

GENERATE TABLE OF CONTENTS OF PDF FILE FROM FEDERAL REGISTER

THIS DOCUMENT IS A GUIDE TO INSTALL AND USE THE SOFTWARE (WRITING IN PYTHON) TO DOWNLOAD, CONVERT AND MERGE THE PDF FILES FROM IRS WITH THE ACTUAL FILES OF FEDERAL REGISTERS.

Table of Contents

Contents **Error! Bookmark not defined.**

How To 3

 Install..... 3

 Use 3

Tutorial for general purpose 4

BUGS..... 6

 JSON loads..... 6

 Description 6

 Solution 6

How To

Install

To install the software you must have this Python libraries in the OS you use:

- **JSON**
- **REQUEST**
- **WEASYPRINT**
- **PyPDF2**
- **XML.DOM**
- **ElementTree**
- **URLLIB**

If you want to install this libraries in your system, type “**pip install *name of the package***” where the name of the package is specified upside. Then you have to create a root folder in your system to allocate the package.

Name the folder as what you want but inside it must have be a folder called “**pdfs_irs**” where the PDF of IRS allocated.

Use

To use the software you have to run the Python interpreter in any OS and run the actual **__init__.py** code. Then follow the instructions inside the terminal.

ATTENTION: For good performance of the software ensure you have a PDF file of the IRS archive.

Tutorial for general purpose

A guide step by step

1. Install all dependencies to ensure the correctly function of the software.
2. Run the software within make `__init__.py` code.
3. In the main menu, select one of two options show up:
 - a. **Option 1** (Search by text)
 - b. **Option 2** (Search by year-document number)
4. If you select the **Option 1**, you have to provide a name to search the document in Federal Register Database.
5. If you select the **Option 2**, you have to provide a document number (if you have it) and the year of emission.

You have to see something like that depends of what option you choose:

Option 1

```

----- GET DOCUMENT FROM THE FEDERAL REGISTER -----
(1) Insert the name of the document (loose, because make a search by name)
(2) Insert the document year and serial number (exact copy of the document)

Enter an option: 1
Enter the exact document name to search it: base erosion
[DOCUMENT NUMBER] 2018-27391
[DOCUMENT TITLE] Base Erosion and Anti-Abuse Tax
IS THIS YOUR DOCUMENT? Press [Y/n] ->

```

Option 2

```

----- GET DOCUMENT FROM THE FEDERAL REGISTER -----
(1) Insert the name of the document (loose, because make a search by name)
(2) Insert the document year and serial number (exact copy of the document)

Enter an option: 2
Enter a YEAR of a document (1990-2020) (FOUR DIGITS): 2018
Enter a DOCUMENT serial number: 27391
[DOCUMENT TITLE] Base Erosion and Anti-Abuse Tax
IS THIS YOUR DOCUMENT? Press [Y/n] ->

```

6. Then, you have to PRESS ENTER to continue and provide the PATH of the PDF IRS, according with the DOCUMENT TITLE of the last steps.

Example:

If a document search by Option 1 have the title **BASE EROSION AND ANTIABUSE TAX**, you should have a file (downloaded from IRS) according with that title.

7. The file named **FINAL_PREAM** and **FINAL_PROPOSED** will be saved in the root folder. The same of the `__init__.py`

Bugs and Error

JSON loads

Description

Estrange bug in the **json.loads(str, encoding)** parameter make an exception when BYTES pass by parameter.

Solution

Use `request.json()` instead in `get_document` variables.

Ex: `get_document_json() = get_document.json()`