script_codici_green

June 22, 2022

1 Make a full list of Green codes

Autori: Fabio Morea, Leyla Vesnic, Andrea Bincoletto Data: 20 giugno 2022 Versione 1.0 GitHub: https://github.com/areasciencepark/CodiciGreen

1.1 load libraries

```python
import pandas as pd
import re
```

1.2 Load data

```python
# reading data and selecting only the column we are interested in
raw_data = pd.read_excel("../2IPCGreenInventoryList_2022WIPO.xlsx", dtype=str)
raw_data = raw_data[[ "IPC"]]
raw_data.dropna(inplace=True)
raw_data.head(5)
```

```python
codes = pd.read_excel("../IPC WIPO unione.xlsx", dtype=str)
codes = list(codes["ipc_class_symbol"])
```

1.3 scan eaach row of the dataframe to generate a list of results

"," is the separator for lists of codes "-" is the separator for sequences and add individual results in a codes_list[]

```python
codes_list = []
i=0

for index, row in raw_data.iterrows():
    parts = row['IPC'].split(" ")

    if len(parts) == 1:
        print(i,j, "this is a short code ", parts[0])
        i+=1
        codes_list.append(parts[0])
    else:
        first_part  = parts[0]
```

```python
        second_part = row['IPC'].split(" ")[1]
        other_parts = second_part.split(",")
        for j in range(0,len(other_parts)):
            this_part = re.sub("\s","",other_parts[j]) #remove whitespaces
            full_code = first_part + this_part
            if not("-" in other_parts[j]):
                i+=1
                codes_list.append(full_code)
                print(i,j,"Adding a single code ", full_code)

            else:#sequenza
                sequence_start = first_part + this_part.split("-")[0]
                sequence_end   = first_part + this_part.split("-")[1]
                print(i,j, "** This is a sequence from ", sequence_start, " to␣
 ↪" , sequence_end)
                # scan the full list of codes and add everything from␣
 ↪sequence_start to sequence_end
                get_also_next_code = False
                for cod in codes:
                    if (cod == sequence_start):

                        print("this is the sequence start",cod)
                        i+=1
                        codes_list.append(cod)
                        get_also_next_code = True
                    elif (cod == sequence_end):
                        print("this is the sequence end",cod)
                        i+=1
                        codes_list.append(cod)
                        get_also_next_code = False
                        break #don't need to search in the rest of the list
                    elif (get_also_next_code==True):
                        print("in the middle of the sequence",cod)
                        i+=1
                        codes_list.append(cod)
print("Done :-D")

codes_list = sorted(list(set(codes_list)))
print(len(codes_list), " codes found")
```

```python
[ ]: print(codes_list)
```

```
[ ]: codes_list
```

2 write the results to a CSV file

```
[ ]: pd.DataFrame(codes_list).to_csv("GreenCodes.csv", index=False, header=False)
```