

# Guide: SDG Keyword Scanner

Author: Thomas Adams: [t.adams4@universityofgalway.ie](mailto:t.adams4@universityofgalway.ie)

Co-Author: Prof. Jamie Goggins: [Jamie.goggins@universityofgalway.ie](mailto:Jamie.goggins@universityofgalway.ie)

Code development: Dr. Muslim Jameel Syed: [muslimjameel.syed@universityofgalway.ie](mailto:muslimjameel.syed@universityofgalway.ie)

Date: June/2023

## Introduction:

This document explains how carry out a keyword text scan using Python code. There are two Excel spreadsheets to accommodate this pdf which should be available for download alongside this doc. The Python code can be found in the appendix of this document. The purpose of the tool is to scan text for keywords related to the UNs Sustainable Development Goals (SDGs) (UN, 2015). The full methodology behind the creation of the SDG keyword list and the tool can be read in the published research article (Adams and Goggins, 2023). This tool has been used to carry out a cross-institutional study of Civil Engineering programmes in Ireland, where learning outcomes for the Civil Engineering bachelor's from 7 different institutions were analysed for SDG coverage. The purpose of the tool is to provide a free, easy to use method for institutions to measure these "SDG baselines". Measuring SDG baselines is very useful for educational institutions to understand what topics of sustainability they currently cover, and what opportunities might lie before them to increase the coverage of sustainability topics.

## How to use the tool:

### Step 1 – Google Collab

Start a new project on [Google Colab](https://colab.research.google.com/), or your preferred programme for running Python script. Copy the code in the appendix of this document and paste it into your new project.

### Step 2 – SDG Keyword Database

Download the premade file titled "SDG Keyword Database" containing the keyword list described in the research paper (Adams and Goggins, 2023). Upload this file to your Google Collab project.

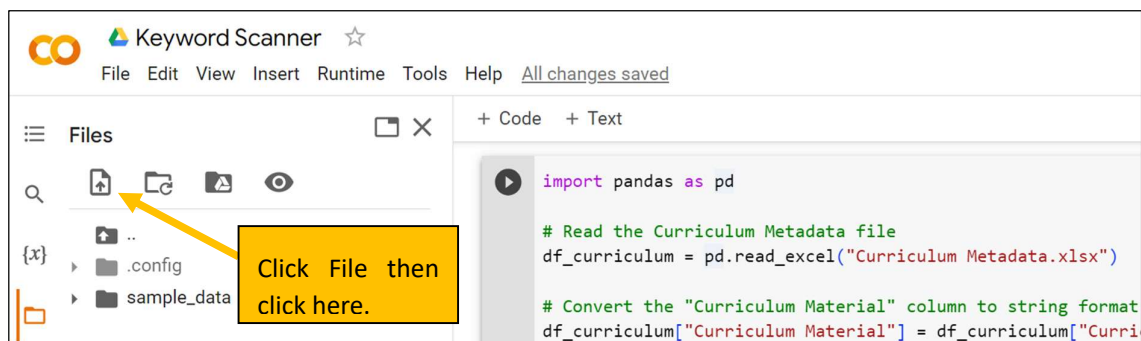


Figure 1 - How to upload the SDG Keyword List and Curriculum Material to a Google Collab project

### Step 3 – Curriculum Metadata

Download the premade file titled “Curriculum Metadata”. This file simply contains column titles as shown in Table 2 below. These headings are just suggestions and do not all need to be filled out. The tool only scans in the “Curriculum Material” column. The other columns are just useful for analysing the results and comparisons etc. The Curriculum Metadata column is highlighted in yellow to emphasise that this is where to paste the text to be scanned.

Table 1 – Column titles in the Curriculum Metadata file

Curriculum Material	Title	Code	Credits	Year	Programme	School	College	Grad Level
---------------------	-------	------	---------	------	-----------	--------	---------	------------

Any type of curriculum material can be pasted here, the tool is set-up simply to take the words from the “Keywords” column in the “SDG Keyword Database” file and scan for them in the text in the “Curriculum Material” column. More columns can be added or changed if required. This needs to be done both in the “Curriculum Metadata” file and in the code. There is a line in the code:

```
requiredfields = ['Curriculum Material', 'Title', 'Code',  
'Credits', 'Year', 'Programme', 'School', 'College', 'Grad Level',]
```

This line defines the column titles in the “Curriculum Metadata” file. Simply match these with the columns in the actual file if you wish to make changes.

### Step 4 – Results

Once the code and the two Excel files are in the Collab notebook, run the code by clicking on the arrow shown in Figure 2. Shortly after this, your results should populate into a downloadable csv file titled “SDG Scan Results”. The “SDG Scan Results” file will be a copy of your “Curriculum Metadata” file with three extra columns added: 1. Keyword, 2. SDG, and 3. Relevancy Label.

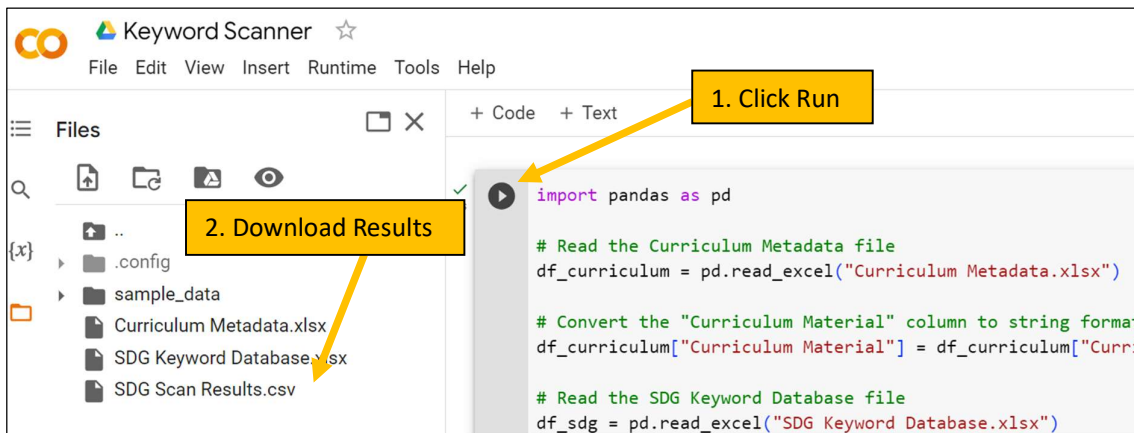


Figure 2 – How to run the code and download your results

**Note: If you run into errors, it can be really useful to copy and paste the error straight into the ChatGPT chat box, it's amazing at solving coding problems!**

## References

Adams, T., Goggins, J., 2023. Education for Sustainable Development: Mapping the SDGs to University Curricula. Sustainability 15, 8340. <https://doi.org/10.3390/su15108340>

UN, 2015. About the Sustainable Development Goals - United Nations Sustainable Development [WWW Document]. Sustain. Dev. Goals. URL <https://www.un.org/sustainabledevelopment/sustainable-development-goals/> (accessed 1.4.20).

## Appendix

```
import pandas as pd

# Read the Curriculum Metadata file
df_curriculum = pd.read_excel("Curriculum Metadata.xlsx")

# Convert the "Curriculum Material" column to string format
df_curriculum["Curriculum Material"] = df_curriculum["Curriculum
Material"].astype(str)

# Read the SDG Keyword Database file
df_sdg = pd.read_excel("SDG Keyword Database.xlsx")

# Convert the "Keywords" column to string format
df_sdg["Keywords"] = df_sdg["Keywords"].astype(str)

def preprocess(df):
    records = df[['Keywords', 'SDG', 'Relevancy
Label']].to_records(index=False).tolist()
    return [(str(word), str(sdg), str(label)) for word, sdg, label in
records]

def replace_nan_in_dataframe(df):
    df = df.fillna("")
    return df

def find_words_matching(df2, records):
    requiredfields = ['Curriculum Material', 'Title', 'Code',
'Credits', 'Year', 'Programme', 'School', 'College', 'Grad Level', ]
    alldictionaries = []
    for i in records:
        print("Searching for word '{}' ...".format(i[0]))
        status = df2['Curriculum Material'].apply(lambda x:
i[0].lower() in x.lower())
        if True in list(status):
            rows = df2.loc[status]
            for index, row in rows.iterrows():
                dictrecord = row[requiredfields].to_dict()
```

```

        dictrecord['keyword'] = i[0]
        dictrecord['Curriculum Material'] = row['Curriculum
Material']

        dictrecord['SDG'] = i[1]
        dictrecord['Relevancy Label'] = i[2]
        alldictionaries.append(dictrecord)
    print("Added record(s)")
    else:
        print("No match found for this word!")
    resultant = pd.DataFrame(alldictionaries)
    return resultant

if __name__=="__main__":
    df = pd.read_excel("SDG Keyword Database.xlsx")
    df2 = pd.read_excel("Curriculum Metadata.xlsx")
    df2 = replace_nan_in_dataframe(df2)
    words = preprocess(df)

    resultantdf = find_words_matching(df2,words)
    resultantdf.to_csv("SDG Scan Results.csv")

```

Curriculum Material	Title	Code	Credits	Year	Programme	School	College	Grad Level
---------------------	-------	------	---------	------	-----------	--------	---------	------------