Package: highlightr (via r-universe)

October 18, 2024

Title Highlight Conserved Edits Across Versions of a Document

Version 1.0.2

Description Input multiple versions of a source document, and receive HTML code for a highlighted version of the source document indicating the frequency of occurrence of phrases in the different versions. This method is described in Chapter 3 of Rogers (2024) <https://digitalcommons.unl.edu/dissertations/AAI31240449/>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.1

Imports dplyr, fuzzyjoin, ggplot2, magrittr, purrr, quanteda, quanteda.textstats, stringi, stringr, tibble, tidyr, tm

Depends R (>= 2.10)

LazyData true

URL https://rachelesrogers.github.io/highlightr/, https://github.com/rachelesrogers/highlightr

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

BugReports https://github.com/rachelesrogers/highlightr/issues

NeedsCompilation no

Author Center for Statistics and Applications in Forensic Evidence [aut, cph, fnd], Rachel Rogers [aut, cre] (<https://orcid.org/0000-0002-4145-9630>), Susan VanderPlas [aut] (<https://orcid.org/0000-0002-3803-0972>)

Maintainer Rachel Rogers <rrogers.rpackages@gmail.com>

Repository CRAN

Date/Publication 2024-10-17 18:10:02 UTC

Contents

`collocate_comments` *Collocation of Comments*

Description

This function provides the frequency of collocations in comments that correspond to the provided transcript.

Usage

```
collocate_comments(transcript_token, note_token, collocate_length = 5)
```

Arguments

`transcript_token`	transcript token to act as baseline for notes, resulting from `token_transcript()`
`note_token`	tokenized document of notes, resulting from `token_comments()`
`collocate_length`	the length of the collocation. Default is 5

Value

data frame of the transcript and corresponding note frequency

Examples

```
comment_example_rename <- dplyr::rename(comment_example, page_notes=Notes)
toks_comment <- token_comments(comment_example_rename[1:100,])
transcript_example_rename <- dplyr::rename(transcript_example, text=Text)
toks_transcript <- token_transcript(transcript_example_rename)
collocation_object <- collocate_comments(toks_transcript, toks_comment)
```

`collocate_comments_fuzzy` *Collocate Comments Fuzzy*

Description

This function provides the frequency of collocations in comments that correspond to the provided transcript, using fuzzy matching.

Usage

```
collocate_comments_fuzzy(transcript_token, note_token, collocate_length = 5)
```

Arguments

`transcript_token`	transcript token to act as baseline for notes, resulting from `token_transcript()`
`note_token`	tokenized document of notes, resulting from `token_comments()`
`collocate_length`	the length of the collocation. Default is 5

Value

data frame of the transcript and corresponding note frequency

Examples

```
comment_example_rename <- dplyr::rename(comment_example, page_notes=Notes)
toks_comment <- token_comments(comment_example_rename)
transcript_example_rename <- dplyr::rename(transcript_example, text=Text)
toks_transcript <- token_transcript(transcript_example_rename)
collocation_object <- collocate_comments_fuzzy(toks_transcript, toks_comment)
```

`collocation_plot` *Map collocation to ggplot object*

Description

This assigns colors based on frequency to the words in the transcript.

Usage

```
collocation_plot(
  frequency_doc,
  n_scenario = 1,
  colors = c("#f251fc", "#f8ff1b")
)
```

Arguments

frequency_doc	document of frequencies (returned from transcript_frequency())
n_scenario	number of scenarios for which this transcript appeared. Defualt is 1
colors	list for color specification for the gradient. Default is c("#f251fc","#f8ff1b")

Value

list of plot, plot object, and frequency

Examples

```
comment_example_rename <- dplyr::rename(comment_example, page_notes=Notes)
toks_comment <- token_comments(comment_example_rename)
transcript_example_rename <- dplyr::rename(transcript_example, text=Text)
toks_transcript <- token_transcript(transcript_example_rename)
collocation_object <- collocate_comments_fuzzy(toks_transcript, toks_comment)
merged_frequency <- transcript_frequency(transcript_example_rename, collocation_object)
freq_plot <- collocation_plot(merged_frequency)
```

---

comment_example *Comment Example Dataset*

Description

Participant comments for the initial description used in the jury perception study

Usage

```
comment_example
```

Format

comment_example**:**

A data frame with 125 rows and 2 columns:

ID Participant Identifier

Notes Participant notes

Source

Jury Perception Study (see Rogers (2024) https://digitalcommons.unl.edu/dissertations/AAI31240449/)

highlighted_text *Create Highlighted Testimony*

Description

Adds html tags to create a highlighted testimony corresponding to word frequency.

Usage

```
highlighted_text(plot_object, labels = c("", ""))
```

Arguments

plot_object	plot object resulting from collocation_plot()
labels	lower and upper labels for the gradient scale

Value

html code for highlighted text

Examples

```
comment_example_rename <- dplyr::rename(comment_example, page_notes=Notes)
toks_comment <- token_comments(comment_example_rename)
transcript_example_rename <- dplyr::rename(transcript_example, text=Text)
toks_transcript <- token_transcript(transcript_example_rename)
collocation_object <- collocate_comments_fuzzy(toks_transcript, toks_comment)
merged_frequency <- transcript_frequency(transcript_example_rename, collocation_object)
freq_plot <- collocation_plot(merged_frequency)
page_highlight <- highlighted_text(freq_plot, merged_frequency)
```

token_comments *Tokenize comments*

Description

This function tokenizes comments that are to be used in collocate_comments_fuzzy() or collocate_comments()

Usage

```
token_comments(comment_document)
```

Arguments

comment_document	document containing notes by individual, where the column containing the notes is named page_notes

Value

tokenized comments

Examples

```
comment_example_rename <- dplyr::rename(comment_example, page_notes=Notes)
toks_comment <- token_comments(comment_example_rename)
```

---

token_transcript *Tokenize Transcript*

Description

This function tokenizes a transcript document that is to be used in `collocate_comments_fuzzy()` or `collocate_comments()`

Usage

```
token_transcript(transcript_file)
```

Arguments

transcript_file	data frame of the transcript, where the transcript text is in a column named text.

Value

a tokenized object

Examples

```
transcript_example_rename <- dplyr::rename(transcript_example, text=Text)
toks_transcript <- token_transcript(transcript_example_rename)
```

---

transcript_example *Transcript Example*

Description

Text corresponding to participant comments

Usage

```
transcript_example
```

Format

`transcript_example`**:**

A data frame with 1 row and 1 column:

Text Transcript text corresponding to the jury perception study

Source

Jury Perception Study (see Rogers (2024) https://digitalcommons.unl.edu/dissertations/AAI31240449/ and Garrett et. al. (2020) doi:10.1037/lhb0000423)

`transcript_frequency` *Mapping Collocation Frequency to Transcript Document*

Description

This function connects the collocation frequency calculated in `collocate_comments_fuzzy()` to the base transcript.

Usage

```
transcript_frequency(transcript, collocate_object)
```

Arguments

`transcript`	transcript document
`collocate_object`	collocation object (returned from `collocate_comments_fuzzy()` or `collocate_comments()`)

Value

a dataframe of the transcript document with collocation values by word

Examples

```
comment_example_rename <- dplyr::rename(comment_example, page_notes=Notes)
toks_comment <- token_comments(comment_example_rename)
transcript_example_rename <- dplyr::rename(transcript_example, text=Text)
toks_transcript <- token_transcript(transcript_example_rename)
collocation_object <- collocate_comments_fuzzy(toks_transcript, toks_comment)
merged_frequency <- transcript_frequency(transcript_example_rename, collocation_object)
```

`wiki_pages` *Wikipedia Edit History for "Highlighter"*

Description

Text corresponding to versions of the Wikipedia article for Highlighter

Usage

```
wiki_pages
```

Format

`wiki_pages`**:**
A data frame with 50 rows and 1 column:

page_notes text of the Wikipedia page for Highlighter

Source

Wikipedia: https://en.wikipedia.org/w/index.php?title=Highlighter&action=history

Index