

Upload to Internet Archive Using API

```
In [2]: from glob import glob
import internetarchive
from internetarchive import upload
from internetarchive import get_item
import re
import os

C:\ProgramData\Anaconda3\lib\site-packages\requests\__init__.py:89: RequestsDependencyWarning: urllib3 (1.26.6)
or chardet (3.0.4) doesn't match a supported version!
  warnings.warn("urllib3 ({}), or chardet ({}), doesn't match a supported "
```

```
In [5]: # configure internet archive on cmd
        # ia configure
```

```
In [3]: # Titles indicating the range of names included in the catalogues

titles = ['A to Amer', 'Ames to Armh', 'Armi to Bal', 'Bam to Bedh', 'Bedi to Bem', 'Ben to Blac', 'Blad to Q.',
          'R. Boyd to I. Brown', 'J. Brown to L. Butler', 'M. Butler to Caq', 'Car to Chaq', 'Char to Clav', 'I.',
          'Con to Conk', 'Conl to Crd', 'Cre to I. Davis', 'J. Davis to Dek', 'Del to Dov', 'Dow to Durg', 'Dun',
          'Elt to Fie', 'Fif to Frn', 'Fro to Geq', 'Ger to Gorl', 'Gorm to Grd', 'Gre to Grh', 'Gri to Hand',
          'Harw to Hem', 'Hen to Holc', 'Hold to Huba', 'Hubb to I. Jackson', 'J. Jackson to I. Johnson', 'J. J.',
          'K to Kira', 'Kirb to Lano', 'Lanp to Ler', 'Les to Lowq', 'Lowr to Mack', 'Macl to Manh', 'Mani to Man',
          'Meth to Mh', 'Mi to Mord', 'More to Mur', 'Mus to Nn', 'No to Oz', 'P to Peo', 'Pep to Pn', 'Po to P.',
          'Ri to Rooj', 'Rook to Sam', 'San to Sem', 'Sen to Simo', 'Simp to Wh. Smith', 'Wi. Smith to So', 'Sp',
          'Ste to Sv', 'Sw to I. Thomas', 'J. Thomas to Tom', 'Ton to Tz', 'U', 'V to Ward', 'Ware to H. White',
          'I. White to R. Williams', 'S. Williams to Wood', 'T. Wood to Wz', 'X to Z']

# PATH TO ALL PDF FILES
pdf_list = [y for x in os.walk(r'C:\Users\heidi\Documents\main_file_pdf_redone') for y in glob(os.path.join(x[0],
                                                             '*.*'))]

# Drawer Numbers for each file
# drawer_no = [re.findall("\d{3}",i)[0] for i in pdf_list]

drawer_no = ['157', '158', '159', '160', '161', '162', '163', '164', '165', '166', '167', '169', '170', '171',
             '209', '210', '211', '212', '213', '214', '215', '217', '218', '219', '220', '221', '222', '223', '224', '225']

# identifier list to be used for uploading each item
identifier_list = ["rubensteinmanuscriptcatalog_" + titles[i].replace(" ", "_") for i in range(len(titles))]
```

```
In [8]: print(drawer_no)

['157', '158', '159', '160', '161', '162', '163', '164', '165', '166', '167', '169', '170', '171', '172', '173', '174', '175', '176', '177', '178', '179', '181', '182', '183', '184', '185', '186', '187', '188', '189', '190', '191', '193', '194', '195', '196', '197', '198', '199', '200', '201', '202', '203', '205', '206', '207', '208', '209', '210', '211', '212', '213', '214', '215', '217', '218', '219', '220', '221', '222', '223', '224', '225', '226', '227', '229', '230', '231']
```

```
In [4]: def upload_to_ia(title_no):
        # Create metadata dictionary
        i=title_no
        md ={'title': "Rubenstein Library Card Catalog Drawer {} {}".format(drawer_no[i],titles[i]),
              'mediatype': 'texts',
              'collection': "opensource",
              'description':
              """This is drawer number {} from the Duke University David M. Rubenstein Rare Book & Manuscript Library.
              This collection contains digitized cards from the former physical catalog and has the authors with
              'subject': 'manuscripts'}

        # create identifier
        idf = "rubensteinmanuscriptcatalog_" + titles[i].replace(" ", "_")

        # If item already exists, print and return
        exists = get_item(idf).exists
        if exists:
            print("{} already exists".format(idf))
            return

        # Uploading item with identifier
        r = upload(idf, files=pdf_list[i], metadata=md,verbose=True,validate_identifier=True)

        # Check for succesful upload
        success = r[0].status_code==200
        print("Drawer no {}".format(drawer_no[i]))
        if success:
            print("---Uploaded Succesfully---")
        else:
            print(r[0].status_code)
```

```
In [ ]: # UPLOAD ALL THE FILES TO IA
        # for i in range(len(titles)):
        #     upload_to_ia(i)
```

```
In [5]: # Change titles to indicate correct alphabetical sequence,
# for e.g from R. Boyd to Boyd, R

titles_2 = ['A to Amer', 'Ames to Armh', 'Armi to Bal', 'Bam to Bedh', 'Bedi to Bem', 'Ben to Blac', 'Blad to Bo',
'Boyd, R to Brown, I', 'Brown, J to Butler, L', 'Butler, M to Caq', 'Car to Chaq', 'Char to Clav', 'C',
'Con to Conk', 'Conl to Crd', 'Cre to Davis, I', 'Davis, J to Dek', 'Del to Dov', 'Dow to Durg', 'Du',
'Elt to Fie', 'Fif to Frn', 'Fro to Geg', 'Ger to Gorl', 'Gorm to Grd', 'Gre to Grh', 'Gri to Hand',
'Hane to Harv', 'Harw to Hem', 'Hen to Holc', 'Hold to Huba', 'Hubb to Jackson, I', 'Jackson, J to J',
'Johnson, J to Jz', 'K to Kira', 'Kirb to Lano', 'Lanp to Ler', 'Les to Lowq', 'Lowr to Mack', 'Macl',
'Mani to Maw', 'Max to Metg', 'Meth to Mh', 'Mi to Mord', 'More to Mur', 'Mus to Nn', 'No to Oz',
'P to Peo', 'Pep to Pn', 'Po to Puk', 'Pul to Rh', 'Ri to Rooj', 'Rook to Sam', 'San to Sem',
'Sen to Simo', 'Simp to Smith, Wh', 'Smith, Wi to So', 'Sp to Std', 'Ste to Sv', 'Sw to Thomas, I',
'Thomas, J to Tom', 'Ton to Tz', 'U', 'V to Ward', 'Ware to White, H', 'White, I to Williams, R',
'Williams, S to Wood', 'Wood, T to Wz', 'X to Z']
```

```
In [ ]: # Update titles of each card into new format with the new titles
# for i in range(len(identifier_list)):
#     item = get_item(identifier_list[i])
#     md = {"title": f"{titles_2[i]}: Drawer {drawer_no[i]}, Rubenstein Library Card Catalog",
#          'description':
#              """This is drawer number {} from the Duke University David M. Rubenstein Rare Book & Manuscript
#              This collection contains digitized cards from the former physical catalog and has the authors w
#     item.modify_metadata(md)
```

```
In [11]: ## Upload new edited version of pdfs
# update = [drawer_no.index(str(i)) for i in [185,196,209,211,212,213,214,215,230]]
# for i in update:
#     md={'title': "Rubenstein Library Card Catalog Drawer {} {}".format(drawer_no[i],titles[i]),
#         'mediatype': 'texts',
#         'collection': "opensource",
#         'description':
#         """This is drawer number {} from the Duke University David M. Rubenstein Rare Book & Manuscript
#         This collection contains digitized cards from the former physical catalog and has the authors w
#         'subject': 'manuscripts'}
#     idf = "rubensteinmanuscriptcatalog_" + titles[i].replace(" ", "_")

#     r = upload(idf, files=pdf_list[i], metadata=md,verbose=True,validate_identifier=True)

#     # Check for succesful upload
#     success = r[0].status_code==200
#     print("Drawer no {}".format(drawer_no[i]))
#     if success:
#         print("---Uploaded Succesfully---")
#     else:
#         print(r[0].status_code)
```

```
uploading main_file_185_gri_hand.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 148/148 [00:05<00:00, 27.06MiB/s]
Drawer no 185
---Uploaded Succesfully---

uploading main_file_196_lanp_ler.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 133/133 [00:11<00:00, 11.89MiB/s]
Drawer no 196
---Uploaded Succesfully---

uploading main_file_209_pep_pn.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 85/85 [00:04<00:00, 17.02MiB/s]
Drawer no 209
---Uploaded Succesfully---

uploading main_file_211_pul_rh.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 89/89 [00:22<00:00, 3.95MiB/s]
Drawer no 211
---Uploaded Succesfully---

uploading main_file_212_ri_rooj.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 166/166 [00:30<00:00, 5.44MiB/s]
Drawer no 212
---Uploaded Succesfully---

uploading main_file_213_rook_sam.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 153/153 [00:06<00:00, 23.08MiB/s]
Drawer no 213
---Uploaded Succesfully---

uploading main_file_214_san_sem.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 158/158 [00:06<00:00, 23.89MiB/s]
Drawer no 214
---Uploaded Succesfully---

uploading main_file_215_sen_simo.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 121/121 [00:09<00:00, 12.83MiB/s]
Drawer no 215
---Uploaded Succesfully---

uploading main_file_230_wood_t_wz.pdf: 100%|██████████████████████████████████████████████████████████████████████████████| 89/89 [00:06<00:00, 12.95MiB/s]
Drawer no 230
---Uploaded Succesfully---
```