error‘ column. The cell-level differences of all outputs are in ‘attr(x, "details")‘.

Examples

```
if (requireNamespace("haven", quietly = TRUE)) {
  d <- system.file("extdata", package = "rtfqc")
  # test.rtf is paired with v_test.sas7bdat in the same folder
  r <- rtf_qc_dir(d, d, pattern = "^test\\.rtf$")
  r
  head(attr(r, "details"))
}
```

---

rtf_read *Read an RTF output*

Description

Parses a clinical RTF table, listing or figure file into pages and table rows. The file is read directly; Word, LibreOffice and Pandoc are not needed.

Usage

```
rtf_read(file)
```

Arguments

file	Path to an ‘.rtf‘ file.

Value

An object of class ‘rtf_doc‘: a list with * ‘file‘: the path; * ‘pages‘: data frame with one row per page: ‘page‘, ‘head‘ (header and footer text, which is where SAS ODS RTF and r2rtf put titles and "Page x of y"), ‘body‘ (all body text); * ‘rows‘: data frame of body table rows: ‘page‘, ‘row‘, ‘header‘ (‘TRUE‘ for repeated column-header rows, RTF ‘\trhdr‘), ‘cells‘ (list column) and ‘indent‘ (list column: paragraph indent of each cell in twips, RTF ‘\li‘ + ‘\fi‘).

Examples

```
f <- system.file("extdata", "demo.rtf", package = "rtfqc")
doc <- rtf_read(f)
doc$pages
```

---

rtf_table *Body table of an RTF output as a data frame*

Description

Rebuilds the body table across pages: repeated column-header rows are dropped (or used as column names) and empty rows are removed.

Usage

```
rtf_table(x, header = c("none", "auto"), keep_indent = FALSE)
```

Arguments

x	An ‘rtf_doc‘ from [rtf_read()] or a path to an RTF file.
header	‘"auto"‘ uses the first header row (‘\trhdr‘) as column names, ‘"none"‘ names the columns ‘COL1‘, ‘COL2‘, ... (the usual QC dataset layout).
keep_indent	If ‘TRUE‘, adds an ‘indent‘ column with the number of leading spaces of the first column (useful for SOC / PT hierarchies).

Value

A data frame of character columns, plus ‘.page‘.

Examples

```
f <- system.file("extdata", "demo.rtf", package = "rtfqc")
rtf_table(f)
rtf_table(f, header = "auto", keep_indent = TRUE)
```

Index

rtf_cell_diff, 2
rtf_check, 3
rtf_check_pages, 3
rtf_check_titles, 4
rtf_compare, 5
rtf_compare_report, 6
rtf_qc_dir, 6
rtf_read, 7
rtf_table, 8