

PDF Structazer Short User Manual

PDF Structazer has been presented at the Black Hat Europe 2008 conference (<http://www.blackhat.com/html/bh-europe-08/bh-eu-08-archives.html#Filiol>) and is dedicated to the precise manipulation of PDF file. Due to some legal restriction, the present version of the application does not include the advanced programming features (the AV community claims that it would help developing PDF malware). This product is free of use provided that the user mentions the Black Hat paper related to it and the name of its authors.

At the present time, there is no application allowing PDF file manipulation and analysis. Products such as Adobe Acrobat, PDF Creator, PDF Converter Professional... (free or proprietary) generally allow the PDF object manipulation only : merging pages, removing objects, form creation... Since they are working at the object, they not suitable for lower analysis at the PDF code level. This is the reason why we have developed such a code level analysis tool, initially for our own purpose.

The main functions available with *PDF Structazer* are:

- PDF File structure analysis.
- Identifying every single object independently from the others.
- Identifying, computing and processing the *Cross Reference Table* and computing the offset of any PDF object within the PDF document. It also enables to locate this table (*startxref* field).
- Computing and processing very easily the length of object streams.

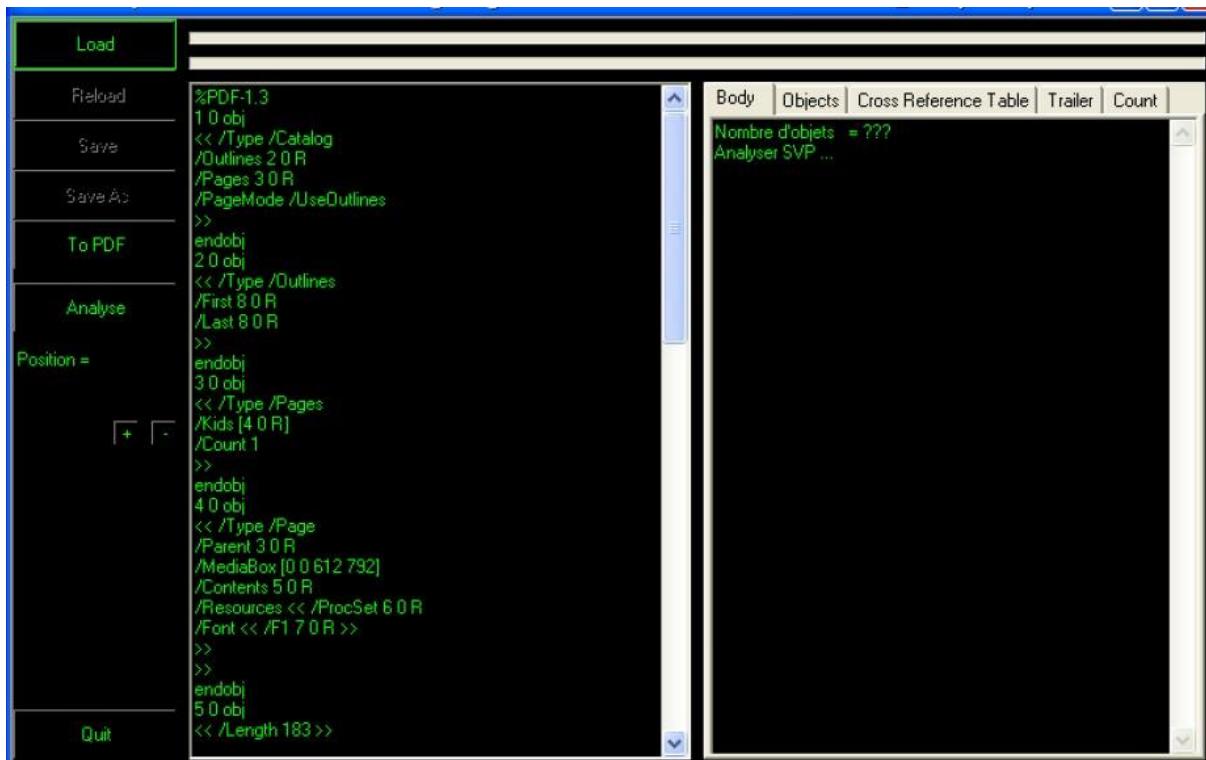
Description of PDF Structazer

PDF Structazer essentially performs four main functionalities (for the open version). It has been built with the Microsoft Visual Studio .NET development framework.

1. Load/Reload

It is possible to load a PDF source code directly from a PDF file or a TXT file (thus the user can directly program PDF language).

The *Body* windows will display useful information whenever the document analysis is launched only. In case of document modification, it is possible to refresh the window content through the Reload function (thus allowing indirect PDF programming in the quite same way at non WYSIWYG application).



2. Save/Save As

Once the file has been modified, it is possible to save the document directly from the *PDF Structazer* application, in any possible directory.

3. To PDF

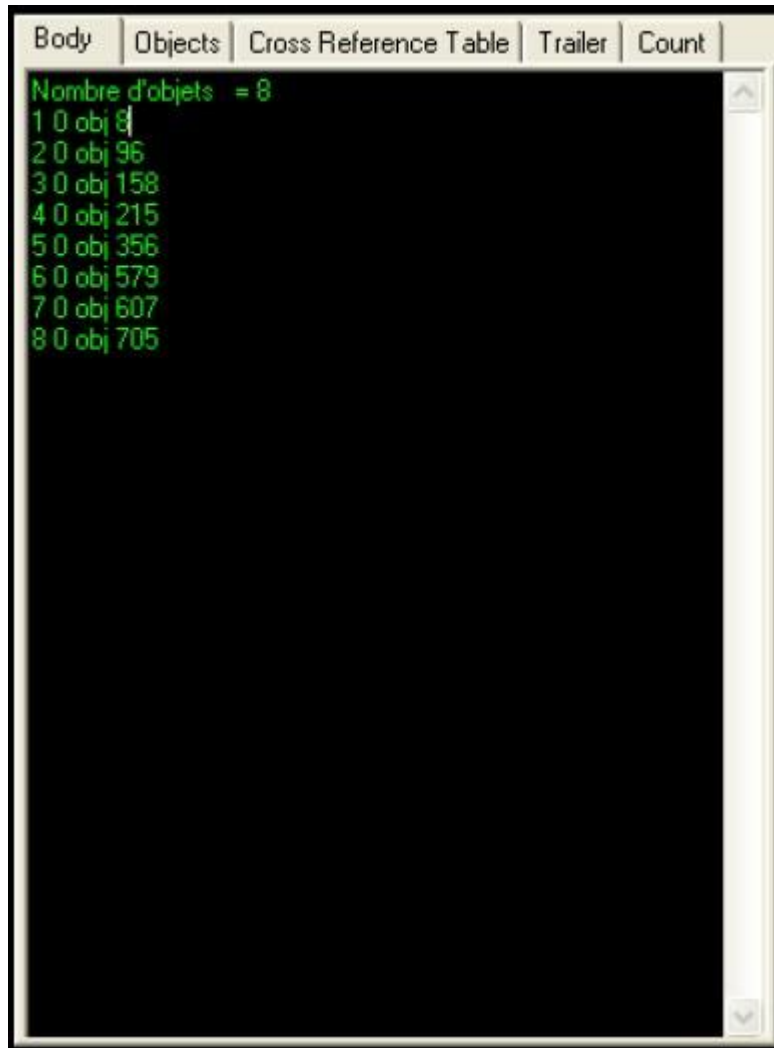
In order to make the PDF document manipulation easier and to directly test the final document rendering and/or the desired effects, this function enables to save the document directly into the PDF format.

Nota: despite the fact that the file manipulation is performed on ASCII 7-bit encoding (text format), file reading and saving functions are bitwise performed in binary format in order to manage any non ASCII character.

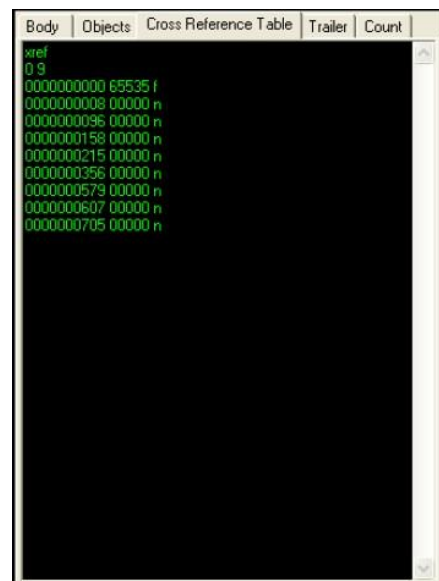
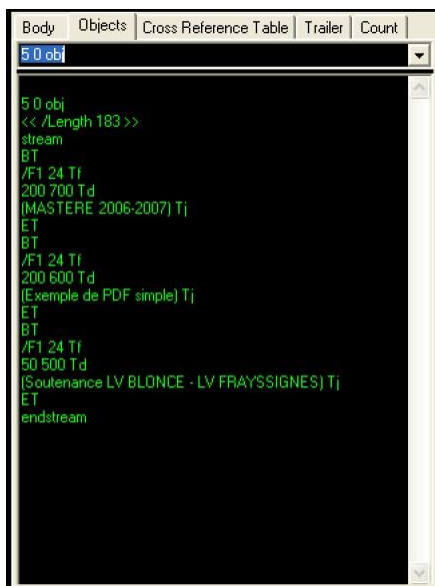
4. Analyze

It is *PDF Structazer's* main function. File analysis enables to extract the exact location of any single PDF object within the PDF file, to precisely specify its location within the file structure and to compute any reference in the *Cross Reference Table*.

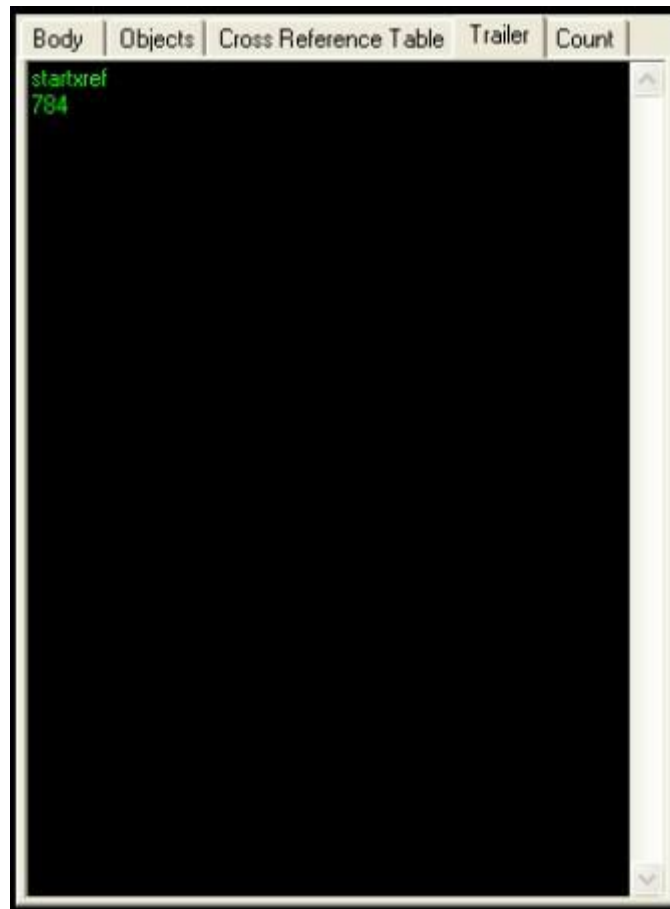
To ease the PDF code line manipulation, *PDF Structazer* enables to extract and isolate any data or PDF object in a separate window. Consequently, any further modification of the object will be far easier and thus independently from the document body or the other objects.



The application determines the content and entries of the *Cross Reference Table* from the exact location of every single PDF object within the document.



This function has been developed in order to compute the initial *Cross Reference Table* before any incremental file modification has been performed.



Additionally, *PDF Structazer* computes the offset of the Cross Reference Table (*startxref* field). This value is a critical one since it is the first value that is read and used by any PDF reader.

Finally, any PDF document contains a lot of PDF objects that in fact are complex structures of objects, called object streams. Whenever we manipulate PDF code within objects, it is very often necessary to actualize the stream length value ("*length*" field). The *Count* tool is very useful at computing this value directly, through a simple copy/paste operation into the working window.

```
Body | Objects | Cross Reference Table | Trailer | Count
stream
BT
/F1 24 Tf
200 700 Td
(MASTERE 2006-2007) Tj
ET
BT
/F1 24 Tf
200 600 Td
(Exemple de PDF simple) Tj
ET
BT
/F1 24 Tf
50 500 Td
(Soutenance LV BLONCE - LV FRAYSSIGNES) Tj
ET
endstream
```

Clear

Count = 189