Compare Standards Lists

This script reads the current UFGS and JES spec sections for comparison

Jupyter Notebook written by Ben Fisher on 2 December 2024
benjamin.s.fisher@usace.army.mil

Imports

The following imports are assumed to have been previously installed (for Notebook installs, use *! pip install ~*)

In [1]:
```
import os, datetime, warnings
from pathlib import Path
import pandas as pd
import numpy as np
import bs4 as bs
```

Directories

Working directories are made relative to the 'current working directory,' which is where the Notebook (.ipynb) file is located.

In [2]:
```
parent_folder = os.getcwd()

ufgs_masters = parent_folder + '\\UFGS Cleaned\\'
jes_masters = parent_folder + '\\JES Cleaned\\'
```

In [3]:
```
warnings.filterwarnings('ignore')
```

Define Helper Functions

In [4]:
```
def get_titles(folder):
    titles = {}
    for file in os.listdir(folder):
        file_path = folder + file
        if Path(file).suffix.lower() == '.sec':
            try:
                with open(file_path, 'r') as doc:
                    soup = bs.BeautifulSoup(doc.read(), 'lxml')
                    section_number = Path(file).stem
                    title = soup.find('stl').text.title()
                    titles.update({section_number: title})
            except:
                titles.update({Path(file).stem})
    return titles
```

In [5]:
```
def get_list(folder):
    new_list = []
```

```
    for file in os.listdir(folder):
        file_path = folder + file
        if Path(file).suffix.lower() == '.sec':
            new_list.append(Path(file).stem)
    return new_list
```

In [6]:
```
def get_full_list(list_a, list_b):
    full_list = list(set(list_a + list_b))
    full_list.sort()
    return full_list
```

In [7]:
```
def compare_list(folder_a, folder_b):
    list_a = get_list(folder_a)
    list_b = get_list(folder_b)
    full_list = get_full_list(list_a, list_b)
    compare = []
    for element in full_list:
        compare.append([element,'', "•" if element in list_a else '', "•" if element i

    titles_a = get_titles(folder_a)
    titles_b = get_titles(folder_b)

    titles_b.update(titles_a)

    for i in range(len(compare)):
        standard = compare[i][0]
        title = titles_b[standard]
        compare[i][1] = title

    return compare
```

In [8]:
```
def get_df(a_list):
    if a_list:
        df = pd.DataFrame(a_list)

        df.rename(columns={0:'Section', 1:'Title', 2:'UFGS', 3:'JES'}, inplace=True)
        df.sort_values(by=['Section'], inplace=True)
        df.index = np.arange(1, len(df) + 1)

        report_name = parent_folder + '\\Section Comparison ' + '{:%Y%m%d %H%M%S}'.for
        df.to_excel(report_name)
    return df
```

Compare Standards

In [9]:
```
comparison = get_df(compare_list(ufgs_masters, jes_masters))

comparison
```

Out[9]:

	Section	Title	UFGS	JES
1	00 01 15	List Of Drawings	•	
2	01 10 00	Description Of Work		•
3	01 11 00	Summary Of Work	•	•
4	01 11 00.00 10	General Contract Requirements		•
5	01 11 30.00 25	Diving		•
...	...	...	...	...
699	48 06 15	Turbine Oil	•	
700	48 14 00	Solar Photovoltaic Systems	•	
701	48 14 13.00 20	Solar Liquid Flat Plate And Evacuated Tube Col...	•	
702	48 15 00	Wind Generator System	•	
703	48 16 00	Landfill Gas Systems	•	

703 rows × 4 columns

In [ ]: