AutoHotkey      Ask for Help                                                    View New Content

## can AHK read pdf?

Started by azure , Apr 05 2012 10:01 PM

hello!

can AHK read the text only of a pdf?

thanks!

Read the text only, of the pdf file?

out of the box, just like it does with txt files?

You know the answer to that one azure. No. You only ask because you can't figure out how to Run pfd2text ;-)

---

I had trouble figuring that out as well. Azure, download and install ghostscript 9.05, paste this code into a AHK file, put that ahk file in the same directory as your pdf files, open a command prompt window, then hit F1. It's not sophisticated, but it works.

```
F1::
Loop,*.pdf
{
    msg="C:\Program Files\gs\gs9.05\bin\gswin32c.exe" -q -dNODISPLAY -dSAFER -dDELAYBIND -dWRITESYSTEMDICT -dSIMPLE -c save -f ps2ascii.ps %A_Loopfilefullpath% -c quit >%A_l
    send %msg%
    sleep, 1000
    send {enter}
    sleep, 500
    loop
    {
        ifexist,%A_loopfilename%.txt
        {
            ;msgbox, exists
            break
        }
        else
        {
            ;msgbox, doesn't exist
            sleep, 2000
            if (A_index>10)
            {
                break
            }
        }
    }
}
```

Lets say I have a list of titles and several PDFs. I need to automatically find which title of the list is part of which PDF.

I managed to do it by converting each PDF in RTF (just by saving it as), clear the formatting and saving it as DOC. Then AHK is actually able to read the DOC like it would be a txt so I can use filereadline with a loop.

The problem now is that I have about 5000 titles and about 90 different PDFs and most probably I will have to repeat this repetitively in the future with different files so I was wandering if anyone knows how to simply read directly PDFs the same way AHK woudl read a txt so to not have to convert all the PDFs in RTF and then DOC all the times.

When I try to run the above, it just send "C:\Program Files\gs\gs9.05\bin\gswin32c.exe" -q -dNODISPLAY -dSAFER -dDELAYBIND -dWRITESYSTEMDICT -dSIMPLE -c save -f ps2ascii.ps %A_Loopfilefullpath% -c quit >%A_loopfilename%.txt" ![sad.png]

Try xpdf (http://www.autohotkey.com/board/topic/34005-xpdf-cmd-good-for-pdf-to-text-conversions/)

Impressive! It converted a 900 pages PDF in couple of seconds!!

Awesome! Thanks a lot! This really speeds this torture up! =)

[https://ahknet.autoh...workingwithpdfs (https://ahknet.autohotkey.com/~faqbot/faq.html#workingwithpdfs)](https://ahknet.autohotkey.com/~faqbot/faq.html#workingwithpdfs)

- Posted 03 February 2013 - 10:23 AM

Back to Ask for Help

AutoHotkey Community  →  AutoHotkey  →  Ask for Help