



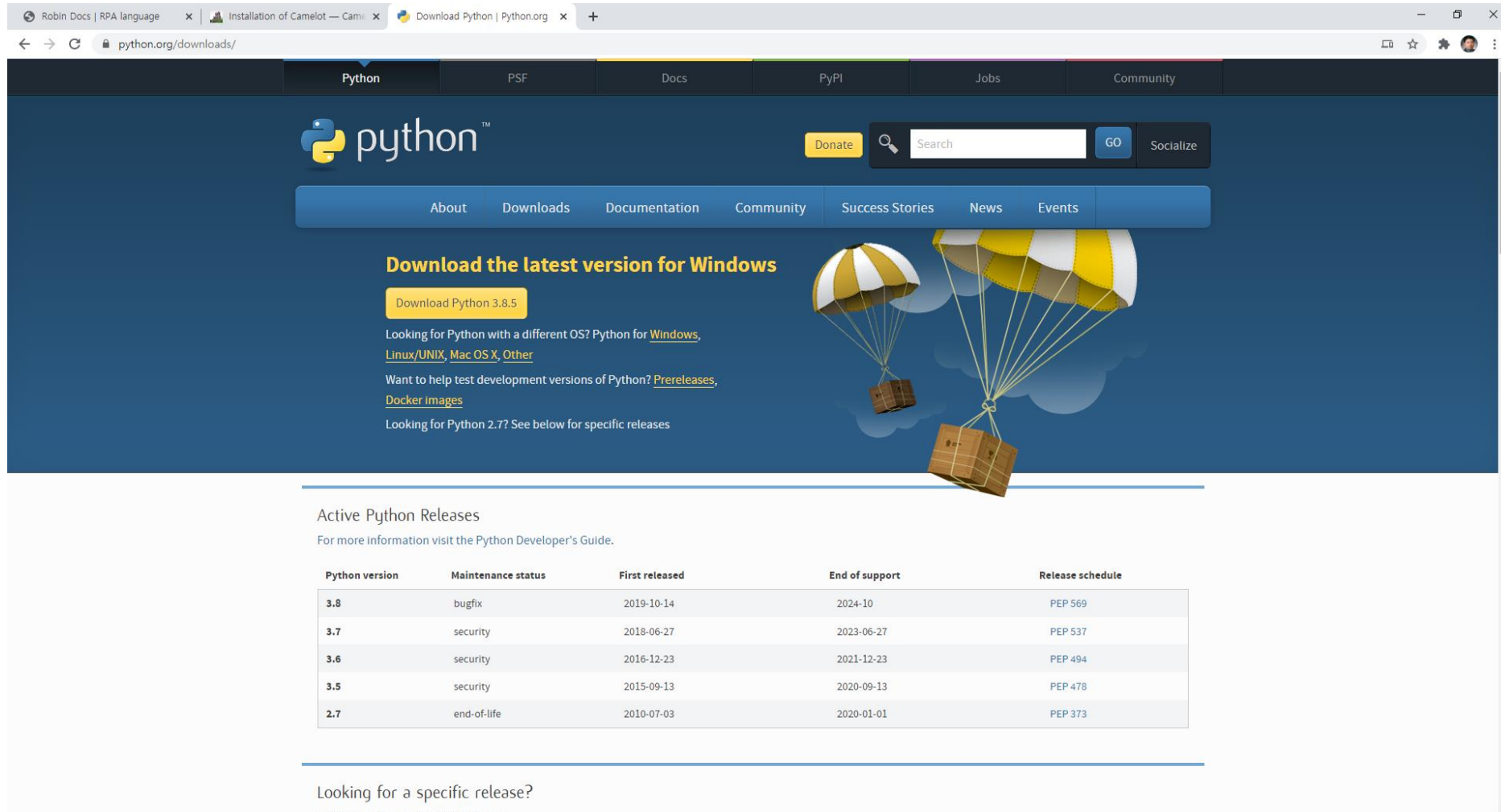
Robotic Process Automation (RPA)

Part 20 – Python을 이용한 PDF 파일 Table 추출

2020. 8.

정 준 수 Ph.D.

Python 설치



The screenshot shows the Python.org website's download page. The browser's address bar displays 'python.org/downloads/'. The page features a dark blue header with the Python logo, a search bar, and navigation links. Below the header, a prominent section titled 'Download the latest version for Windows' includes a 'Download Python 3.8.5' button and links for other operating systems. To the right of this text is an illustration of two parachutes carrying crates. Further down, a section titled 'Active Python Releases' provides a table with details on various Python versions, including their maintenance status, release dates, and end-of-support dates. At the bottom, there is a link to find specific releases.

Download the latest version for Windows

[Download Python 3.8.5](#)

Looking for Python with a different OS? Python for [Windows](#), [Linux/UNIX](#), [Mac OS X](#), [Other](#)

Want to help test development versions of Python? [Prereleases](#), [Docker images](#)

Looking for Python 2.7? See below for specific releases

Active Python Releases

For more information visit the Python Developer's Guide.

Python version	Maintenance status	First released	End of support	Release schedule
3.8	bugfix	2019-10-14	2024-10	PEP 569
3.7	security	2018-06-27	2023-06-27	PEP 537
3.6	security	2016-12-23	2021-12-23	PEP 494
3.5	security	2015-09-13	2020-09-13	PEP 478
2.7	end-of-life	2010-07-03	2020-01-01	PEP 373

Looking for a specific release?

[Python releases by version number](#)

<https://www.python.org/downloads/>

Camelot 설치

Camelot install again using this command.

- > pip install camelot-py
- > pip install camelot-py[cv]
- > pip install camelot-py[all]

Ghostscript install again using this command.

<https://www.ghostscript.com/download/gsdnld.html>

Camelot: PDF 파일내에 있는 Table 가져오기

```
import camelot
```

```
import sys
```

```
tables = camelot.read_pdf(sys.argv[1])
```

```
tables[0].to_excel(sys.argv[2]) # to_json, to_excel, to_html
```

Examples and Codes

<https://github.com/JSJeong-me/SDF-RPA>

정 준 수 / Ph.D. (heinem@naver.com)

- 삼성전자 연구원
- 삼성의료원 (삼성생명과학연구소)
- 삼성SDS (정보기술연구소)
- (사)한국인공지능협회, AI, 머신러닝 전문강의
- 한국소프트웨어산업협회, AI, 머신러닝 SW전문강의
- 한성대학교 교수(겸)

