
Working With PDFs in Ruby

Working With PDFs in Ruby

Thomas Leitner

@gettalong

@_gettalong

Note

Shameless plug for one of my projects ;)

... but **you** might find it useful, too!

Curent Situation

- *Prawn* generates PDFs
- *pdf-reader* reads PDFs
- *combinepdf* has basic PDF merge functionality
- *origami* is intended for auditing PDFs
- Various Java libraries for JRuby (e.g. PDFBox)

Problems

- No complete, integrated solution
- PDF specification only partly supported
- Performance/Memory consumption

What if...?

- there was a library for **reading and writing PDFs**,
- that supports **most parts of the PDF specification**,
- supports content generation similar to Prawn,
- can easily be extended if needed,
- and is written in **pure Ruby**?

Meet HexaPDF

Goal: **One library for all things PDF**

- Reading, modifying, writing PDFs
- Encryption

- Content Generation
- Basic validation of PDFs
- No Dependencies
- Fast and memory efficient

Fully Tested

Read-Optimize-Write Performance

e.pdf	Time	Memory	File size
hexapdf	1.006ms	46.960KiB	21.770.465
origami	3.301ms	153.628KiB	21.796.847
pdftk	681ms	123.152KiB	21.874.883
qpdf-comp	1.517ms	65.172KiB	21.787.558
smpdf	37.539ms	647.440KiB	21.188.516

Full benchmark at <https://gist.github.com/gettalong/>

Text Output Performance

Comparison with Python's reportlab and Prawn, using reportlab's odyssey demo as benchmark.

reportlab w/ C extension	0.27s	930 pages/second
reportlab w/o C extension	0.67s	350 pages/second
Prawn	6.46s	35 pages/second
HexaPDF	0.96s	249 pages/second

Full benchmark at <https://gist.github.com/gettalong/>

Less Talk — More Code

Examples of what **HexaPDF** can do **today**

PDF Compression

```
HexaPDF:::Document.open(ARGV.shift) do |doc|
  doc.task(:optimize, compact: true,
          object_streams: :generate, compress_pages: false)
  doc.write(ARGV.shift, validate: true)
end
```

Basic PDF File Merging

```
target = HexaPDF::Document.new
output_name = ARGV.shift

ARGV.each do |file|
  pdf = HexaPDF::Document.new(io: File.open(file, 'rb'))
  pdf.pages.each_page do |page|
    target.pages.add_page(target.import(page))
  end
end

target.trailer.info[:Title] = 'Merged document'
target.write(output_name)
```

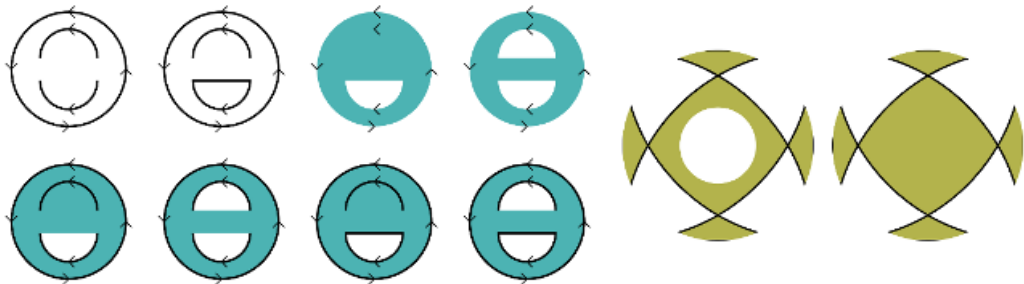
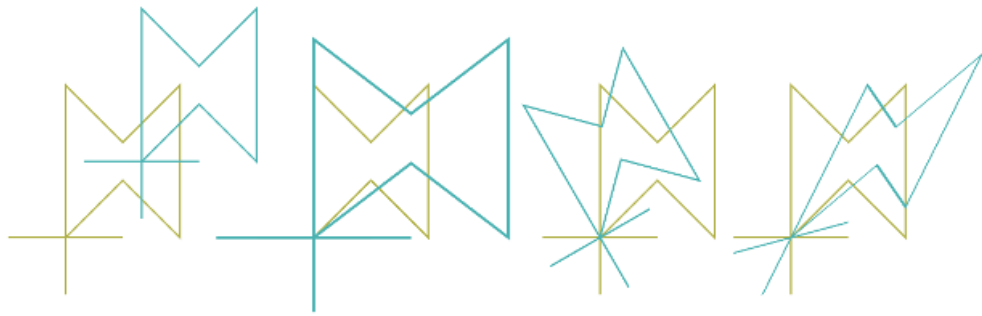
Content Generation

```
doc = HexaPDF::Document.new
canvas = doc.pages.add_page.canvas

canvas.fill_color(255, 255, 0)
canvas.line(100, 100, 200, 200).stroke
canvas.ellipse(520, 50, a: 30, b: 15, inclination: 45).fill_stroke
canvas.image("picture.jpg", at: [300, 300], width: 100)
canvas.image("vectors.pdf", at: [500, 700], height: 200)
canvas.font("Times", size: 20, variant: :bold)
canvas.text("Works!", at: [100, 750])

doc.write("graphics.pdf")
```

Example PDFs



Provided by The Internet Classics Archive.
See bottom for copyright. Available online at
<http://classics.mit.edu/Homer/odyssey.html>

The Odyssey
By Homer

Translated by Samuel Butler

BOOK I

Tell me, O muse, of that ingenious hero who travelled far and wide after he had sacked the famous town of Troy. Many cities did he visit, and many were the nations with whose manners and customs he was acquainted; moreover he suffered much by sea while trying to save his own life and bring his men safely home; but do what he might he could not save his men, for they perished through their own sheer folly in eating the cattle of the Sun-god Hyperion; so the god prevented them from ever reaching home. Tell me, too, about all these things, O daughter of Jove, from whatsoever source you may know them.

So now all who escaped death in battle or by shipwreck had got safely home except Ulysses, and he, though he was longing to return to his wife and country, was detained by the goddess Calypso, who had got him into a large cave and wanted to marry him. But as years went by, there came a time when the gods settled that he should go back to Ithaca; even then, however, when he was among his own people, his troubles were not yet over; nevertheless all the gods had now begun to pity him except Neptune, who still persecuted him without ceasing and would not let him get home.

Now Neptune had gone off to the Ethiopians, who are at the world's end, and lie in two halves, the one looking West and the other East. He had gone there to accept a hecatomb of sheep and oxen, and was enjoying himself at his festival; but the other gods met in the house of Olympian Jove, and the sire of gods and men spoke first. At that moment he was thinking of Aegisthus, who had been killed by Agamemnon's son Orestes; so he said to the other gods:

"See now, how men lay blame upon us gods for what is after all nothing but their own folly. Look at Aegisthus; he must needs make love to

Text Processing

```
class SampleProcessor < HexaPDF::Content::Processor
  def show_text(str)
    boxes = decode_text_with_positioning(str)
    # do something with the character boxes
  end
end
```

```
end

doc = HexaPDF::Document.open(ARGV.shift)
processor = SampleProcessor.new
doc.pages.each_page do |page|
  puts "Processing page"
  page.process_contents(processor)
end
doc.write('output.pdf')
```

Text Processing Sample

Homer's Odyssey

Provided by The Internet Classics Archive.
See bottom for copyright. Available online at
<http://classics.mit.edu/Homer/odyssey.html>

The Odyssey
By Homer

Translated by Samuel Butler

.....

BOOK I

Tell me, O muse, of that ingenious hero who travelled far and wide
after he had sacked the famous town of Troy. Many cities did he visit,
and many were the nations with whose manners and customs he was acquainted;
moreover he suffered much by sea while trying to save his own life
and bring his men safely home; but do what he might he could not save
his men, for they perished through their own sheer folly in eating
the cattle of the Sun-god Hyperion, so the god prevented them from
even reaching home. Tell me, too, about all these things, O daughter
of Jove, from whatsoever source you may know them.

So now all who escaped death in battle or by shipwreck had got safely
home except Odysseus, and he, though he was longing to return to his
wife and country, was detained by the goddess Calypso, who had got
him into a large cave and wanted to marry him. But as years went by,
there came a time when the gods settled that he should go back to
Ithaca; even then, however, when he was among his own people, his
troubles were not yet over; nevertheless all the gods had now begun
to pity him except Neptune, who still persecuted him without ceasing
and would not let him get home.

Now Neptune had gone off to the Ethiopians, who are at the world's
end, and lie in two halves, the one looking West and the other East.
He had gone there to accept a hecatomb of sheep and oxen, and was
enjoying himself at his festival; but the other gods met in the house
of Olympian Jove, and the sire of gods and men spoke first. At what
moment he was thinking of Aegisthus, who had been killed by Agamemnon's
son Orestes; so he said to the other gods:

"See now, how men lay blame upon us gods for what is after all nothing
but their own folly. Look at Aegisthus; he must needs make love to

So... what's the catch?

Not yet released :-)

But in the next few weeks :-)

Thank you!

Slides available at

<http://talks.gettalong.org/euruko2016/>