

Package ‘antiword’

January 30, 2024

Type Package

Title Extract Text from Microsoft Word Documents

Version 1.3.3

Description Wraps the 'AntiWord' utility to extract text from Microsoft Word documents. The utility only supports the old 'doc' format, not the new xml based 'docx' format. Use the 'xml2' package to read the latter.

Imports sys (>= 2.0)

URL <https://docs.ropensci.org/antiword/>

BugReports <https://github.com/ropensci/antiword/issues>

License GPL-2

Encoding UTF-8

NeedsCompilation yes

Author Jeroen Ooms [aut, cre] (<<https://orcid.org/0000-0002-4035-0289>>),
Adri van Os [cph] (Author 'antiword' utility)

Maintainer Jeroen Ooms <jeroen@berkeley.edu>

Repository CRAN

Date/Publication 2024-01-30 12:40:05 UTC

R topics documented:

antiword	2
Index	3

antiword

Antiword

Description

Wraps the antiword utility. Takes a path to an word file and returns text from the document.

Usage

```
antiword(file = NULL, format = FALSE)
```

Arguments

file	path or url to your word file
format	format the output text (-f parameter)

Examples

```
text <- antiword("https://jeroen.github.io/files/UDHR-english.doc")  
cat(text)
```

Index

antiword, [2](#)