Package: disclosuR (via r-universe)

October 6, 2024

Type Package

Title Text Conversion from Nexis Uni PDFs to R Data Frames

Version 0.6.0

Date 2024-01-07

Description Transform 'newswire' and earnings call transcripts as PDF obtained from 'Nexis Uni' to R data frames. Various 'newswires' and 'FairDisclosure' earnings call formats are supported. Further, users can apply several pre-defined dictionaries on the data based on Graffin et al. (2016)<doi:10.5465/amj.2013.0288> and Gamache et al. (2015)<doi:10.5465/amj.2013.0377>.

License GPL-3

Imports dplyr, lubridate, pdftools, qdap, SentimentAnalysis, stringi, stringr, syuzhet, zoo, SnowballC, tm, stats, rlang

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Author Jonas Röttger [aut, cre]

Maintainer Jonas Röttger <jonas.roettger@gmx.net>

Repository CRAN

Date/Publication 2024-01-09 12:40:10 UTC

Contents

conference_call_segmenter *Earnings call segmenter*

Description

Converts one earnings call transcript from 'FairDisclosure' obtained from 'NexisUni' to an R data frame.

Usage

```
conference_call_segmenter(
  file,
  sentiment = FALSE,
  emotion = FALSE,
  regulatory_focus = FALSE,
  laughter = FALSE,
  narcissism = FALSE
)
```

Arguments

file	The name of the PDF file which the data are to be read from. If it does not contain an absolute path, the file name is relative to the current working directory, getwd().
sentiment	Performs dictionary-based sentiment analysis based on the analyzeSentiment function (default: FALSE)
emotion	Performs dictionary-based emotion analysis based on the get_nrc_sentiment function (default: FALSE)
regulatory_focus	Calculates the number of words indicative for promotion and prevention focus based on the dictionary developed by Gamache et al., 2015 (default: FALSE)
laughter	Counts the number of times laughter was indicated in a quote. (default: FALSE)
narcissism	Counts the number of pronoun usage and calculates the ratio of first-person singular to first-person plural pronouns. This measure is derived from Zhu & Chen, (2015 (default: FALSE)

Value

An R data frame with each row representing one quote. The columns indicate the quarter, year, section (presentation versus Q&A), the speaker's name, role, affiliation, and also three binary indicators on whether the speaker is the host company's (1) CEO, (2) CFO, and/or (3) Chairman.

Examples

```
earnings_calls_df <- conference_call_segmenter(file = system.file("inst",
"examples",
"earnings_calls", "earnings_example_01.pdf",
package = "disclosuR"));
earnings_calls_df_sentiment <- conference_call_segmenter(file = system.file("inst",
"examples",
"newswire", "earnings_example_01.pdf",
package = "disclosuR"),
sentiment = TRUE);
```

`conference_call_segmenter_folder`

Earnings call segmenter (multiple files)

Description

Converts all 'FairDisclosure' earnings call transcripts obtained from 'NexisUni' in a folder to an R data frame.

Usage

```
conference_call_segmenter_folder(
  folder_path,
  sentiment = FALSE,
  emotion = FALSE,
  regulatory_focus = FALSE,
  laughter = FALSE,
  narcissism = FALSE
)
```

Arguments

`folder_path`	The name of the folder which the data are to be read from. If it does not contain an absolute path, the file name is relative to the current working directory.
`sentiment`	Performs dictionary-based sentiment analysis based on the `analyzeSentiment` function (default: FALSE)
`emotion`	Performs dictionary-based emotion analysis based on the
`regulatory_focus`	Calculates the number of words indicative for promotion and prevention focus based on the dictionary developed by Gamache et al., 2015 (default: FALSE)
`laughter`	Counts the number of times laughter was indicated in a quote. (default: FALSE)
`narcissism`	Counts the number of pronoun usage and calculates the ratio of first-person singular to first-person plural pronouns. This measure is derived from Zhu & Chen, (2015 (default: FALSE)

Value

An R data frame with each row representing one quote. The columns indicate the quarter, year, section (presentation versus Q&A), the speaker's name, role, affiliation, and also three binary indicators on whether the speaker is the host company's (1) CEO, (2) CFO, and/or (3) Chairman.

Examples

```
earnings_calls_df <- conference_call_segmenter_folder(
folder_path = system.file("inst",
"examples",
"earnings_calls",
package = "disclosuR"));
earnings_calls_df_sentiment <- conference_call_segmenter_folder(
folder_path = system.file("inst",
"examples",
"newswire",
sentiment = TRUE,
package = "disclosuR"));
```

---

impression_offsetting *Impression offsetting*

Description

Takes an event data set containing of dates and CUSIPs which have to correspond to a press data frame compiled by the function newswire_segmenter_folder.

Usage

```
impression_offsetting(event_data, press_data_categorized)
```

Arguments

event_data	An R data that contains three columns which have to be labeled "date_announced", "cusip", and "ID". The date_announced column contains the dates of the events for which impression offsetting is calculated. The cusip column contains the 8-digit cusip of the companies for which impression offsetting is calculated. The ID column should contain a unique ID that identifies the specific event.
press_data_categorized	An R data frame with each row representing one 'newswire' article. The columns indicate the title, text, 'newswire', date, and weekday. It should be the outcome of newswire_segmenter in which both the argument sentiment and text_clustering have been set to TRUE.

Value

An R data frame which contains the column of the event_data plus three columns for the baseline announcements (positive, neutral, and negative) and three columns for the impression offsetting announcements (positive, neutral, and negative).

Examples

```
## Not run:
impression_offsetting(event_data, press_data)

## End(Not run)
```

`newswire_segmenter` *Newswire segmenter*

Description

Takes a PDF document containing a 'newswire' document obtained from 'NexisUni' and transforms it into an R data frame consisting of one row

Usage

```
newswire_segmenter(
  file,
  sentiment = FALSE,
  emotion = FALSE,
  regulatory_focus = FALSE,
  laughter = FALSE,
  narcissism = FALSE,
  text_clustering = FALSE
)
```

Arguments

`file`	The name of the PDF file which the data are to be read from. If it does not contain an absolute path, the file name is relative to the current working directory, getwd().
`sentiment`	Performs dictionary-based sentiment analysis based on the `analyzeSentiment` function (default: FALSE)
`emotion`	Performs dictionary-based emotion analysis based on the `get_nrc_sentiment` function (default: FALSE)
`regulatory_focus`	Calculates the number of words indicative for promotion and prevention focus based on the dictionary developed by Gamache et al., 2015 (default: FALSE)
`laughter`	Counts the number of times laughter was indicated in a quote. (default: FALSE)

narcissism Counts the number of pronoun usage and calculates the ratio of first-person singular to first-person plural pronouns. This measure is derived from Zhu & Chen, (2015 (default: FALSE)

text_clustering Applies a document categorization using a dictionary developed based on the framework developed by Graffin et al., 2016. (default: FALSE)

Value

An R data frame with each row representing one 'newswire' article. The columns indicate the title, text, 'newswire', date, and weekday.

Examples

```
newswire_df <- newswire_segmenter(
file = system.file("inst",
"examples",
"newswire", "newswire_example_01.pdf",
package = "disclosuR"));
newswire_df_sentiment <- newswire_segmenter(
file = system.file("inst",
"examples",
"newswire", "newswire_example_01.pdf",
sentiment = TRUE,
package = "disclosuR"));
```

---

newswire_segmenter_folder

Newswire segmenter (multiple files)

Description

Takes all PDF documents in a folder containing 'newswire' documents obtained from 'NexisUni' and transforms them into an R data frame consisting of one row per document.

Usage

```
newswire_segmenter_folder(
  folder_path,
  sentiment = FALSE,
  emotion = FALSE,
  regulatory_focus = FALSE,
  laughter = FALSE,
  narcissism = FALSE,
  text_clustering = FALSE
)
```

Arguments

folder_path	The path to the folder in which the 'newswire' PDFs reside. If it does not contain an absolute path, the folder name is relative to the current working directory, getwd().
sentiment	Performs dictionary-based sentiment analysis based on the analyzeSentiment function (default: FALSE)
emotion	Performs dictionary-based emotion analysis based on the get_nrc_sentiment function (default: FALSE)
regulatory_focus	Calculates the number of words indicative for promotion and prevention focus based on the dictionary developed by Gamache et al., 2015 (default: FALSE)
laughter	Counts the number of times laughter was indicated in a quote. (default: FALSE)
narcissism	Counts the number of pronoun usage and calculates the ratio of first-person singular to first-person plural pronouns. This measure is derived from Zhu & Chen, (2015 (default: FALSE)
text_clustering	Applies a document categorization using a dictionary developed based on the framework developed by Graffin et al., 2016. (default: FALSE)

Value

An R data frame with each row representing one 'newswire' article. The columns indicate the title, text, 'newswire', date, and weekday. (default: FALSE)

An R data frame with each row representing one 'newswire' article. The columns indicate the title, text, 'newswire', date, and weekday. Depending on the additional arguments, the output data can also contain sentiment, emotion, regulatory focus, laughter, narcissism and text cluster based on the Graffin et al. categories.

Examples

```
newswire_df <- newswire_segmenter_folder(
folder_path = system.file("inst",
"examples",
"newswire",
package = "disclosuR"));
newswire_df_sentiment <- newswire_segmenter_folder(
folder_path = system.file("inst",
"examples",
"newswire",
package = "dislosuR"), sentiment = TRUE);
```

Index