Ghent Workgroup

10 things you might not know about fonts, but really should

David van Driessche, Christian Blaise
Executive Director & co-chair Specifications Subcommittee,
Marketing Officer & co-chair Packaging Subcommittee

September 23, 2020

David

- Started working at Enfocus in 1996
  (Tailor, PitStop…)
- Helped to start the Ghent Workgroup in June 2002
- Co-founded Gradual Software in 2004
  (Switch)
- CTO at Four Pees from 2008
  (working with callas software, axaio software, Elpical, InSoft…)

Ghent Workgroup

Christian

- 30 years of experience in several companies and roles:
  - Gap Systems (Implementation Consultant & COO)
  - Enfocus (International Sales Account Manager)
  - Nestlé (eBusiness Project Manager & Graphic Chain Expert)
  - Société Alsacienne d'Aluminium, now AMCOR (Prepress Manager)
- Founder & Owner of bleuprocess (2007), founder & CEO of agileStreams (2018), offering consultancy in process automation & in packaging

- 1 -

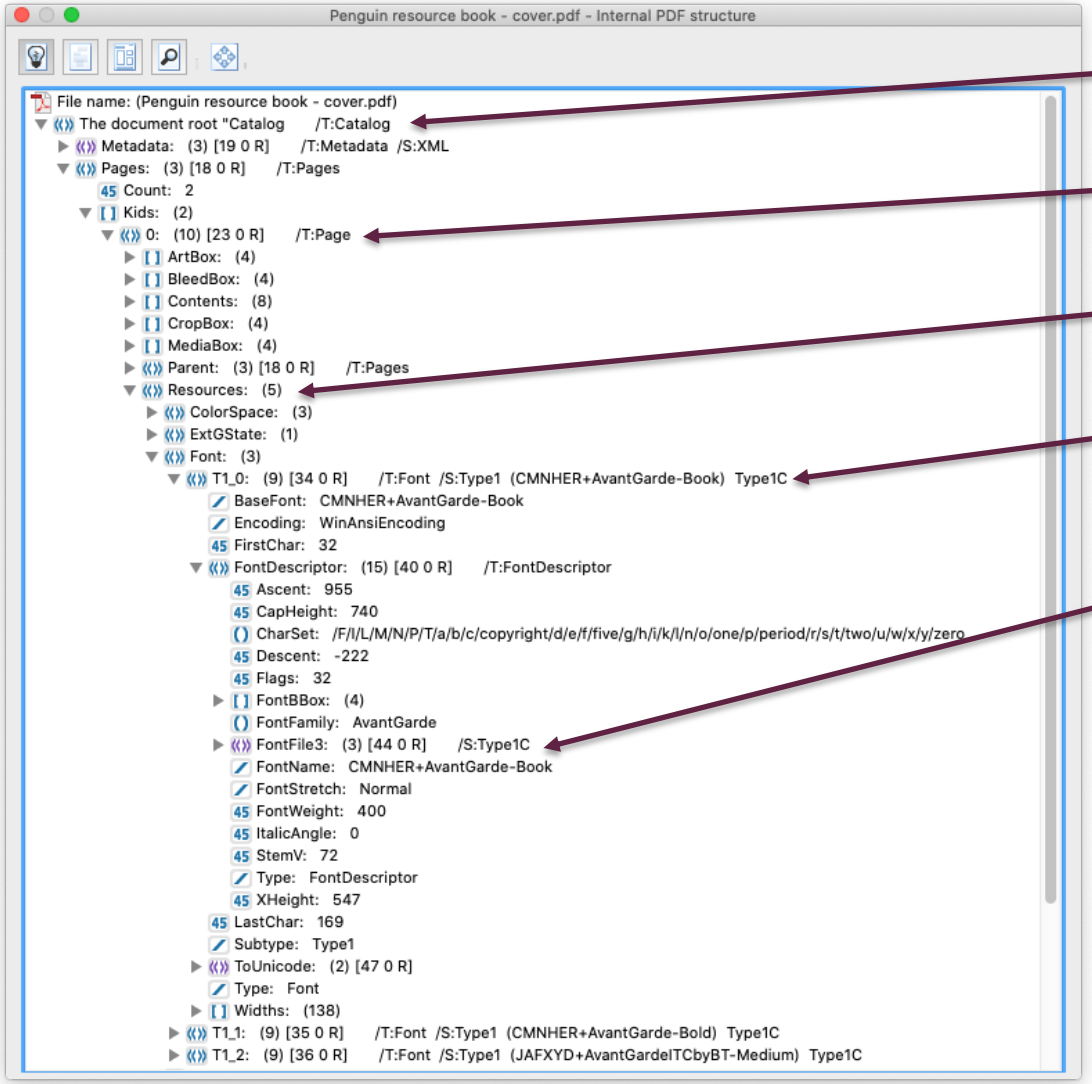
Where Do fonts live in PDF?

Ghent Workgroup

On your computer

Ghent Workgroup

In a PDF file



Document

Page

Resources

Font

Embedded font file

What's the essence

- Each font appears in the resources of pages where it's used
  - As an "object", or "entity"
- It consists of two big parts
  - Metadata about the font that is stored directly in the PDF as PDF objects
  - An embedded font file
    - It's optional
    - It's not the same as what lives on your system

- 2 -

Glyphs

Ghent Workgroup

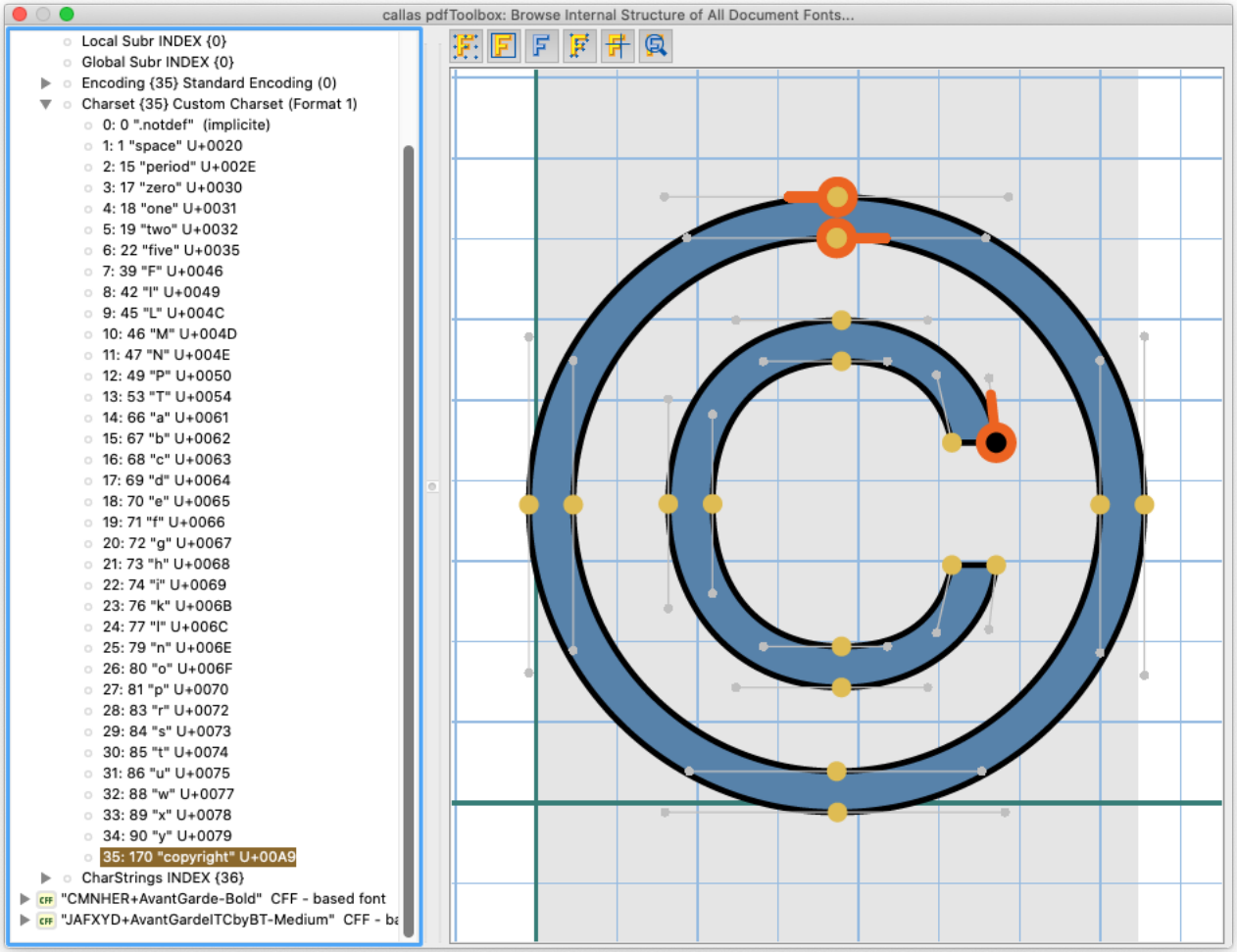
What is a glyph?

A simple definition

- A glyph is a shape used to draw a certain character.
- When talking about glyphs:
  - We're talking about the shape of the character, not the meaning of the character
  - The same character can be represented by different glyphs 

Ghent Workgroup

How does that look inside of a font?



- 3 -

Encodings

How do we go from fonts in the PDF, to glyphs shown on the page?

- We know PDF contains font information in the document.
- We also know fonts contain glyphs, which can be used to draw on the page…

- How does PDF tie those two together?

Ghent Workgroup

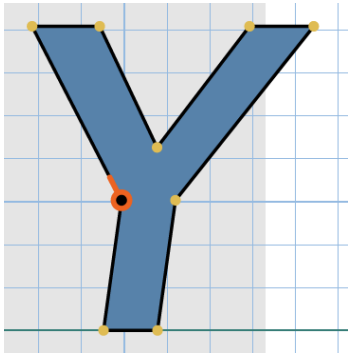
Page description to printed page

```
BT
  .037 .145 .886 0 k
  /TT0 1 Tf
  72 0 0 72 254.495193481 774.448791503 Tm
  [(Y)92(ou!)]TJ
  30 0 0 30 188.514404296 702.448791503 Tm
  (have been invited...)Tj
ET
```

Begin text object
Set CMYK yellow
Set font “TT0”
Set text matrix (location / size)
Draw “You!” (with kerning)
Set text matrix (location / size)
Draw more text
End text object

In details for “Y”

- Page description contains “Y”, which corresponds with number “89”.
- Font with name “TT0” contains a character – to – glyph table, that maps “89” to “glyph 59”.
- Font with name “TT0” contains a glyph table that defines “glyph 59” as



Ghent Workgroup

That's complex…

- You don't know the half of it. Different font types contain different mapping mechanisms. There are a lot of moving parts. But…

- In essense: each font contains a mechanism to translate numbers on the page to the glyph they correspond with. The collective name for this is *encoding.*
  (just know that's a bit of an over-simplification)

- 4 -

Font Size & x-Height

Ghent Workgroup

Readability and legibility

- Deliver a message: it’s not only about printing content… it is also about being able to read this content afterwards

- Texts must be readable and legible
  Oh wait, isn’t it the same? Well… no!

Readability

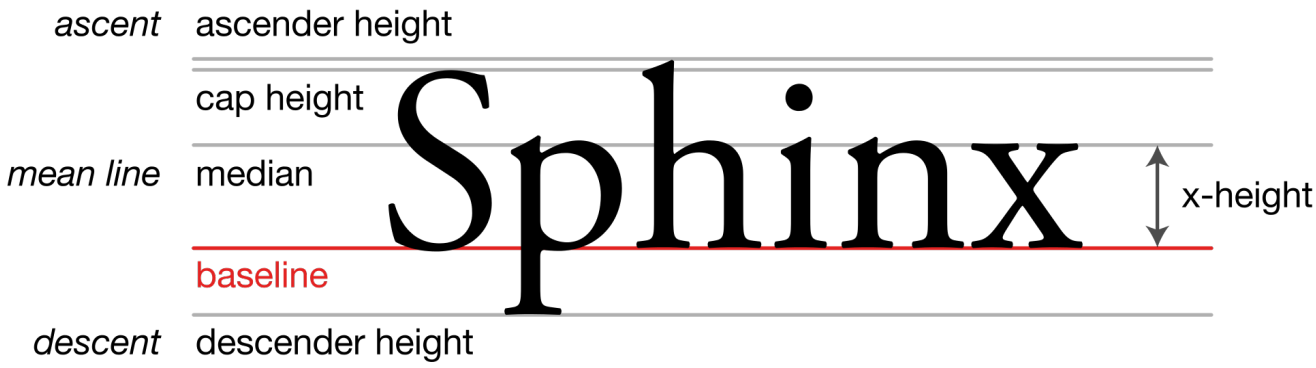
- Readability is related to how the type is arranged, or typeset, and therefore is controlled by the designer
  - Font size
  - Case
  - Line spacing
  - Line length
  - Color & contrast

Legibility

- The legibility of a typeface is a product of its design, and relates to the ability to distinguish one glyph from another when reading
  - x-height
  - Character width
  - Weight
  - Serif or lack thereof

Font size & x-height

- There are 2 vertical metrics that are key to readability and legibility
  - Font size: the size of text characters
  - x-height: distance between baseline and mean line of a lowercase letters (typically the letter x)



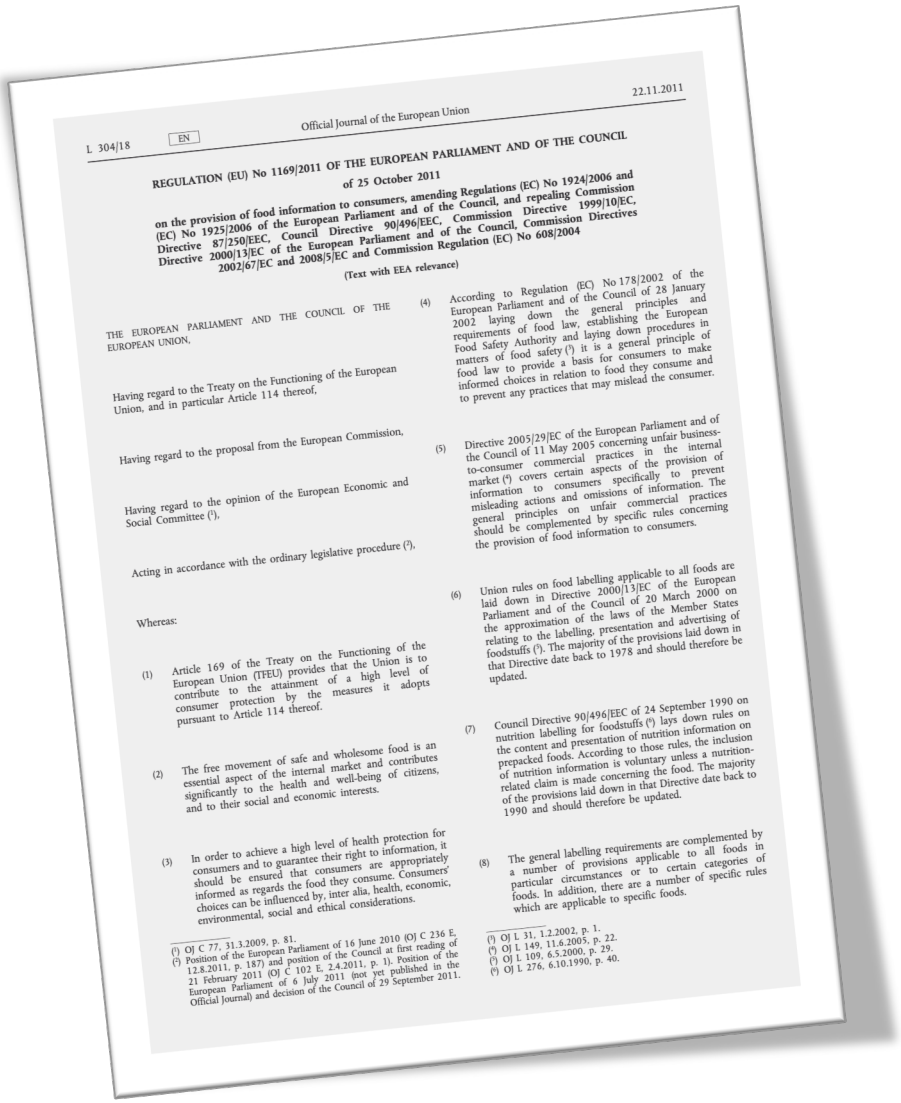
Checking size of a text

- To check the « quality » of a text, combining checks on font size and x-height is a sensible approach

- Checking x-height is more accurate, as font size can be affected in some conditions and would give bad preflight results
  - x-height is used by regulations, especially in packaging to ensure a minimum text size for important information on the package

Ghent Workgroup

EU Regulation 1169/2011 (INCO)



L 304/18 EN Official Journal of the European Union 22.11.2011

REGULATION (EU) No 1169/2011 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 25 October 2011

on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004

(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 114 thereof,

Having regard to the proposal from the European Commission,

Having regard to the opinion of the European Economic and Social Committee [1],

Acting in accordance with the ordinary legislative procedure [2],

Whereas:

(1) Article 169 of the Treaty on the Functioning of the European Union (TFEU) provides that the Union is to contribute to the attainment of a high level of consumer protection by the measures it adopts pursuant to Article 114 thereof.

(2) The free movement of safe and wholesome food is an essential aspect of the internal market and contributes significantly to the health and well-being of citizens, and to their social and economic interests.

(3) In order to achieve a high level of health protection for consumers and to guarantee their right to information, it should be ensured that consumers are appropriately informed as regards the food they consume. Consumers' choices can be influenced by, inter alia, health, economic, environmental, social and ethical considerations.

(4) According to Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety [3] it is a general principle of food law to provide a basis for consumers to make informed choices in relation to food they consume and to prevent any practices that may mislead the consumer.

(5) Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market [4] covers certain aspects of the provision of information to consumers specifically to prevent misleading actions and omissions of information. The general principles on unfair commercial practices should be complemented by specific rules concerning the provision of food information to consumers.

(6) Union rules on food labelling applicable to all foods are laid down in Directive 2000/13/EC of the European Parliament and of the Council of 20 March 2000 on the approximation of the laws of the Member States relating to the labelling, presentation and advertising of foodstuffs [5]. The majority of the provisions laid down in that Directive date back to 1978 and should therefore be updated.

(7) Council Directive 90/496/EEC of 24 September 1990 on nutrition labelling for foodstuffs [6] lays down rules on the content and presentation of nutrition information on prepacked foods. According to those rules, the inclusion of nutrition information is voluntary unless a nutrition-related claim is made concerning the food. The majority of the provisions laid down in that Directive date back to 1990 and should therefore be updated.

(8) The general labelling requirements are complemented by a number of provisions applicable to all foods in particular circumstances or to certain categories of foods. In addition, there are a number of specific rules which are applicable to specific foods.

[1] OJ C 77, 31.3.2009, p. 81.
[2] Position of the European Parliament of 16 June 2010 (OJ C 236 E, 12.8.2011, p. 187) and position of the Council at first reading of 21 February 2011 (OJ C 102 E, 2.4.2011, p. 1). Position of the European Parliament of 6 July 2011 (not yet published in the Official Journal) and decision of the Council of 29 September 2011.
[3] OJ L 31, 1.2.2002, p. 1.
[4] OJ L 149, 11.6.2005, p. 22.
[5] OJ L 109, 6.5.2000, p. 29.
[6] OJ L 276, 6.10.1990, p. 40.

Article 13.2 & 13.3

2. Without prejudice to specific Union provisions applicable to particular foods, when appearing on the package or on the label attached thereto, the mandatory particulars listed in Article 9(1) shall be printed on the package or on the label in such a way as to ensure clear legibility, in characters using a font size where the x-height, as defined in Annex IV, is equal to or greater than 1,2 mm.

3. In case of packaging or containers the largest surface of which has an area of less than 80 cm^2, the x-height of the font size referred to in paragraph 2 shall be equal to or greater than 0,9 mm.

DEMO

- 5 -

Font Types in PDF

My PDF has a font you say?

- When PDF was invented, multiple font systems existed
  - PDF adopted three different “simple” font formats
- To allow the use of PDF in environments with *many* characters (think: Chinese, Japanese, Korean…)
  - PDF adopted “composite” fonts, normally referred to as Type 0 fonts

Simple fonts

- A simple font is a font that
  - Has a maximum of 256 usable characters
  - Has a single writing direction
- There are three types
  - TrueType
  - Type 1
  - Type 3

Composite fonts

- Only one type, called Type 0
  - Uses a complex system of included font information
  - Can address more than 256 characters
  - Has the capability to support more than one writing direction

Some more things to know

- Can a PDF file embed an OpenType font?
- Can a PDF file embed a multiple-master font?
- What are city fonts?

- 6 -

Opentype SVG fonts & Type 3

Ghent Workgroup

About Opentype SVG fonts

- These are new fonts where:
  - glyphs are SVG files
  - multiple colors can be used within the font and/or glyphs
    - Complex drawings, emoji fonts or variants 
  - Colors are RGB-based
- Examples
  - Emoji One 
  - Trajan Color 

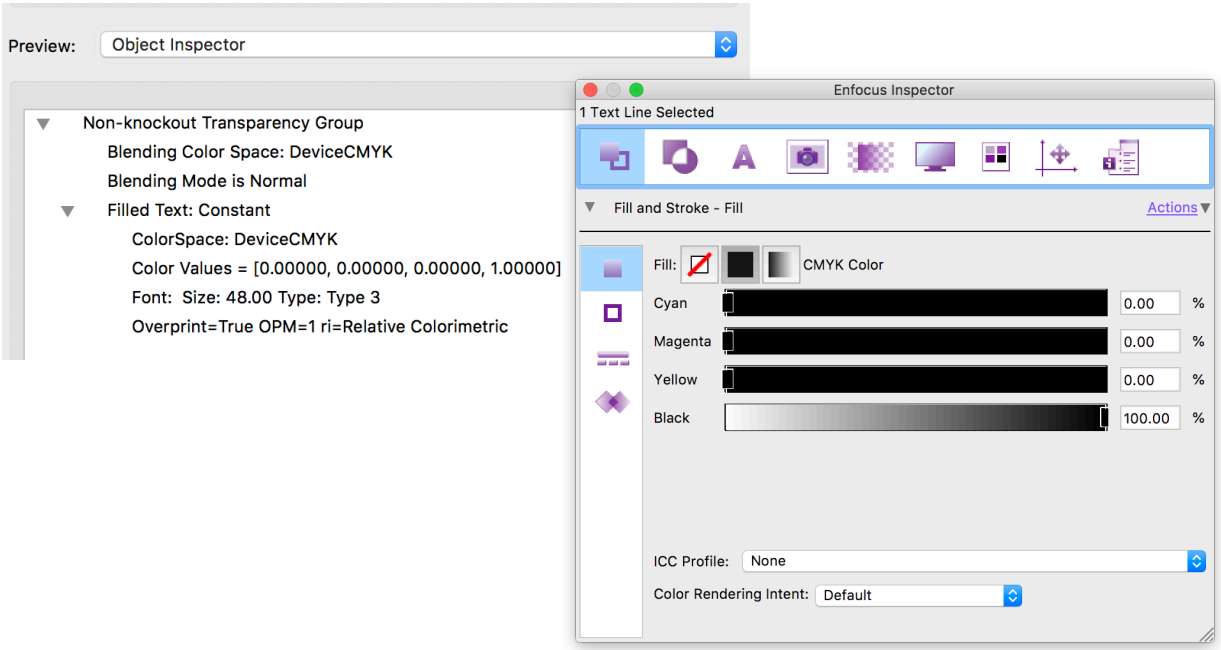
What does happen in PDF

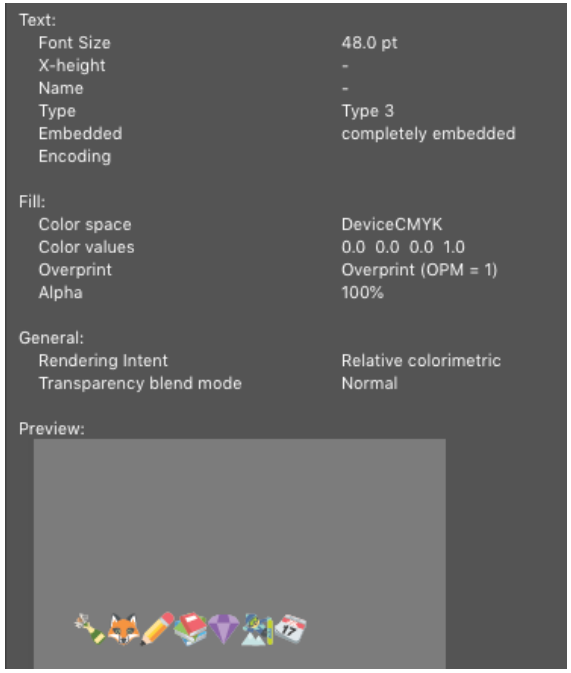
- When the original document from the authoring application is converted to PDF, the Opentype SVG is converted into a Type 3 font
  - Can also contain some parts in Type 1
  - Depending on the authoring application or the PDF standard chosen, it can also be converted to outlines or rastered
- A glyph in a Type 3 font can be almost anything!
  - From a character to a complete complex drawing

Ghent Workgroup

Color checks give…
…unexpected results!

- Text object seen as 100K
  - difficult to show a consistent information,
    as each glyph can contain many colors, fill and/or stroke

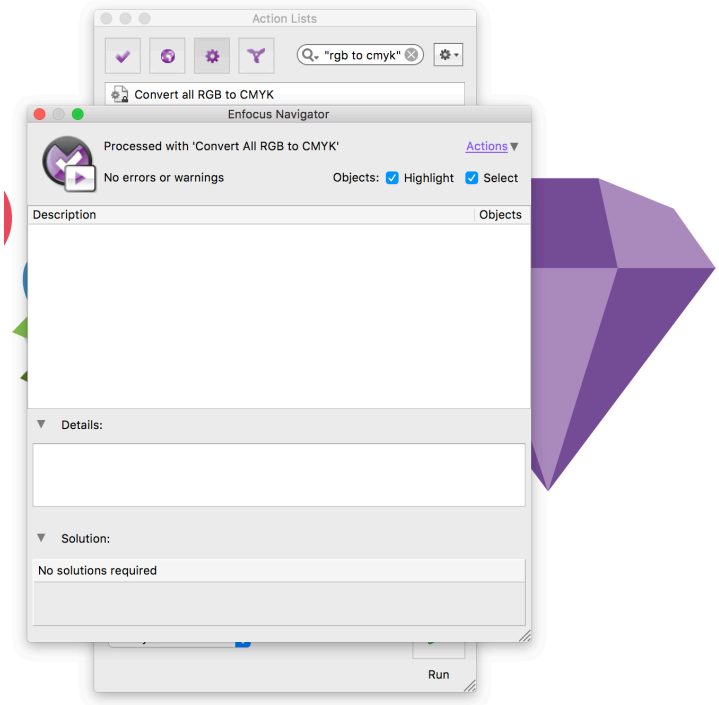


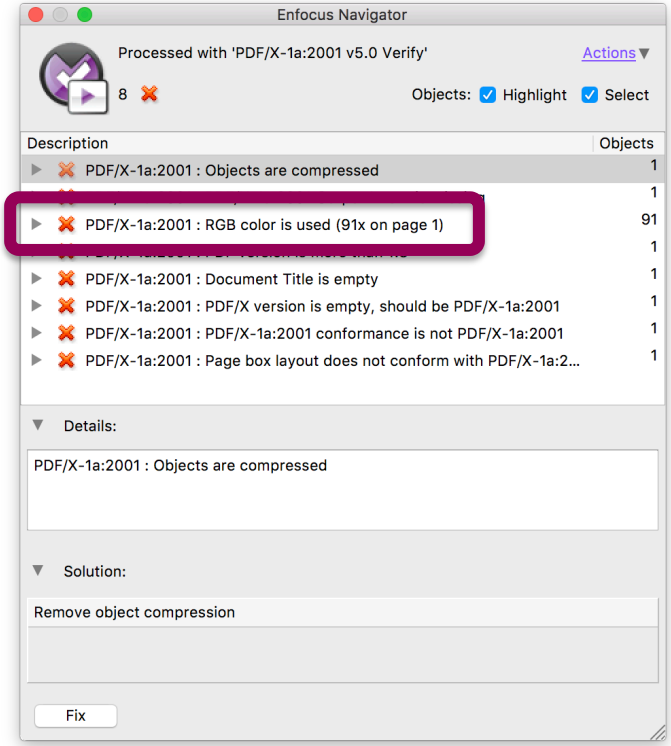


Ghent Workgroup

Color conversion & PDF standard compliance

- Some example with Enfocus PitStop
  - Nothing found to be converted, but…





Issues fixed

- Outlining the font or converting from RGB to CMYK wasn't possible or was completely destroying the file
- There were errors on opening the files in Acrobat when PitStop Pro was installed
  - the PDF engine used in InDesign or Illustrator and Acrobat are using 2 different ways to encode the font (and both were compliant to the specification!)
- Thanks to our testing and findings, PDF tools have been updated to overcome these issues
  - New engines, new corrections

DEMO

- 7 -

Character Widths and their importance

Ghent Workgroup

Page description to printed page

```
BT
  .037 .145 .886 0 k
  /TT0 1 Tf
  72 0 0 72 254.495193481 774.448791503 Tm
  [(Y)92(ou!)]TJ
  30 0 0 30 188.514404296 702.448791503 Tm
  (have been invited...)Tj
ET
```

Begin text object
Set CMYK yellow
Set font “TT0”
Set text matrix (location / size)
Draw “You!” (with kerning)
Set text matrix (location / size)
Draw more text
End text object

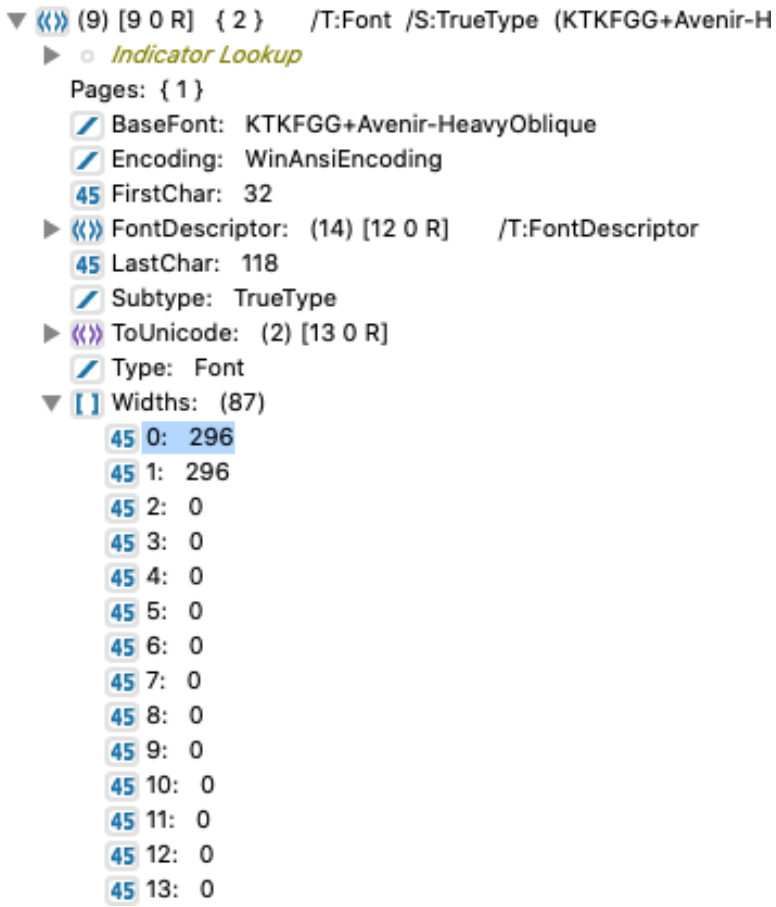
Positioning using widths

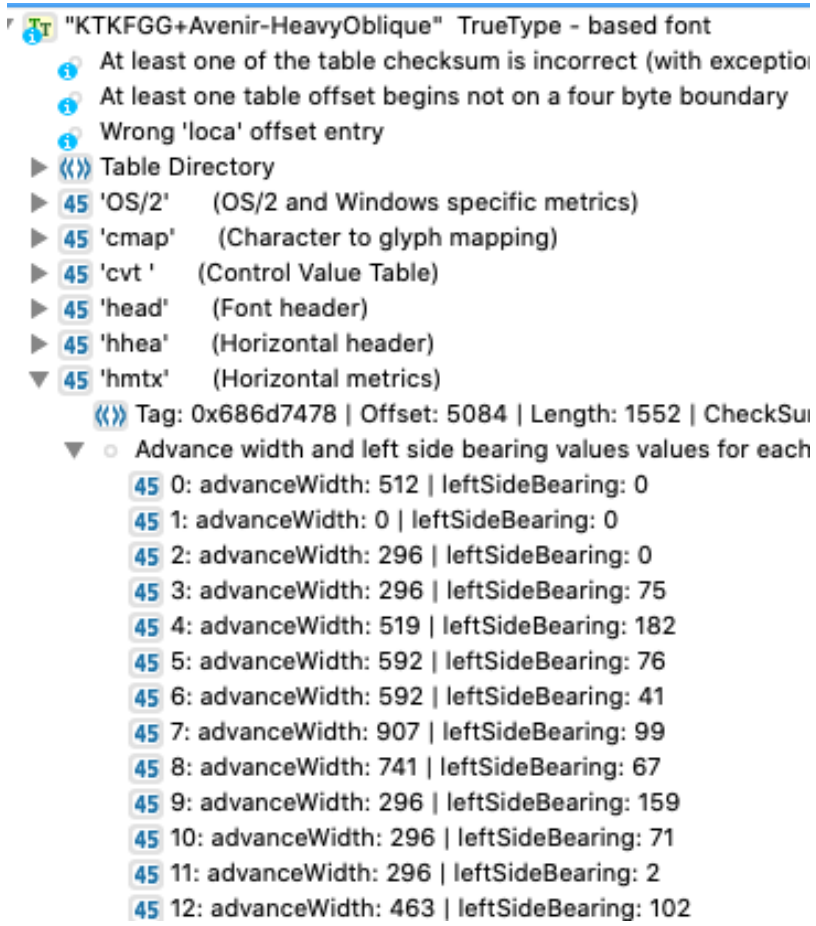
have been

430 512 315 …

Ghent Workgroup

Where do we find those widths





Potential problems

- A PDF writer might make mistakes and those two pieces of information don’t match up…
- When embedding a font, the font isn’t exactly what was already in the PDF as metadata
- The font is changed (during embedding or replacing)
  - Widths change
  - Character position on the page changes

- 8 -

Embedding And subsetting

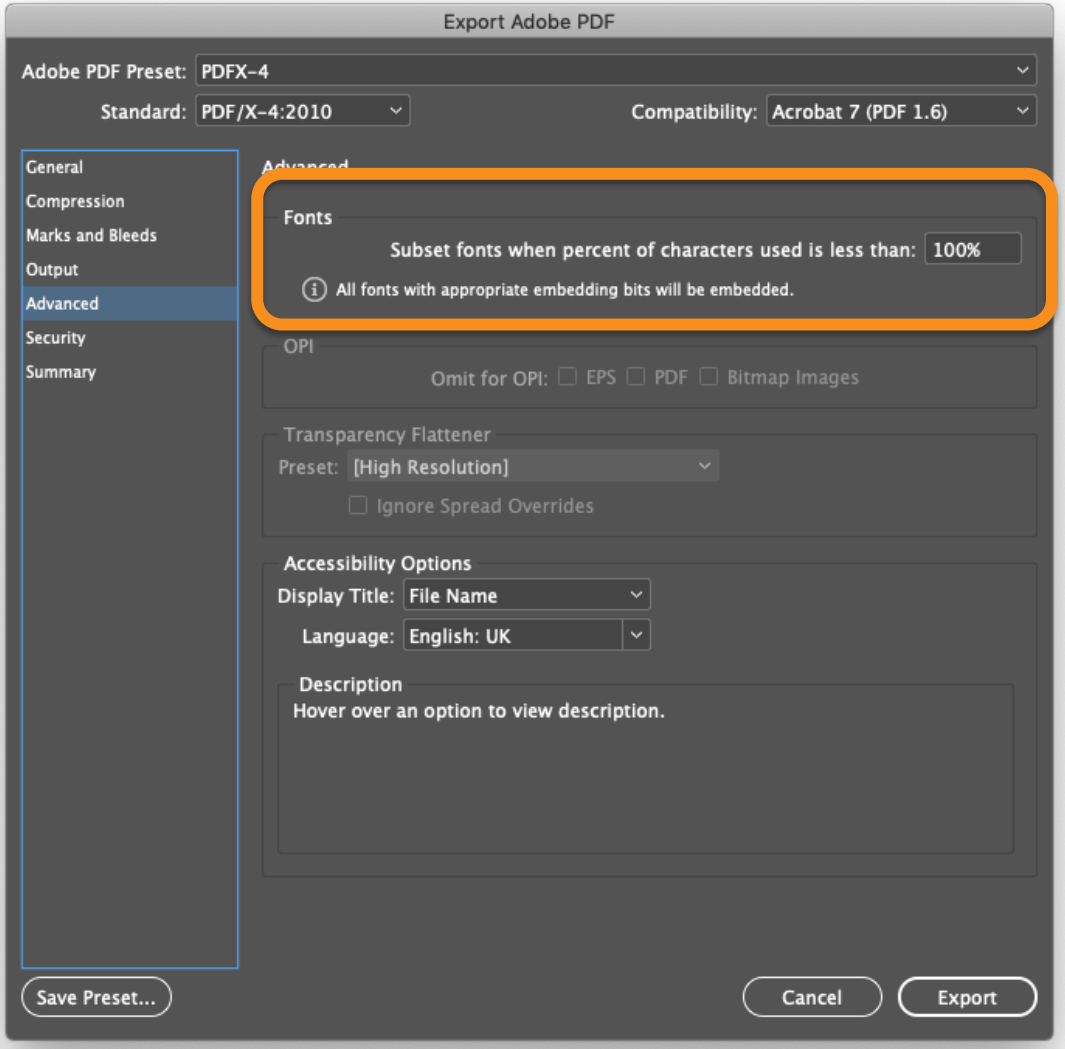
A good PDF contains fonts!

- PDF is a self-contained format, which allows to exchange a file with someone without having to deal with software versions, external resources (images, fonts, etc.), etc.

- Adding fonts to the PDF is called embedding. There are 2 ways of embedding fonts in a PDF:
  - Complete set
  - Subset

Ghent Workgroup

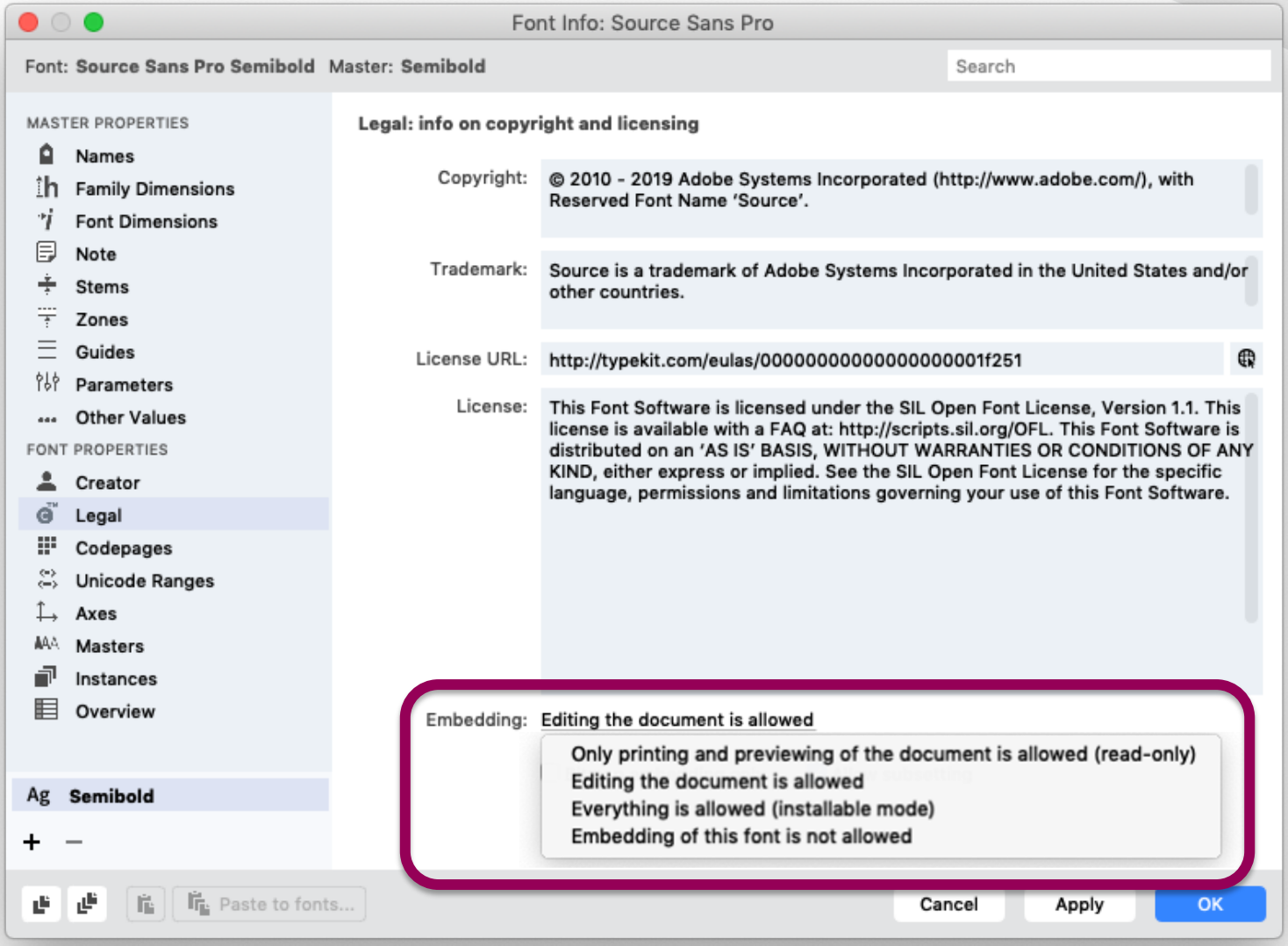
Export PDF – Font Embedding



- Use 100% to subset the font in the PDF
- Use 0% to completely embed the font /!\ Doesn't always work…
- NEVER use another number (e.g. 50%) as there is no way to expect the result!

Ghent Workgroup

Font properties - Legal



Ghent Workgroup

Subsetting a font generates a unique font name

Glyphs used in PDF:

U+0020 CID 32 space	, U+002C CID 44 comma	- U+002D CID 45 hyphen	0 U+0030 CID 48 zero	1 U+0031 CID 49 one	2 U+0032 CID 50 two	4 U+0034 CID 52 four	A U+0041 CID 65 A
B U+0042 CID 66 B	C U+0043 CID 67 C	D U+0044 CID 68 D	E U+0045 CID 69 E	F U+0046 CID 70 F	G U+0047 CID 71 G	H U+0048 CID 72 H	I U+0049 CID 73 I
N U+004E CID 78 N	O U+004F CID 79 O	P U+0050 CID 80 P	R U+0052 CID 82 R	S U+0053 CID 83 S	T U+0054 CID 84 T	W U+0057 CID 87 W	X U+0058 CID 88 X
Z U+005A CID 90 Z	a U+0061 CID 97 a	b U+0062 CID 98 b	c U+0063 CID 99 c	d U+0064 CID 100 d	e U+0065 CID 101 e	g U+0067 CID 103 g	h U+0068 CID 104 h
i U+0069 CID 105 i	k U+006B CID 107 k	l U+006C CID 108 l	n U+006E CID 110 n	o U+006F CID 111 o	p U+0070 CID 112 p	r U+0072 CID 114 r	s U+0073 CID 115 s
t U+0074 CID 116 t	u U+0075 CID 117 u	w U+0077 CID 119 w	y U+0079 CID 121 y	° U+00B0 CID 176 degree			

Font properties:

Type 1 font: "SourceSansPro-Regular"
PostScript name: SourceSansPro-Regular
Unicode CMap is present
Encoding: WinAnsiEncoding
Font is a subset
Embedded: True
Font descriptors
Maximum height above the baseline: 1000
y-coordinate of the top of flat capital letters: 660
Maximum depth below the baseline: -400
Italic angle in degrees counterclockwise: 0
Width of the dominant vertical stems of glyphs: 84
Average width of characters: 0
Desired spacing between lines: 0
Maximum width of characters: 0
Width to use for un-encoded characters.: 0
Vertical width of the dominant horizontal stems of glyphs: 0
y-coordinate of the top of flat non-ascending lowercase letters: 486
PostScript name: XFNXLC+SourceSansPro-Regular
Various properties
Fixed-width font: False
Serif font: False
Symbolic font: False
Script font: False
Uses the Adobe Standard Roman Character Set: True
Italic: False
All-caps font: False
Small-caps font: False
Force bold at small text sizes: False

- When the font is embedded as subset in a PDF, this generates a unique font name
- This ensures that the font will not be replaced by another one with same name later in the process, e.g. during the ripping process

DEMO

- 9 -

The Mystery of the Missing Glyph

Missing glyphs

- What happens if you attempt to draw a character, but the font doesn't have a glyph for it?
  - (Yes, that should never happen in the first place)
  - (but of course it does!)
- In that case, fonts use glyph 0
  - Usually a box
  - Sometimes with something in it

What's the problem with this?

- "Character references .notdef glyph"
- This means that the font incorrectly uses this glyph 0, usually as an "optimisation" to draw spaces.
- So what happens when this font is now replaced by another – correct – font?
  - (mayhem!)

- 10 -

The Meaning of Text - ToUnicode

Ghent Workgroup

www.gwg.org

The problem with meaning

- PDF was developed as a “visual” format. Optimized for on-screen viewing or printing. So when it was developed:
  - PDF didn’t retain document structure. There is no notion of words, sentences, lines of text or paragraphs
  - There was no definition of reading order. Text can be on the page in any order (often illogical)
  - Text was optimized for drawing, without what was necessary to understand the meaning of it

Ghent Workgroup

Why is that a problem?

A simple font encoding – straightforward to see what the text might mean

```
BT
  .037 .145 .886 0 k
  /TT0 1 Tf
  72 0 0 72 254.495193481 774.448791503 Tm
  [(Y)92(ou!)]TJ
  30 0 0 30 188.514404296 702.448791503 Tm
  (have been invited...)Tj
ET
```

A complex font encoding – these numbers map to glyphs, but what do they mean?

```
BT                                                                    (29)
  1 0 0 -1 0 0 Tm                                                       (30)
  [(\000M\000 \000A\0001\000a\000i\000n\000 \000J\000A\000M\000E\000S)]TJ  (31)
ET                                                                    (32)
```

ToUnicode table

- To solve this problem, fonts *can* contain an ToUnicode table
  - Gives a translation from character codes to Unicode
- Usually you know this is missing if:
  - You can’t search a PDF
  - You get nonsense characters if you
    - Extract the text from the PDF
    - Copy and paste the text into a text editor

DEMO

Ghent Workgroup