ndaa_book

January 13, 2021

1 FY2020 Quick Search Analysis

1.1 GC Document Processing to JSON required before hand

```
[74]: import json
      import pandas as pd
      with open('NDAA_FY2020.json') as f:
        data = json.load(f)
```

1.2 Defining Keywords for analysis

Current keywords: transform, reform

Choosing stem words to catch the largest amount of variance

```
[75]: keywords = ['transform', 'reform']
      key_dict = {}
```

```
[76]: for par in data['paragraphs']:
          for key in keywords:
              if key in par['par_raw_text_t']:
                  if key in key_dict:
                      key_dict[key].append(par)
                  else:
                      key_dict[key] = [par]
```

```
[77]: all_df = []
      out_dict = []
      for keyword in keywords:
          for item in key_dict[keyword]:
              entity = []
              for ent in item['entities']:
                  for ent_s in item['entities'][ent]:
                      entity.append(ent_s)
              out_dict.append({
                  'keyword': keyword,
                  'filename':item['filename'],
                  'page_num':item['page_num_i'],
```

```
                'par_raw_text_t':item['par_raw_text_t'],
                'entities':entity,
            })
    df = pd.DataFrame(out_dict)
    df
```

[77]:
```
        keyword          filename  page_num  \
    0   transform  NDAA_FY2020.pdf       110
    1   transform  NDAA_FY2020.pdf       353
    2   transform  NDAA_FY2020.pdf       551
    3   transform  NDAA_FY2020.pdf       657
    4   transform  NDAA_FY2020.pdf      1065
    5      reform  NDAA_FY2020.pdf         2
    6      reform  NDAA_FY2020.pdf        10
    7      reform  NDAA_FY2020.pdf        11
    8      reform  NDAA_FY2020.pdf        23
    9      reform  NDAA_FY2020.pdf        66
    10     reform  NDAA_FY2020.pdf        96
    11     reform  NDAA_FY2020.pdf       334
    12     reform  NDAA_FY2020.pdf       340
    13     reform  NDAA_FY2020.pdf       341
    14     reform  NDAA_FY2020.pdf       359
    15     reform  NDAA_FY2020.pdf       373
    16     reform  NDAA_FY2020.pdf       412
    17     reform  NDAA_FY2020.pdf       413
    18     reform  NDAA_FY2020.pdf       414
    19     reform  NDAA_FY2020.pdf       543
    20     reform  NDAA_FY2020.pdf       551
    21     reform  NDAA_FY2020.pdf       553
    22     reform  NDAA_FY2020.pdf       568
    23     reform  NDAA_FY2020.pdf       580
    24     reform  NDAA_FY2020.pdf       581
    25     reform  NDAA_FY2020.pdf       587
    26     reform  NDAA_FY2020.pdf       589
    27     reform  NDAA_FY2020.pdf       928
    28     reform  NDAA_FY2020.pdf      1112

                                       par_raw_text_t  \
    0   H . R . 6395-111 ( d )AGENCY PARTICIPATION.-Th…
    1   H . R . 6395-354 and applicability of such tec…
    2   H . R . 6395-552 ' ' ( vi ) transformation of …
    3   H . R . 6395-658 ( B ) notify the appropriate …
    4   H . R . 6395-1066 SEC . 4201 .RESEARCH , DEVEL…
    5   H . R . 6395-3 Sec . 212 .Disclosure requireme…
    6   H . R . 6395-11 Sec . 757 .Study on force mix …
    7   H . R . 6395-12 Sec . 838 .Comptroller General…
    8   H . R . 6395-24 Sec . 2503 .Execution of proje…
```

```
9    H . R . 6395-67 Sec . 212 .Disclosure requirem…
10   H . R . 6395-97 Defense as the Secretary consi…
11   H . R . 6395-335 ( 4 )The effect of any such l…
12   H . R . 6395-341 ( 3 )The increase , as compar…
13   H . R . 6395-342 Sec . 832 .Extension of pilot…
14   H . R . 6395-360 ( 5 )SECRETARY CONCERNED.-The…
15   H . R . 6395-374 SEC . 838 .COMPTROLLER GENERA…
16   H . R . 6395-413 an analysis of the process us…
17   H . R . 6395-414 of the Department of Defense …
18   H . R . 6395-415 General of the United States …
19   H . R . 6395-544 ' ' ( 1 ) A staffing plan to …
20   H . R . 6395-552 ' ' ( vi ) transformation of …
21   H . R . 6395-554 ( D ) an assessment of the ex…
22   H . R . 6395-569 ( 6 ) affirms the commitment …
23   H . R . 6395-581 ( C )1 member selected by agr…
24   H . R . 6395-582 intelligence services and str…
25   H . R . 6395-588 of title 10 , United States C…
26   H . R . 6395-590 involved in the illicit trade…
27   H . R . 6395-929 Subtitle B-Military Family Ho…
28   H . R . 6395-1113 SEC . 4501 .OTHER AUTHORIZAT…


                                         entities
0    [the Food and Drug Administration, the Nationa…
1    [the Chief Information Office, the Defense Adv…
2    [State, the Department of Defense, SEC, C the …
3    [National Security Space Launch, the Space For…
4    [T&E, RAND ARROYO CENTER, MLRS, CHINOOK, SEC, …
5    [the Joint Artificial Intelligence Center, the…
6    [Department of Defense, Department of Veterans…
7    [Congress, the Small Business Technology Trans…
8    [Department of Defense, Department of Defense …
9    [the Joint Artificial Intelligence Center, the…
10   [the Joint Artificial Intelligence Center, D Q…
11   [the Federal Government, nent, United States C…
12   [Department of Defense, the Joint Artificial I…
13   [Department of Defense, the Small Business Tec…
14   [the Department of Defense, SEC, United States…
15   [the Defense Science Board Task Force, the Dep…
16   [Congress, Subtitle BOther Department of Defen…
17   [Congress, the Department of Defense, SUBMITTA…
18   [Department of Defense, the Department of Defe…
19   [Congress, Hizballah, Islamic Revolutionary Gu…
20   [State, the Department of Defense, SEC, C the …
21   [Congress, State, air force, the Committee on …
22   [State, Congress, the Department of Defense, S…
23   [TRANSITIONAL PERIOD.The, the General Intellig…
24   [SEC, the Child Soldiers Prevention Act, Congr…
```

```
25  [Department of Defense, United Nations, SEC, t…
26  [State, SEC, Fed, FORM.The, Cabinet, the Sover…
27  [Department of Defense, Guam Realignment, Neva…
28  [PROC REPLACEMENT & MODERNIZATION, OFFICE OF T…
```

1.3 Writing simple table to file

Output File is in attached CSV at NDAA_FY2020_QuickSearch.csv

```python
[78]: df.to_csv('NDAA_FY2020_QuickSearch.csv')
```

1.4 Further Analysis

```
[ ]:
```

```
[ ]:
```