Subject: Read a PDF
Posted by Chrisparr on Sat, 30 Jul 2022 17:25:19 GMT
View Forum Message <> Reply to Message

Subject: Re: Read a PDF
Posted by jjacksonRIAB on Thu, 11 Aug 2022 08:15:59 GMT
View Forum Message <> Reply to Message

I don't think so. As far as I know U++ cannot read pdf files, it can only output them from other formats. You'll want to find an external library for that.

Subject: Re: Read a PDF
Posted by Chrisparr on Thu, 11 Aug 2022 13:48:31 GMT
View Forum Message <> Reply to Message

Hi,
I've managed to extract text from a pdf. To do that I needed the Zlib library which is part of Ultimate++. What was needed was the Zdecompress function. It took a while but it's working fine now.

Subject: Re: Read a PDF
Posted by jjacksonRIAB on Thu, 11 Aug 2022 15:02:24 GMT
View Forum Message <> Reply to Message

Cool.

How do you handle when the text you're extracting from the file doesn't match the order it appears on the rendered page?

Subject: Re: Read a PDF
Posted by Chrisparr on Thu, 11 Aug 2022 15:16:54 GMT
View Forum Message <> Reply to Message

I had a very specific problem.
The pdf's were of a question and answer type.
A web page posed questions and the user provided answers.
When the questions were done a pdf was generated.
When decoded I pick up the question answer pairs.
I am ignoring titles, and various other stuff which doesn't concern me.
So far all the pdf's of this type have decoded well.

I am making no claim to have created a generalised pdf translation tool

---

Subject: Re: Read a PDF
Posted by Chrisparr on Fri, 12 Aug 2022 08:16:19 GMT
View Forum Message <> Reply to Message

Maybe you have something much better, but I've extracted the code which does the job for me. No guarantees but as I said it does the job for me. That is all I can say

File Attachments
1) Extract.cpp, downloaded 239 times

Subject: Re: Read a PDF
Posted by jjacksonRIAB on Fri, 12 Aug 2022 08:39:04 GMT
View Forum Message <> Reply to Message

If it works for you every time you need it to and it does just what you need without importing a bunch of stuff you don't, then it's the best solution. I only raise the issue of it being used as a general solution because the topic question was pretty broad and I don't want to risk trivializing it for others reading, that's all.

Glad you got it working and thanks for posting your code. :d