General

Why isn't my PDF file being imported?

PDF has been referred to as a "write-only" file format[1], since it's so difficult for applications to extract useful information from it. A PDF file is only required to contain what's necessary to reproduce the visual representation of a document, but it notably isn't required that there's any orderly record of what's in the document.

PDF documents also are by design not particularly editable because nothing is ever deleted from them: even tools that are able to modify/edit documents can only do so by adding new information. The old information — text, images, structure, etc. — is still retained within the file. And it can get to be a bit of a mess in there.

Fade In is, however, able to import a wide range of PDF screenplays. Sometimes, however, it may run into problems importing, and there are at least a couple of reasons why that might be.

The first reason is related to screenplays that were scanned from a printed copy. Scanned PDFs are generally just comprised of images of the text rather than the text itself. (Unless during the scanning process the additional step of OCR — Optical Character Recognition — was taken to add an separate, invisible layer of actual text reflecting the contents of the images of text.)

Another reason that a PDF file may not be imported correctly is this:

```
59 0 obj
<</BaseFont/Helvetica/Encoding<</BaseEncoding/WinAnsiEncoding/Type/Encoding>>/Name/OPBaseFont3/Subtype/Type1/Type/Fo
endobj
26 0 obj
<</CreationDate(D:20160714003812-08'00')/Creator(PaperPort 14)/ModDate<443A32303136303731343030333831382D3038273030
endobj
60 0 obj
<</Filter/FlateDecode/Length 1009>>
stream
xúçV]o,8¿}ÁW<}¢#Å'é¿«y¿-Yf¿Ì¿´í™™4/i0$-â´fYf¸˙ΩŒ¿¿Ç/™ïJU˙ú{œπÁ∫∑·¿¿˙Ul€¿w¿
(âÌ'
íìÄÉO]¬ Ã¡'¿Ñ{¿¸7¿_ã¡◊Â?∑Q©~Ë<‡aF•/õ¡Õ„Ú¿œO¿•}yOCX.`>Õ/·Û    ®G<¿Ñ¿Ñz¿5¿D0¿zÇGÅ:-fil¿¿¯sÚ∞Z.Ü+¿èa¿B¯s:ªÉpr˙kJ`¿?¶¿0_Œ
¿∫B)¸∏„ÿµ¿r¿'Ï¿÷¿∂;ŒO"ÏµÛÅµ]ê·G=j,C‡>ª,]@¿vœ∏¿˘t¿ÏáYÄX◊A/¿:ÉΩ.^œιkßX¿=f•#±Q#«iv¿jfiT~_%#»±"'¶K,◊T6IΩj-¿ó¿¨Au/}˙ûL8FL◊
¨Å–zs¿§
wà„¿s¿".6Â2!∞–¿¿–èó\ÔGm™0I∫/&¿ßr>À¿¬·\3Gœ¡…&ñ¿¶¢t¿©©s¿A˘™,4/]–ÕE≈EÎ[;r:é£≤^
#ÿ'iúX§¥D%z°ƒf}&¡ι=āÙ°{H»êP•7¯d√∏Ë|€jΩ&'-X
jÛ˙Ph°ãèƒq}À&ò'´ç¿πlT¿,¿‡∫Ú¿∆v˝¸T∞≈%_!áMXb_Æ$5Ê¿ÒN„N+{ˆ3+ìã¿∑¯å˝î∏¶(Æ{zûƒ0âløÀøõ¿ÿi˝¿õ4WWË¿¿•>nqœ´∞OÒù†¯Z°mŸfÿ~filLpØ¸
¸◊¡ÙM‡ˆ5¢¿˙\¿Îm¿8>€{ÃÃ¿ö–¨œi8¯¿∑Ã\L
endstream
endobj
```

That's what the inside of a PDF document looks like.

There really are no requirements for structure or layout within PDF: it's essentially just a container for an arbitrary collection of "objects". These objects can be stored in any number of ways, and in any order. For instance, a PDF document may store the words "quick", "The", "fox", and "brown", in that order, along with instructions to display them onscreen as: "The" "quick" "brown" "fox". (And some PDF creators basically do that.)

The problem then comes with trying to figure out exactly what the layout of the objects is supposed to represent in logical terms. (This is one of the things that

General

human beings are way better at than computers. For now.)

There's also no requirement that PDF store the word "fox" as "fox". It could store it as "D&!" or almost anything else, as long as it also stores the — sometimes byzantine — rules of transliteration.

The greatest obstacle to reliable PDF import is the wide variety of PDF creators and methods of creation which can result in all sorts of variables to be taken into account when trying to extract the content. There may still be a few instances where things don't work perfectly.

In those cases, what you may have to/be able to do is select the text from a PDF viewer and paste it into Fade In. With that method, however, a certain amount of reformatting is almost guaranteed.

[1]By this author

Unique solution ID: #1094
Author: Fade In Knowledge Base
Last update: 2020-06-07 04:35