Package ‘readOffice’

July 23, 2025

Type Package

Title Read Text Out of Modern Office Files

Version 0.2.2

Author Mark Ewing

Maintainer Mark Ewing <b.mark@ewingsonline.com>

Description Reads in text from 'unstructured' modern Microsoft Office files (XML based files) such as Word and PowerPoint.
This does not read in structured data (from Excel or Access) as there are many other great packages to that do so already.

License Unlimited

Encoding UTF-8

LazyData true

Imports xml2 (>= 1.0.0), rvest (>= 0.3.2), purrr (>= 0.2.2), magrittr (>= 1.5)

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-03-08 08:22:32

Contents

read_docx *Read data from a Modern Word File*

Description

Read data from a Modern Word File

Usage

```
read_docx(docx)
```

Arguments

docx	The .docx file to read

Details

Only accepts one file at a time and only .docx files. Modifying file extensions will not work.

Text is broken out into the XML defined paragraphs in the vector.

Value

Vector of document text

Examples

```
read_docx(docx = system.file('extdata','example.docx',package='readOffice'))
```

read_pptx *Read data from a Modern PowerPoint File*

Description

Read data from a Modern PowerPoint File

Usage

```
read_pptx(pptx)
```

Arguments

pptx	The .pptx file to read

Details

Only accepts one file at a time and only .pptx files. Modifying file extensions will not work.

The returned list contains named lists of the elements on the slide, each element of which is either a data.frame or a matrix containing the text and minor details about the structure on the page.

Data frames will contain the text in addition to the following columns: "Bulleted" indicates if the text is part of a bulleted or numbered list on the slide. "Hierarchy" indicates the tabbed depth of the element in a bulleted or numbered list (NA if not bulleted).

Alternatively, returns a matrix for tables on the slide.

Value

List containing slide elements.

Examples

```
read_pptx(system.file('extdata','example.pptx',package='readOffice'))
```

Index