Atitit word office文档转换txt 纯文本 docx转txt 解决方案cmd doctotext v3 u66.docx

Atitit word office文档转换txt 纯文本 doctotext

D:\doctotext\doctotext.exe "C:\Users\Administrator\OneDrive\桌面\prjs hsu6\2020docx\sumdoc hsu6\Atitit 稳定性提升总结kok案例.docx"

for %I in ("C:\Users\Administrator\OneDrive\桌面\prjs hsu6\2020docx\sumdoc hsu6\\*.\*") do D:\doctotext\doctotext.exe "%I" >"%I.txt"

## 命令行 doctotext.exe

D:\doctotext>doctotext.exe

DocToText v4.0.1512

Converts DOC, XLS, PPT, RTF, ODF (ODT, ODS, ODP), OOXML (DOCX, XLSX, PPTX) and HTML documents to plain text

Copyright (c) 2006-20112 SILVERCODERS(R)

http://silvercoders.com

Usage: doctotext [OPTION]... [FILE]

FILE name of file to convert

Options:

--meta extract metadata instead of text.

--rtf try to parse as RTF document first.

--odf try to parse as ODF/OOXML document first.

--ooxml try to parse as ODF/OOXML document first.

--xls try to parse as XLS document first.

--xlsb try to parse as XLSB document first.

--iwork try to parse as iWork document first.

--ppt try to parse as PPT document first.

--doc try to parse as DOC document first.

--html try to parse as HTML document first.

--pdf try to parse as PDF document first.

--eml try to parse as EML document first.

--odfxml try to parse as ODFXML (Open Document Flat XML) document first.

--fix-xml try to fix corrupted xml files (ODF, OOXML)

--strip-xml strip xml tags instead of parsing them (ODF, OOXML)

--unzip-cmd=[COMMAND] use COMMAND to unzip files from archives (ODF, OOXML)

instead of build-in decompression library.

%d in the command is substituted with destination directory path

%a in the command is substituted with name of archive file

%f in the command is substituted with name of file to extract

Example: --unzip-cmd="unzip -d %d %a %f"

--verbose turn on verbose logging

--log-file=[PATH] write logs to specified file.

D:\doctotext>D:\doctotext\doctotext.exe "D:\ati\sum doc txted\p2015sumdoc txted\Atitit 《摩奴法典》不是由国王 颁布的，而是 僧侣编制.doc" >ra2txt.txt

Using DOC parser.

D:\doctotext\doctotext.exe c:\\Atitit OLTP之道 attilax著.docx

效果良好。。目录内容都齐全

## Openoffice模式 应该也可 但是麻烦些。体积大

需要启动一个后台服务。。使用程序语言连接。不能直接调用cli？？

java 调用OpenOffice将word格式文件转换为pdf格式 - warrior1234 - 博客园.mhtml