

Join Stack Overflow in [Standing Up for Net Neutrality](#)



x Dismiss

Join the Stack Overflow Community

Stack Overflow is a community of 7.4 million programmers, just like you, helping each other. Join them; it only takes a minute:

Sign up

How to convert webpage into PDF by using Python

Ask Question

Get started

28

I was finding solution to print webpage into local file PDF, using Python. one of the good solution is to use Qt, found here, <https://bharatikunal.wordpress.com/2010/01/>.

18

It didn't work at the beginning as I had problem with the installation of PyQt4 because it gave error messages such as 'ImportError: No module named PyQt4.QtCore', and 'ImportError: No module named PyQt4.QtCore'.

It was because PyQt4's not installed properly. I used to have the libraries located at C:\Python27\Lib however it's not for PyQt4.

In fact, it simply needs to download from <http://www.riverbankcomputing.com/software/pyqt/download> (mind the correct Python version you are using), and install it to C:\Python27 (my case). That's it.

Now the scripts runs fine so I want to share it. for more options in using Qprinter, please refer to <http://qt-project.org/doc/qt-4.8/qprinter.html#Orientation-enum>.

python pdf webpage qprinter

share improve this question

add a comment

asked Apr 29 '14 at 8:10



Mark K

1,486 2 13 32

5 Answers

active oldest votes

64

You also can use [pdfkit](#):

```
import pdfkit
pdfkit.from_url('http://google.com', 'out.pdf')
```

share improve this answer

answered May 20 '14 at 13:24



NorthCat

4,525 14 25 37

thanks, NorthCat :) – [Mark K](#) May 22 '14 at 4:52

This is awesome, way easier than messing around with reportlab or using a print drive to convert. Thanks so much. – [Dowlers](#) May 27 '15 at 18:56

@NorthCat can u give another example about converting html tables with pdfkit ? – [Egzona](#) Aug 12 '15 at 14:44

The render is very good and can convert form very well. – [Laurent LAPORTE](#) Nov 9 '15 at 9:56

This should be the accepted answer. – [Blaig23](#) Oct 14 '16 at 20:32

I want to print it out as pdf in downloadable format. – [Hariharan Srinivasan](#) May 9 at 10:19

asked 3 years, 2 months ago

viewed 39,585 times

active 3 months ago

BLOG

[Introducing Channels: Private Q&A for Your Team](#)

Customizable HTML5 Document Viewer & Processing Suite



Download Free Trial

accusoft

Jobs near you

DevOps Engineer at Bay Area Tech Startup

Pinn Redwood City, CA

linux database

Senior Engineer, E-Commerce Integrations

AdRoll San Francisco, CA

REMOTE

python javascript

[More jobs near San Mateo...](#)

Linked

56

[Add text to Existing PDF using Python](#)

2

[CssParser Error, while converting html to pdf ,PISA & Python](#)

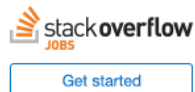
2

[pisa.CreatePDF hangs with huge html file](#)

0

add a comment

More jobs
means more choice



thanks to below posts, and I am able to add on the webpage link address to be printed and present time on the PDF generated, no matter how many pages it has.

14

[Add text to Existing PDF using Python](#)

https://github.com/disflux/django-mtr/blob/master/pdfgen/doc_overlay.py

To share the script as below:

```
import time
from pyPdf import PdfFileWriter, PdfFileReader
import StringIO
from reportlab.pdfgen import canvas
from reportlab.lib.pagesizes import letter
from xhtml2pdf import pisa
import sys
from PyQt4.QtCore import *
from PyQt4.QtGui import *
from PyQt4.QtWebKit import *

url = 'http://www.yahoo.com'
tem_pdf = "c:\\tem_pdf.pdf"
final_file = "c:\\younameit.pdf"

app = QApplication(sys.argv)
web = QWebView()
#Read the URL given
web.load(QUrl(url))
printer = QPrinter()
#setting format
printer.setPageSize(QPrinter.A4)
printer.setOrientation(QPrinter.Landscape)
printer.setOutputFormat(QPrinter.PdfFormat)
#export file as c:\tem_pdf.pdf
printer.setOutputFileName(tem_pdf)

def convertIt():
    web.print_(printer)
    QApplication.exit()

QObject.connect(web, SIGNAL("loadFinished(bool)"), convertIt)

app.exec_()
sys.exit

# Below is to add on the weblink as text and present date&time on PDF generated
```

share improve this answer

edited May 23 at 12:18



Community ♦
1 1

answered Apr 30 '14 at 7:31



Mark K
1,486 2 13 32

Thanks for sharing your code! Any advice for making this work for local pdf files? Or is it as easy as prepending "file:///" to the url? I'm not very familiar with these libraries... thanks – [user2426679](#) Oct 31 '14 at 18:02

@user2426679, you mean convert online PDF into local PDF files? – [Mark K](#) Nov 25 '14 at 1:48

thanks for your reply... sorry for my tardiness. I ended up using wkhtmltopdf since it was able to handle what I was throwing at it. But I was asking how to load a pdf that was local to my hdd. Cheers – [user2426679](#) Dec 28 '14 at 23:15

@user2426679 sorry I still don't get you. maybe because I am a newbie to Python too. You meant read local PDF files in Python? – [Mark K](#) Jan 22 '15 at 8:05

There were some issues with `html5lib`, which is used by `xhtml2pdf`. This solution fixed the problem: github.com/xhtml2pdf/xhtml2pdf/issues/318 – [Blair23](#) Oct 14 '16 at 21:22

add a comment

here is the one working fine:

7

```
import sys
from PyQt4.QtCore import *
from PyQt4.QtGui import *
from PyQt4.QtWebKit import *

app = QApplication(sys.argv)
web = QWebView()
web.load(QUrl("http://www.yahoo.com"))
printer = QPrinter()
printer.setPageSize(QPrinter.A4)
```

Python: Excel to Web to PDF

0

[How to convert my form into PDF by using Python](#)

0

[python cgi html pdf](#)

Related

300

[Python module for converting PDF to text](#)

2395

[How to merge two Python dictionaries in a single expression?](#)

3375

[How do I check whether a file exists using Python?](#)

3743

[What is a metaclass in Python?](#)

6718

[What does the "yield" keyword do in Python?](#)

1181

[Convert HTML + CSS to PDF with PHP?](#)

784

[Converting integer to string in Python?](#)

547

[Merge / convert multiple PDF files into one PDF](#)

1062

[How to convert string to lowercase in Python?](#)

1

[How to convert an HTML document containing an SVG to a PDF file in Python](#)

Hot Network Questions

[Would like <...> to expand to \text{...}](#)

[What is the sci fi story \(novella?\) where a guy is reanimated multiple times?](#)

[Is there an existing data structure that is of fixed size, and will push the oldest/last element out if a new element is inserted?](#)

[Fixed distance buffer around box gives rounded corners in QGIS](#)

[How would humans deal with living in an environment with dinosaurs?](#)

[Foreground-Background Perception of a Logo](#)

[The Original Number](#)

[Help! Where am I?](#)

[Are Global Variables Evil in Arduino?](#)

[Why can "populace" be used as an adjective in "most populace cities"?](#)

[How did the refereeing system of Gösta Mittag-Leffler's Acta Mathematica function from 1882 to \(about\) 1918?](#)

[Was 2016 the Deadliest year for the LGBTQ community?](#)

[The Physiology of people living on a world with 9 year days](#)

[Distribute a number into a list of values as equal as possible whose sum is equal to that number](#)

[Find the Serialized Integer](#)

[Yet Unused Pairs](#)

[Word for Pretentiously Academic but Useless](#)

```
printer.setOutputFormat(QPrinter.PdfFormat)
printer.setOutputFileName("fileOK.pdf")

def convertIt():
    web.print_(printer)
    print "Pdf generated"
    QApplication.exit()

QObject.connect(web, SIGNAL("loadFinished(bool)"), convertIt)
sys.exit(app.exec_())
```

share improve this answer

answered Apr 29 '14 at 8:11



Mark K

1,486 2 13 32

Interestingly, the web page links are generated as text rather than links in the generated PDF. – [amergin](#) Nov 24 '14 at 17:16

Anyone know why this would be generating blank pdfs for me? – [boson](#) Oct 4 '16 at 14:33

[add a comment](#)

WeasyPrint

7

```
pip install weasyprint

python
>>> pdf = weasyprint.HTML('http://www.google.com').write_pdf()
>>> len(pdf)
92059
>>> file('google.pdf', 'w').write(pdf)
```

share improve this answer

answered Dec 23 '15 at 15:04



John Mudd

6,858 2 14 14

[add a comment](#)

Here is a simple solution using QT. I found this as part of an answer to a different question on StackOverflow. I tested it on Windows.

3

```
from PyQt4.QtGui import QTextDocument, QPrinter, QApplication

import sys
app = QApplication(sys.argv)

doc = QTextDocument()
location = "c://python//Jim//html//notes.html"
html = open(location).read()
doc.setHtml(html)

printer = QPrinter()
printer.setOutputFileName("foo.pdf")
printer.setOutputFormat(QPrinter.PdfFormat)
printer.setPageSize(QPrinter.A4);
printer.setPageMargins (15,15,15,15,QPrinter.Millimeter);

doc.print_(printer)
print "done!"
```

share improve this answer

edited Mar 12 '15 at 13:31

answered Jan 20 '15 at 20:38



Jim Paul

91 3

[add a comment](#)

Your Answer

Only one administrator can connect at this time (MS SQL Server error 18461)

Why do professors give talks on published work without their coauthors present at the seminars?

Is there an organization for adults to learn outdoor skills?

Multiply a string and a number!

Is it possible to create food allergies in Minecraft?

Could Wasps and Ants reach the stone age?

What are the features needed in a DSLR for astrophotography?

Sign up or [log in](#)

Post as a guest

Sign up using Google

Sign up using Facebook

Sign up using Email and Password

Name

Email

required, but never shown

Post Your Answer

By posting your answer, you agree to the [privacy policy](#) and [terms of service](#).

Not the answer you're looking for? Browse other questions tagged [python](#) [pdf](#) [webpage](#) [qprinter](#) or [ask your own question](#).

[question feed](#)



Questions

Jobs

Developer Jobs Directory

Documentation

Help

Mobile

**STACK OVERFLOW
BUSINESS**

Talent

Ads

Enterprise

Insights

COMPANY

About

Press

Work Here

Legal

Privacy Policy

Contact Us

**STACK EXCHANGE
NETWORK**

Technology

Life / Arts

Culture / Recreation

Science

Other

[Blog](#) • [Facebook](#) • [Twitter](#) • [LinkedIn](#)

site design / logo ©2017 Slack Exchange Inc;
user contributions licensed under [cc by-sa 3.0](#)
with attribution required. rev 2017.7.12.1275