```
In [1]: %pip install pandas
        %pip install spacy
        %pip install difflib
        %pip install numpy
        %pip install sklearn
```

Requirement already satisfied: pandas in c:\users\user\appdata\local\package
s\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packag
es\python310\site-packages (2.2.1)
Requirement already satisfied: numpy<2,>=1.22.4 in c:\users\user\appdata\loc
al\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\lo
cal-packages\python310\site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\user\appda
ta\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localca
che\local-packages\python310\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\user\appdata\local\p
ackages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-
packages\python310\site-packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\user\appdata\local
\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\loca
l-packages\python310\site-packages (from pandas) (2024.1)
Requirement already satisfied: six>=1.5 in c:\users\user\appdata\local\packa
ges\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-pack
ages\python310\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: C:\Users\user\AppData\Local\Microsoft\WindowsApps\P
ythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\python.exe -m pip install
--upgrade pip

Requirement already satisfied: spacy in c:\users\user\appdata\local\packages
\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-package
s\python310\site-packages (3.8.4)Note: you may need to restart the kernel to
use updated packages.

[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: C:\Users\user\AppData\Local\Microsoft\WindowsApps\P
ythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\python.exe -m pip install
--upgrade pip
```

Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in c:\users\user
\appdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\l
ocalcache\local-packages\python310\site-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\user
\appdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\l
ocalcache\local-packages\python310\site-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\user\ap
pdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\loca
lcache\local-packages\python310\site-packages (from spacy) (1.0.12)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\user\appdata
\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcach
e\local-packages\python310\site-packages (from spacy) (2.0.11)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\user\appdat
a\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcac
he\local-packages\python310\site-packages (from spacy) (3.0.9)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in c:\users\user\appdata
\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcach
e\local-packages\python310\site-packages (from spacy) (8.3.4)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in c:\users\user\appdata
\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcach
e\local-packages\python310\site-packages (from spacy) (1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in c:\users\user\appdata
\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcach
e\local-packages\python310\site-packages (from spacy) (2.5.1)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\user\appd
ata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localc
ache\local-packages\python310\site-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in c:\users\user\appdata
\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcach
e\local-packages\python310\site-packages (from spacy) (0.4.1)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in c:\users\user\appdata
\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcach
e\local-packages\python310\site-packages (from spacy) (0.15.2)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\user\appdata
\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcach
e\local-packages\python310\site-packages (from spacy) (4.67.1)
Requirement already satisfied: numpy>=1.19.0 in c:\users\user\appdata\local
\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\loca
l-packages\python310\site-packages (from spacy) (1.26.4)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\user\appd
ata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localc
ache\local-packages\python310\site-packages (from spacy) (2.32.3)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in c:\us
ers\user\appdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2k
fra8p0\localcache\local-packages\python310\site-packages (from spacy) (2.10.
6)
Requirement already satisfied: jinja2 in c:\users\user\appdata\local\package
s\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packag
es\python310\site-packages (from spacy) (3.1.2)
Requirement already satisfied: setuptools in c:\program files\windowsapps\py
thonsoftwarefoundation.python.3.10_3.10.3056.0_x64__qbz5n2kfra8p0\lib\site-p
ackages (from spacy) (65.5.0)
Requirement already satisfied: packaging>=20.0 in c:\users\user\appdata\loca
l\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\loc
al-packages\python310\site-packages (from spacy) (24.0)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in c:\users\user\appd

ata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localc
ache\local-packages\python310\site-packages (from spacy) (3.5.0)
Requirement already satisfied: language-data>=1.2 in c:\users\user\appdata\l
ocal\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache
\local-packages\python310\site-packages (from langcodes<4.0.0,>=3.2.0->spac
y) (1.3.0)
Requirement already satisfied: annotated-types>=0.6.0 in c:\users\user\appda
ta\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localca
che\local-packages\python310\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.
0,>=1.7.4->spacy) (0.7.0)
Requirement already satisfied: pydantic-core==2.27.2 in c:\users\user\appdat
a\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcac
he\local-packages\python310\site-packages (from pydantic!=1.8,!=1.8.1,<3.0.
0,>=1.7.4->spacy) (2.27.2)
Requirement already satisfied: typing-extensions>=4.12.2 in c:\users\user\ap
pdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\loca
lcache\local-packages\python310\site-packages (from pydantic!=1.8,!=1.8.1,<
3.0.0,>=1.7.4->spacy) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\user\app
data\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\local
cache\local-packages\python310\site-packages (from requests<3.0.0,>=2.13.0->
spacy) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in c:\users\user\appdata\local\p
ackages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-
packages\python310\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.1
0)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\user\appdata\l
ocal\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache
\local-packages\python310\site-packages (from requests<3.0.0,>=2.13.0->spac
y) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\user\appdata\l
ocal\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache
\local-packages\python310\site-packages (from requests<3.0.0,>=2.13.0->spac
y) (2024.12.14)
Requirement already satisfied: blis<1.3.0,>=1.2.0 in c:\users\user\appdata\l
ocal\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache
\local-packages\python310\site-packages (from thinc<8.4.0,>=8.3.4->spacy)
(1.2.0)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\user\app
data\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\local
cache\local-packages\python310\site-packages (from thinc<8.4.0,>=8.3.4->spac
y) (0.1.5)
Requirement already satisfied: colorama in c:\users\user\appdata\local\packa
ges\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-pack
ages\python310\site-packages (from tqdm<5.0.0,>=4.38.0->spacy) (0.4.6)
Requirement already satisfied: click>=8.0.0 in c:\users\user\appdata\local\p
ackages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-
packages\python310\site-packages (from typer<1.0.0,>=0.3.0->spacy) (8.1.8)
Requirement already satisfied: shellingham>=1.3.0 in c:\users\user\appdata\l
ocal\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache
\local-packages\python310\site-packages (from typer<1.0.0,>=0.3.0->spacy)
(1.5.4)
Requirement already satisfied: rich>=10.11.0 in c:\users\user\appdata\local
\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\loca
l-packages\python310\site-packages (from typer<1.0.0,>=0.3.0->spacy) (13.9.
4)

Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in c:\users\user\a
ppdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\loc
alcache\local-packages\python310\site-packages (from weasel<0.5.0,>=0.1.0->s
pacy) (0.21.0)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in c:\users\user\app
data\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\local
cache\local-packages\python310\site-packages (from weasel<0.5.0,>=0.1.0->spa
cy) (7.1.0)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\user\appdata\loca
l\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\loc
al-packages\python310\site-packages (from jinja2->spacy) (2.1.1)
Requirement already satisfied: marisa-trie>=1.1.0 in c:\users\user\appdata\l
ocal\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache
\local-packages\python310\site-packages (from language-data>=1.2->langcodes<
4.0.0,>=3.2.0->spacy) (1.2.1)
Requirement already satisfied: markdown-it-py>=2.2.0 in c:\users\user\appdat
a\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcac
he\local-packages\python310\site-packages (from rich>=10.11.0->typer<1.0.0,>
=0.3.0->spacy) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in c:\users\user\appd
ata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localc
ache\local-packages\python310\site-packages (from rich>=10.11.0->typer<1.0.
0,>=0.3.0->spacy) (2.17.2)
Requirement already satisfied: wrapt in c:\users\user\appdata\local\packages
\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-package
s\python310\site-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.
1.0->spacy) (1.17.0)
Requirement already satisfied: mdurl~=0.1 in c:\users\user\appdata\local\pac
kages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-pa
ckages\python310\site-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->t
yper<1.0.0,>=0.3.0->spacy) (0.1.2)
Note: you may need to restart the kernel to use updated packages.
ERROR: Could not find a version that satisfies the requirement difflib (from
versions: none)

[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: C:\Users\user\AppData\Local\Microsoft\WindowsApps\P
ythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\python.exe -m pip install
--upgrade pip
ERROR: No matching distribution found for difflib
Requirement already satisfied: numpy in c:\users\user\appdata\local\packages
\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-package
s\python310\site-packages (1.26.4)
Note: you may need to restart the kernel to use updated packages.
[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: C:\Users\user\AppData\Local\Microsoft\WindowsApps\P
ythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\python.exe -m pip install
--upgrade pip
Collecting sklearn
  Using cached sklearn-0.0.post12.tar.gz (2.6 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'error'
Note: you may need to restart the kernel to use updated packages.

```
error: subprocess-exited-with-error

× python setup.py egg_info did not run successfully.
│ exit code: 1
╰─> [15 lines of output]
    The 'sklearn' PyPI package is deprecated, use 'scikit-learn'
    rather than 'sklearn' for pip commands.

    Here is how to fix this error in the main use cases:
    - use 'pip install scikit-learn' rather than 'pip install sklearn'
    - replace 'sklearn' by 'scikit-learn' in your pip requirements files
      (requirements.txt, setup.py, setup.cfg, Pipfile, etc ...)
    - if the 'sklearn' package is used by one of your dependencies,
      it would be great if you take some time to track which package uses
      'sklearn' instead of 'scikit-learn' and report it to their issue tra
cker
    - as a last resort, set the environment variable
      SKLEARN_ALLOW_DEPRECATED_SKLEARN_PACKAGE_INSTALL=True to avoid this
error

    More information is available at
    https://github.com/scikit-learn/sklearn-pypi-package
    [end of output]

  note: This error originates from a subprocess, and is likely not a problem
with pip.

[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: C:\Users\user\AppData\Local\Microsoft\WindowsApps\P
ythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\python.exe -m pip install
--upgrade pip
error: metadata-generation-failed

× Encountered error while generating package metadata.
╰─> See above for output.

note: This is an issue with the package mentioned above, not pip.
hint: See above for details.
```

In [ ]:
```python
import pandas as pd
import spacy
from difflib import get_close_matches
from sklearn.metrics.pairwise import cosine_similarity
import numpy as np

# Descargar el modelo de lenguaje en español
!python -m spacy download es_core_news_md

# Cargar modelo de lenguaje
nlp = spacy.load("es_core_news_md")

# Cargar los archivos Excel
file_ceibal_combine = "../data/Ceibal.xlsx"
file_codeorg_combine = "../data/CodeOrg.xlsx"
file_thetech_combine = "../data/TheTech.xlsx"
file_restantes_combine = "../data/Restantes.xlsx"
```

```
file2 = "Enlaces.xlsx"

df_ceibal = pd.read_excel(file_ceibal_combine)
df_codeorg = pd.read_excel(file_codeorg_combine)
df_thetech = pd.read_excel(file_thetech_combine)
df_restantes = pd.read_excel(file_restantes_combine)

df1 = pd.concat([df_ceibal, df_codeorg, df_thetech, df_restantes], ignore_in
df2 = pd.read_excel(file2, header=None, names=["Name", "URL"])  # Datos con
```

```
[notice] A new release of pip is available: 24.3.1 -> 25.0.1
[notice] To update, run: C:\Users\user\AppData\Local\Microsoft\WindowsApps\P
ythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\python.exe -m pip install
--upgrade pip
```
Collecting es-core-news-md==3.8.0
  Downloading https://github.com/explosion/spacy-models/releases/download/es
_core_news_md-3.8.0/es_core_news_md-3.8.0-py3-none-any.whl (42.3 MB)
     ---------------------------------------- 0.0/42.3 MB ? eta -:--:--
     - -------------------------------------- 1.8/42.3 MB 20.2 MB/s eta 0:0
0:03
     --------- ------------------------------ 10.0/42.3 MB 32.7 MB/s eta 0:0
0:01
     --------------- -------------------- 18.4/42.3 MB 36.2 MB/s eta 0:0
0:01
     ------------------------ ------------ 27.3/42.3 MB 37.6 MB/s eta 0:0
0:01
     ------------------------------- ------ 35.7/42.3 MB 38.4 MB/s eta 0:0
0:01
     ------------------------------------ 42.2/42.3 MB 38.9 MB/s eta 0:0
0:01
     ------------------------------------ 42.2/42.3 MB 38.9 MB/s eta 0:0
0:01
     ------------------------------------ 42.2/42.3 MB 38.9 MB/s eta 0:0
0:01
     ------------------------------------ 42.3/42.3 MB 26.9 MB/s eta 0:0
0:00
✔ Download and installation successful
You can now load the package via spacy.load('es_core_news_md')
```

In [3]:
```
# Pasarlos a string
df1 = df1.astype(str)
df2 = df2.astype(str)

# Comprobar el tipo de nuevo
print(df1.dtypes)
print(df2.dtypes)
```

```
Filename              object
Name                  object
Resource type         object
Values                object
Key competences       object
Knowledege areas      object
Cc concepts           object
Ct skills             object
Duration              object
School years          object
Languages             object
dtype: object
Name    object
URL     object
dtype: object
```

In [4]:
```python
# Crear una nueva columna URL en df1
df1["URL"] = ""

# Crear un diccionario basado en los nombres de df2
url_dict = {row["Name"].lower(): row["URL"] for _, row in df2.iterrows()}

def get_semantic_match(target, choices):
    """Encuentra la mejor coincidencia semántica usando embeddings de spaCy.
    target_doc = nlp(target)
    best_match = None
    best_score = 0.0

    for choice in choices:
        choice_doc = nlp(choice)
        similarity = target_doc.similarity(choice_doc)
        if similarity > best_score:
            best_score = similarity
            best_match = choice

    return best_match if best_score > 0.7 else None  # Umbral de 0.7 para co

# Intentar coincidencias exactas basadas en el Filename
for idx, row in df1.iterrows():
    filename = str(row["Filename"]).strip()
    name = str(row["Name"]).strip().lower()

    # Buscar coincidencia basada en filename
    for key, url in url_dict.items():
        if filename in url:
            df1.at[idx, "URL"] = url
            break

    # Si no encontró por filename, buscar por name (semántico)
    if not df1.at[idx, "URL"]:
        best_match = get_semantic_match(name, url_dict.keys())
        if best_match:
            df1.at[idx, "URL"] = url_dict[best_match]
        else:
            # Si no hay coincidencia semántica, buscar coincidencia basada e
            match = get_close_matches(name, url_dict.keys(), n=1, cutoff=0.6
```

```
        if match:
            df1.at[idx, "URL"] = url_dict[match[0]]

# Guardar el resultado en un nuevo archivo Excel
output_file = "output.xlsx"
df1.to_excel(output_file, index=False)

print(f"Archivo guardado en {output_file}")
```

C:\Users\user\AppData\Local\Temp\ipykernel_14136\2413517763.py:15: UserWarning: [W008] Evaluating Doc.similarity based on empty vectors.
  similarity = target_doc.similarity(choice_doc)
Archivo guardado en output.xlsx

In [5]:
```
# Mostrar el resultado
df1.head()
```

Out[5]:

| | Filename | Name | Resource type | Values | comp |
|---|---|---|---|---|---|
| 0 | Tarjetas-Bebras-I.pdf | Desafío Bebras: Retos de Pensamiento Computaci... | Actividad enchufada | Fomento de la creatividad y del espíritu cient... | Com Li Con |
| 1 | Tarjetas-Bebras-Tanda-2.pdf | Desafío Bebras | Actividad desenchufada | Pensamiento crítico, Educación para la Salud, ... | Con Con Mat |
| 2 | Building_ai_education_ori_digital-2.pdf | Construyendo la Inteligencia Artificial para l... | Actividad enchufada, Curso | Fomento de la creatividad y del espíritu cient... | Con Con Mat |
| 3 | 2023-Numero-la-rana_Guia-docentes.pdf | La rana que aprendió a contar | Programación visual | Fomento de la creatividad, Pensamiento crítico... | Con Con Mat |
| 4 | 2024-Nuestro-Show-Programado-Guia-docentes.pdf | Nuestro Show programado | Programación visual, Curso | Fomento de la creatividad y del espíritu cient... | Com Li Con |