Explore PDF

The "Explore PDF" functionality gives you an insight into the internal structure of PDF files.
Usually, this is not needed for common use, but might be helpful if you encounter a damaged file or if you are simply interested in learning more about the internal structure of a PDF file. The entry "Explore PDF..." can be found in the "Plug-Ins" menu of Acrobat ("Miscellaneous") or the "Tools" menu in the Standalone version.

With the "Explore PDF" tool you can have a look into the data structure of a PDF file, exposing the several commands and details that are forming the page objects.
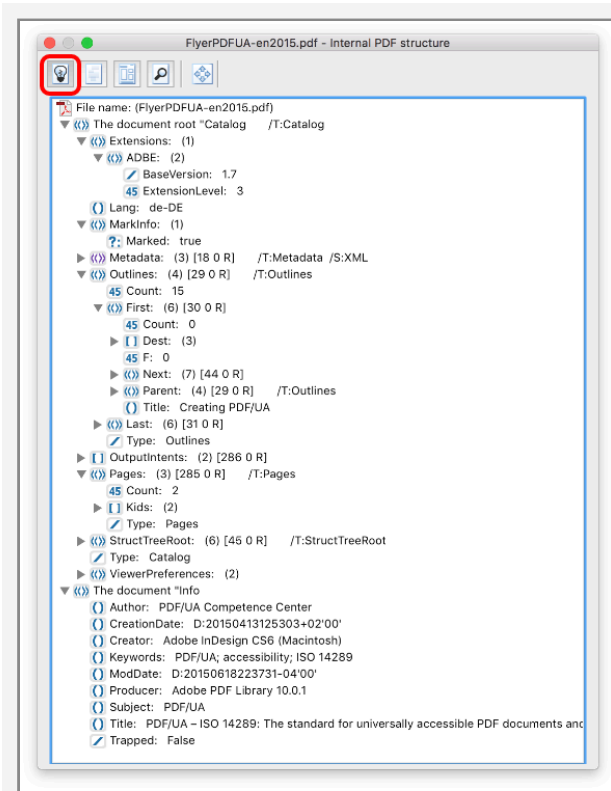
Document Structure view

The first button opens the "Document Structure" view, which lists 2 items:

- The document root "Catalog"

The root of a document's object hierarchy is the "Catalog" dictionary. The catalog contains references to other objects, defining the document's contents, outlines, article threads, named destinations and other attributes. In addition, it contains information about how the document shall be displayed on screen, such as its outline and thumbnail page images shall be displayed automatically and whether some location other than the first page shall be shown when the document is opened.
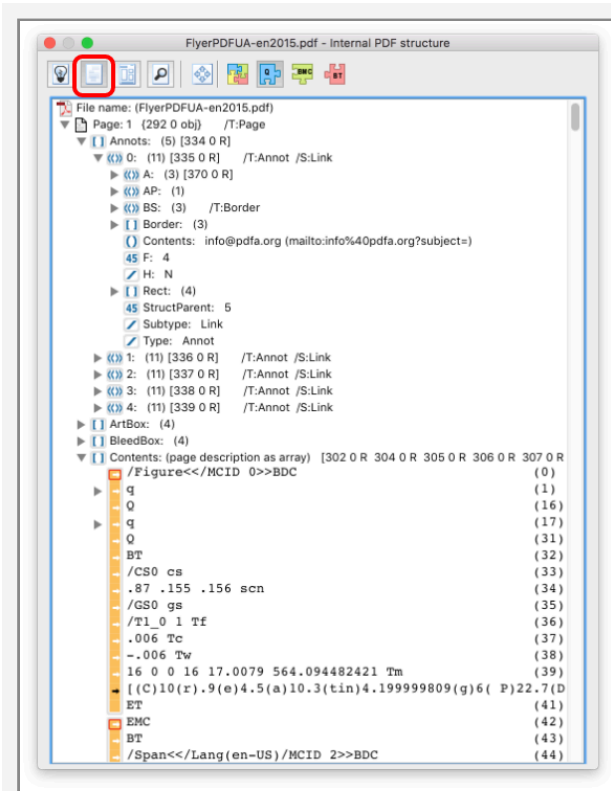
- The document info

The document info area lists some basic information about the file, like title, author, creation date, producer, creator, keywords and so on

callas

Logical Structure view

While the "Document Structure" view contains the complete view of the documents content, the "Logical Structure" view offers a page-by-page view of the different properties of a page like page geometry boxes, used resources, content stream and more as well as other optional attributes like annotations or thumbnails.
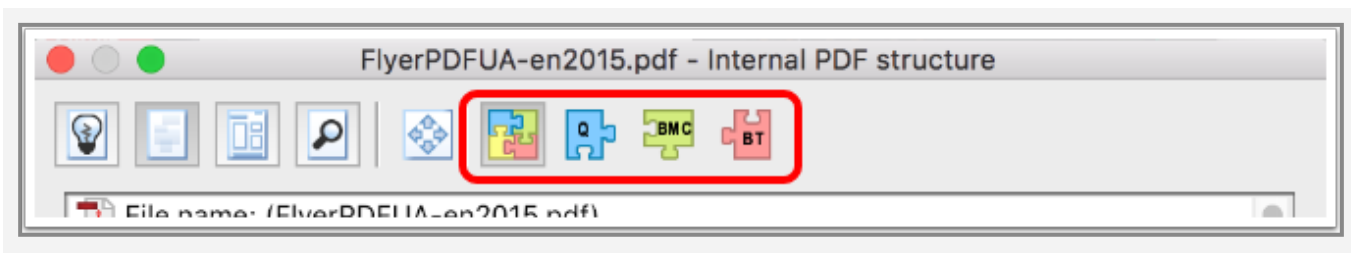
Different views of the content stream

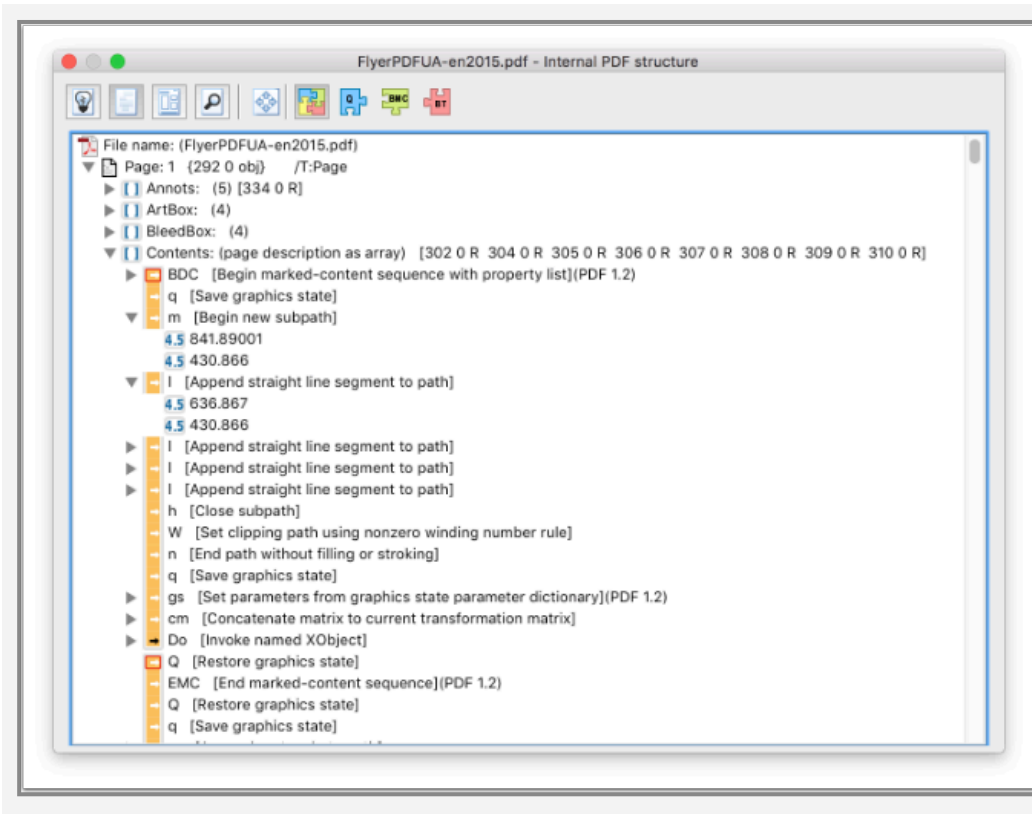The "Logical Structure" view offers 4 different representations of the content stream:

- Content Stream snippets: explained
- Content Stream snippets: q/Q pairs
- Content Stream snippets: Marked content
- Content Stream snippets: text

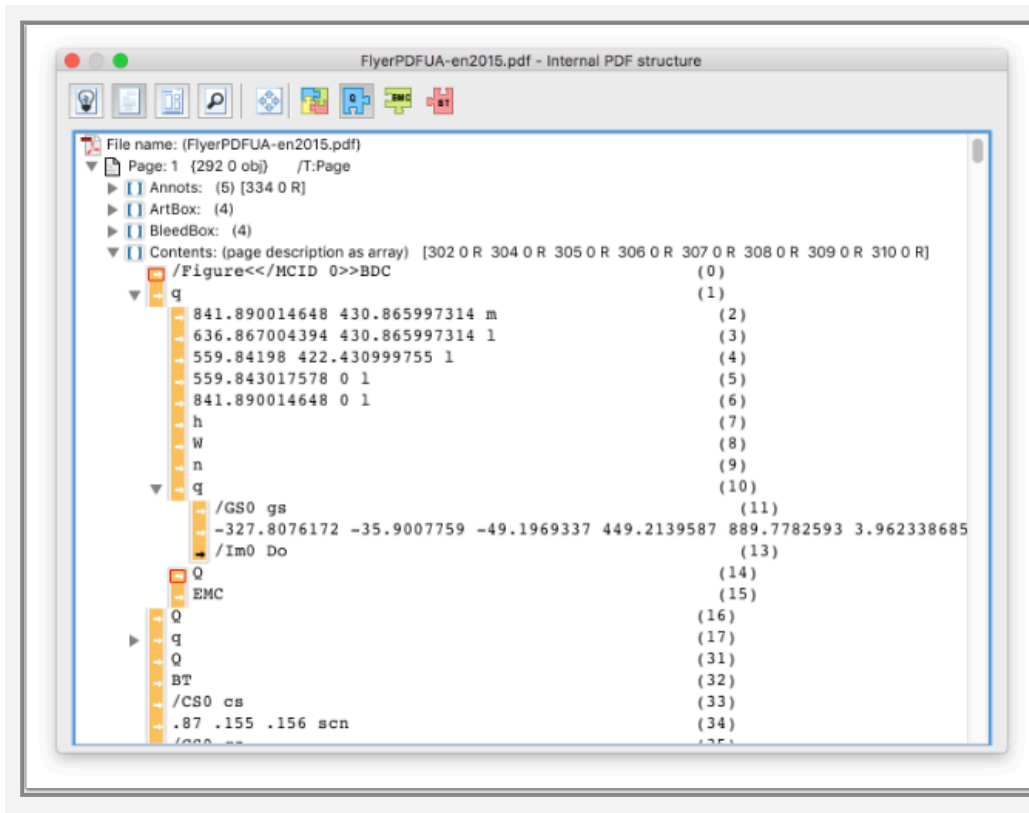These views can be selected using the 4 colored buttons on the right of the selection bar:

Content Stream snippets: explained

This mode shows minute explanations for all painting sequences of the content stream and makes it easy to understand the way the content of a page is composed step by step.
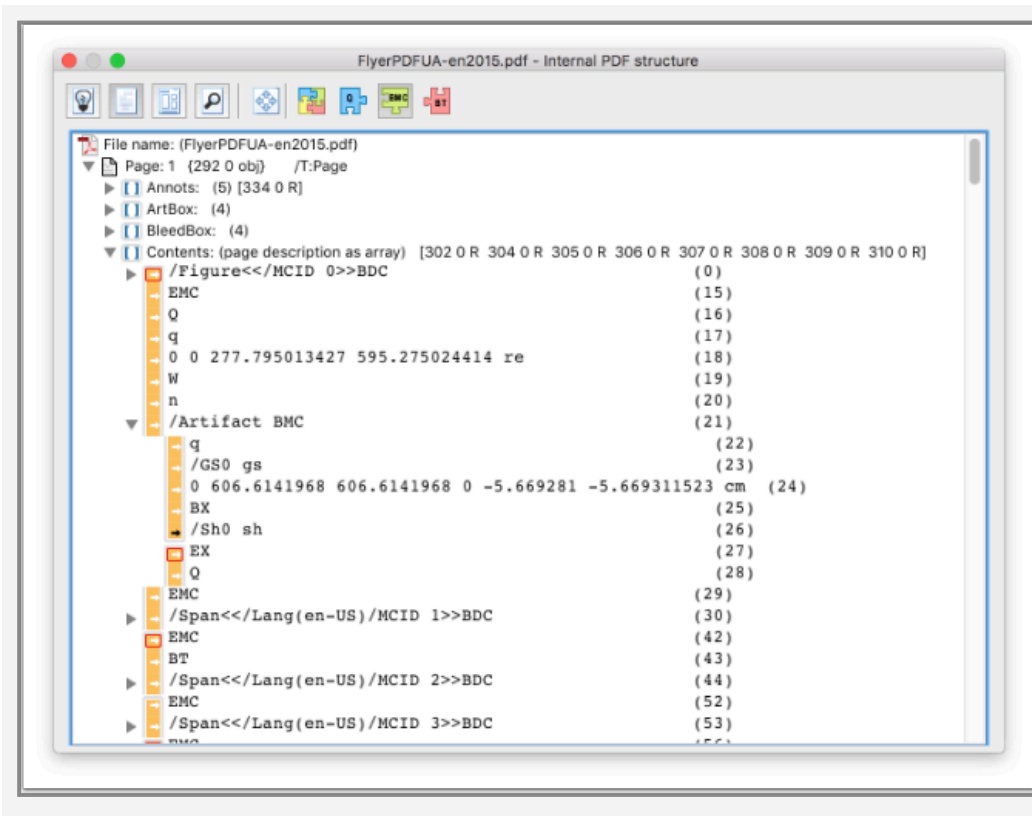

Content Stream snippets: q/Q pairs

The "q/Q pairs" view shows the content stream in a more condensed way, as the painting sequences are sorted in their respective graphic state nesting, also known as q/Q pairs.
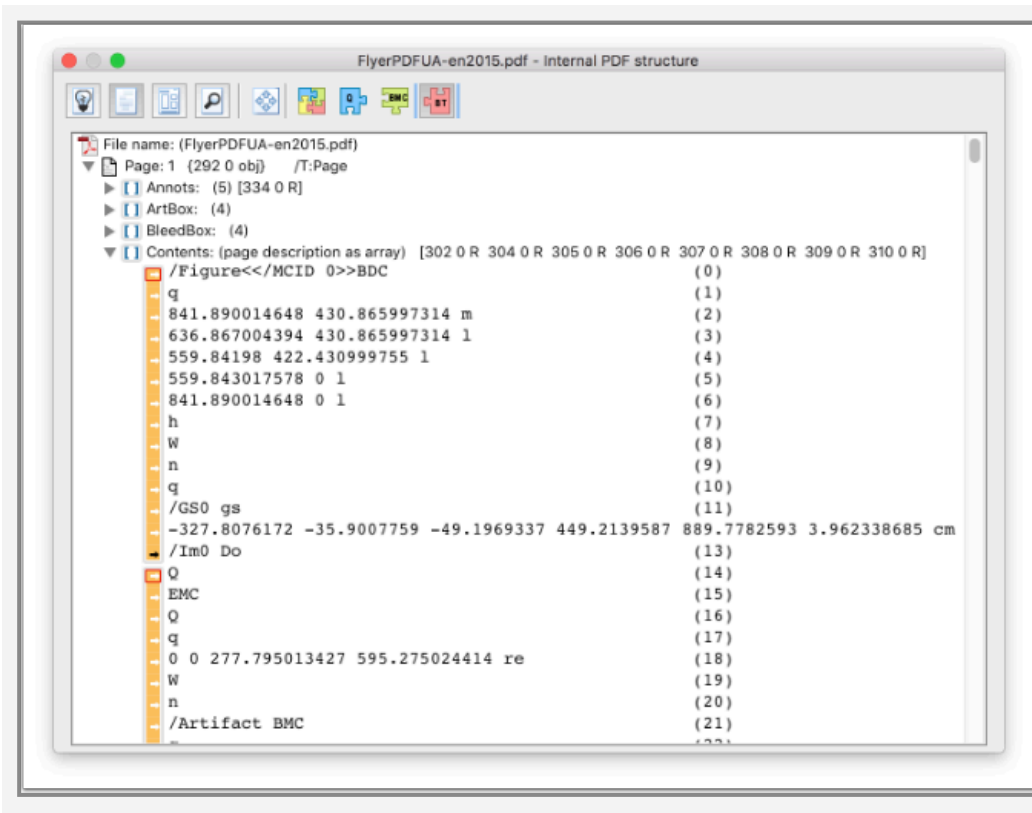
callas

Content Stream snippets: Marked content

The "Marked content" view shows the content stream from the tagging or marked content perspective, as the content stream is grouped by the respective BMC or BDC properties.
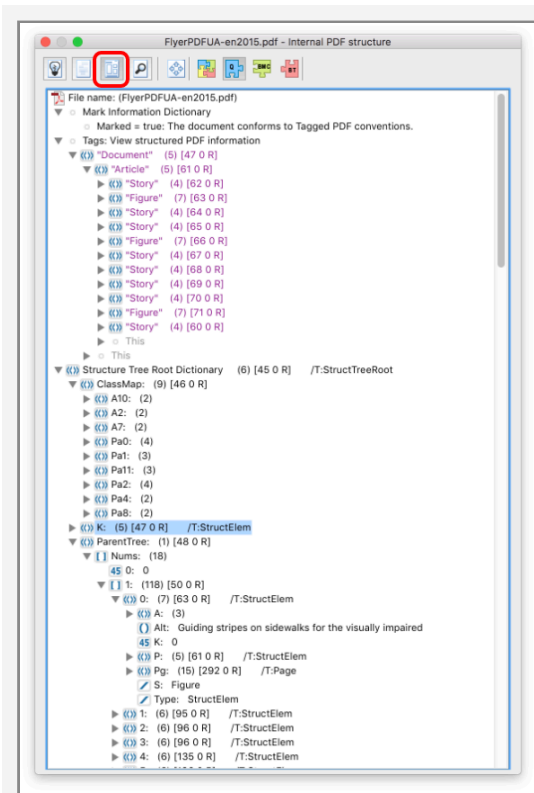
Content Stream snippets: text

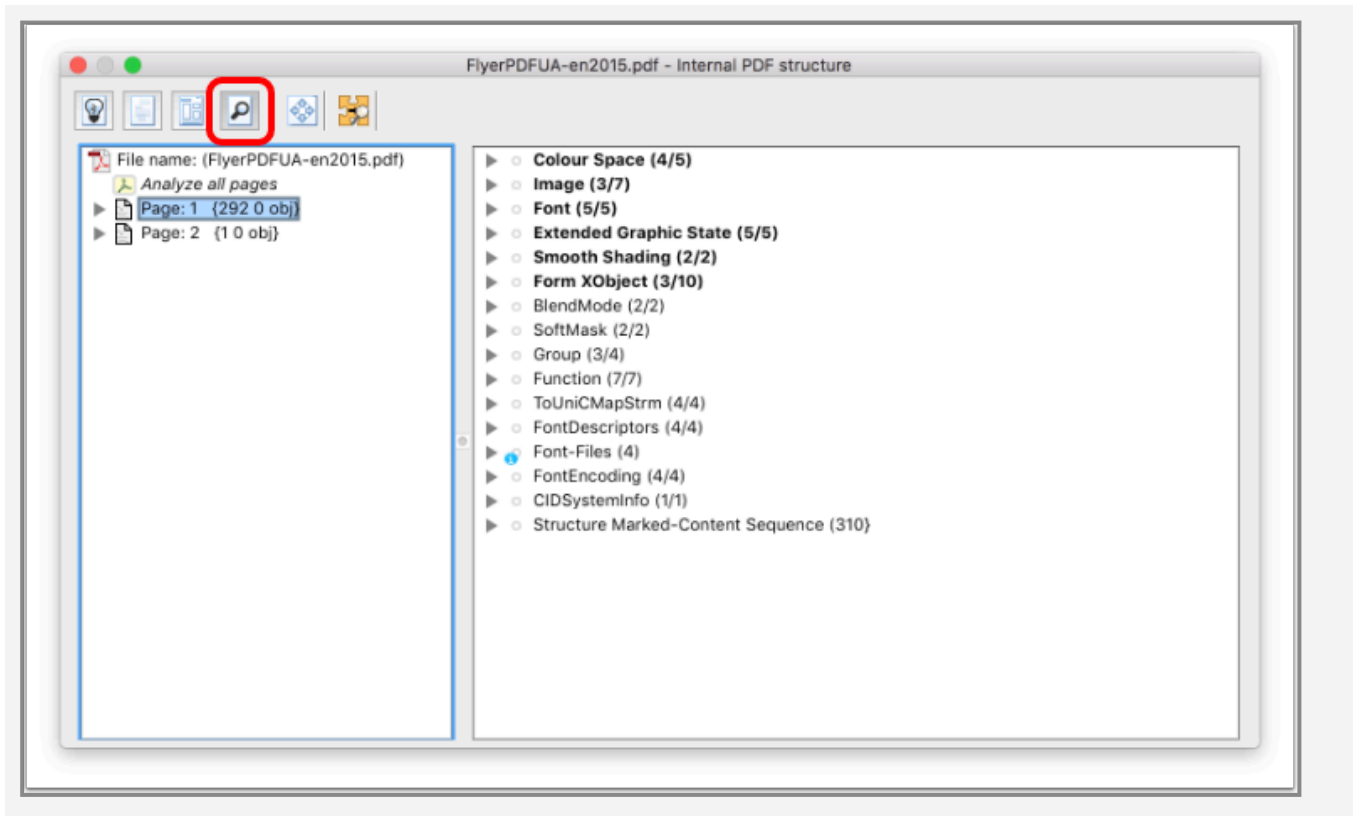The fourth view offers a plain text view of the content stream
in a readable fashion.

Tagging Stucture view

When a PDF file contains tagging information, this view gives a detailed overview about all the details like the ClassMap, a structural view of the various tags as well as other interesting details.
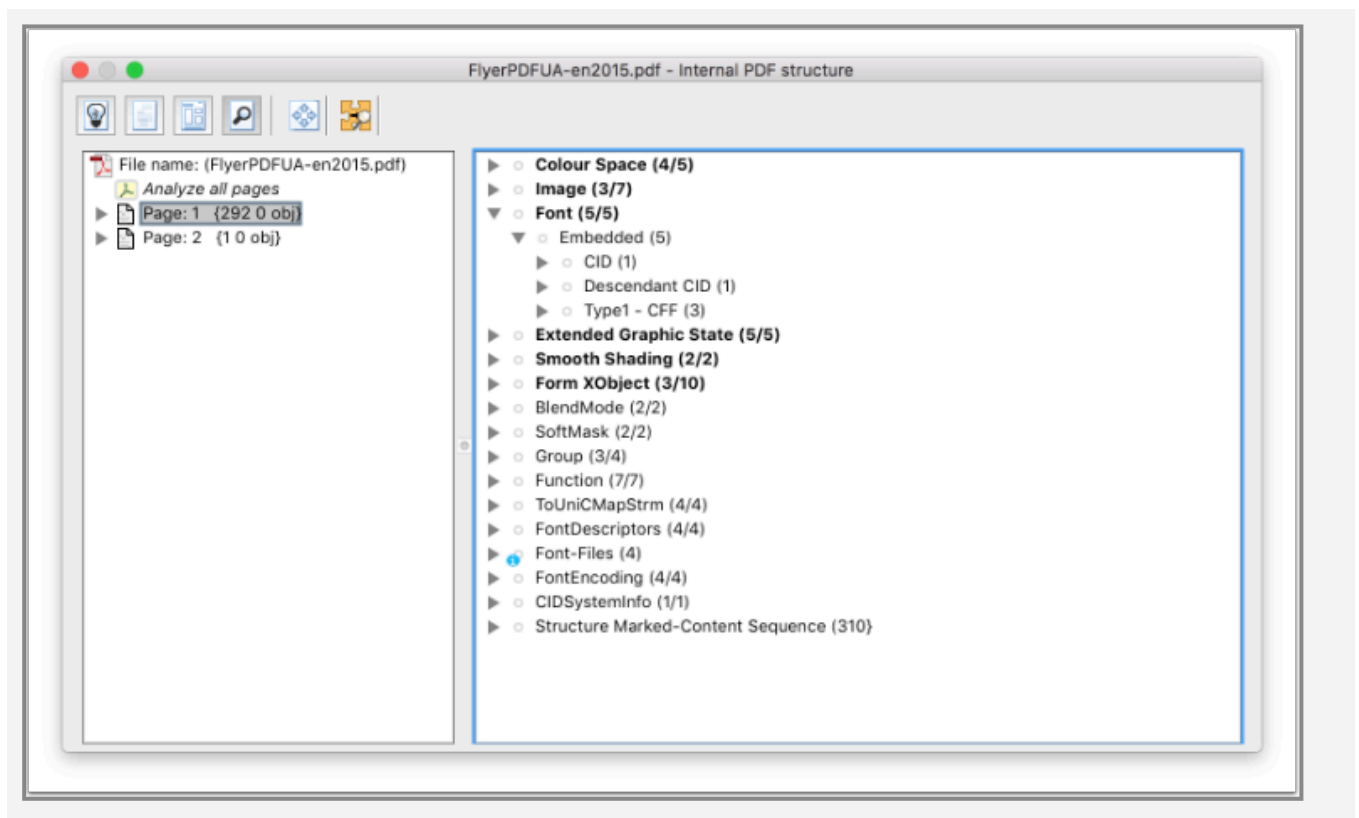
Resource view

While the Logical view gives you (amongst others) a view into the content stream, the "Resource" view offers a page-by-page listing of all painting objects (like images, shadings, text, vector objects, ...).
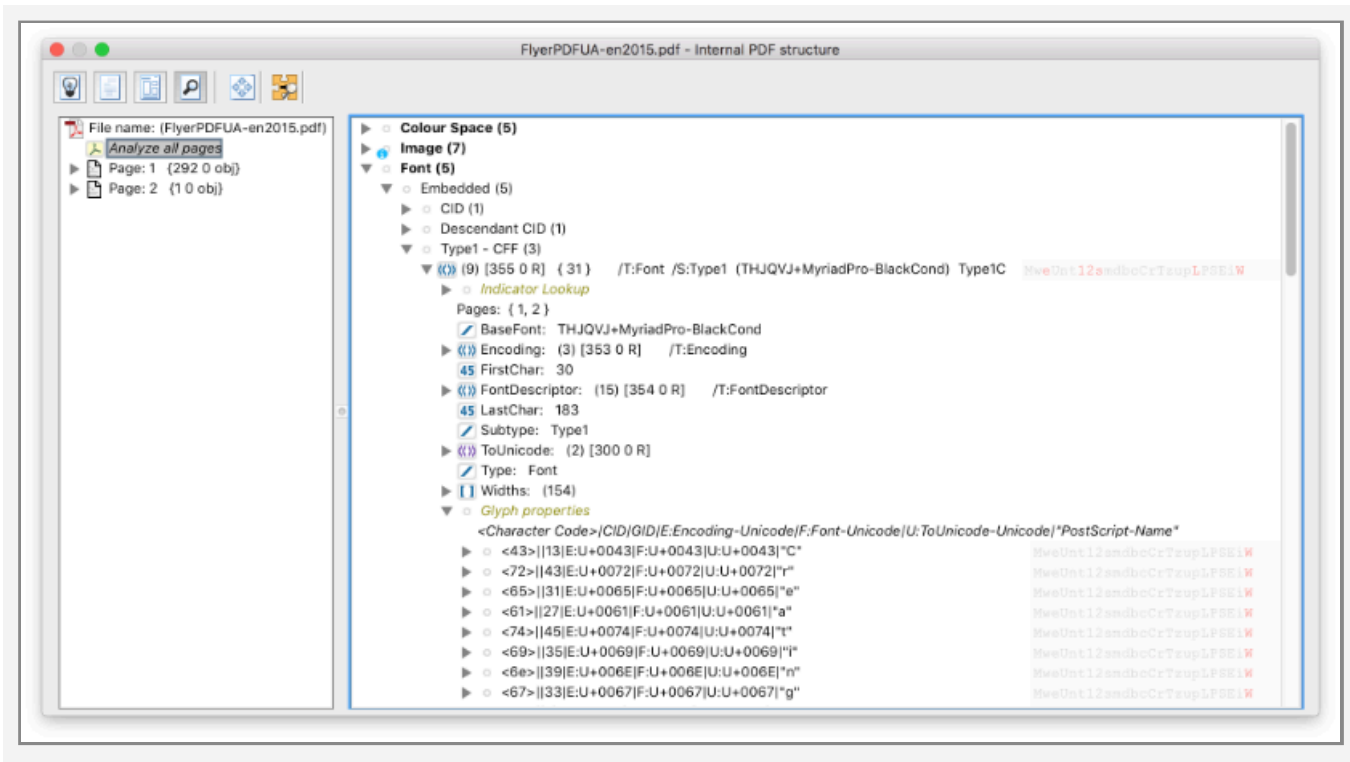
Each of the resources has a specific substructure with further information. Also all resources are listed that are used on the selected page, independent from whether they are specified in the page resource or in Form XObject resources.

The "Font" section is specifically rich with many results from pdfToolbox' font engine.

Detailed glyph information for embedded fonts

The font section shows embedded and not embedded fonts and then font types. The results of the font engine are available for the whole font and for each glyph by a list of indicators behind the respective entries.

callas

When selecting a font, a lot of detailed information about the font file itself and the contained glyphs is available. Depending on if the selection on the left pane is on a specific page or on "Analyze all pages", the fonts used on that page or in the whole document are listed.

For all glyphs of an embedded font, there are several indicators behind each glyph. If such an indicator is red, this means that the corresponding property of the indicator applies to the glyph. This does not have to be a problem right away, it can help to make the different properties of the glyphs quickly accessible.

In this example for almost all glyphs a capital "W" and for some glyphs, also "e" and "s" are active - as indicated by the indicator being red. The section "Indicator lookup" explains the indicators.
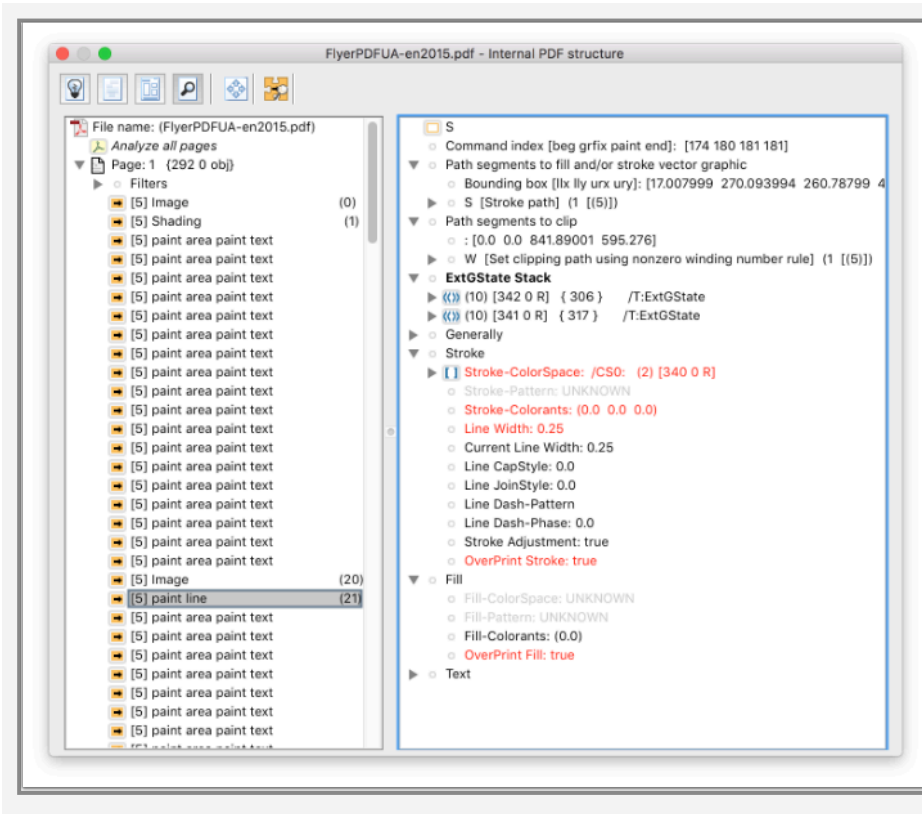
A list of the available indicators can be found at "Indicator lookup" entry. Please note, that the order of the indicators has to be considered.
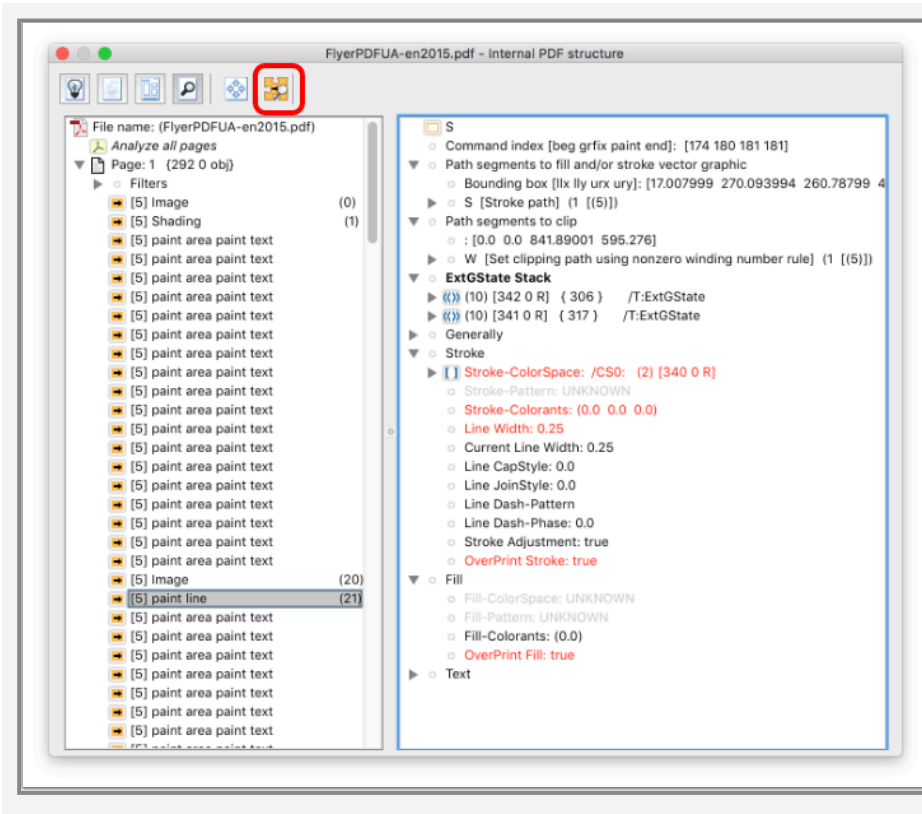
The indicator lookup informs us that the capital "W" means that the glyph width is used for positioning (the glyph is not positioned using coordinates but the width of a previous glyph). "e" stands for glyphs without contour and "s" is for such empty glyphs with a width, so in fact the respective glyph is a whitespace.

Analyze snippets and export selected snippets to a new PDF

For each object, a detailed view shows the painting and clipping area, the used color space as well as a lot of other information of the current Extend GraphicState, a used transformation matrix, for text the used font and font size, blend spaces, overprint modes and much much more.

The "open snippet as PDF" icon at the top of the Resource
view allows for creating a PDF from the current selection in
the left pane. The selection might also include several objects
and Filters are provided to select all object of a certain type.
Such PDF parts can be used for analysis to simplify a PDF.

The new PDF will be opened separately and can be used for further investigation of the PDF.
You'll see the functionality indicated by the red rectangle in the screen shot above.

PDF sample file used

The attached file has been used to show the various views of "Explore PDF".
This file has been created by the PDF Association.

FlyerPDFUA-en2015.pdf