ZADANIE KONKURSOWE DLA KANDYDATÓW NA STAŻ W ZESPOLE IN-SILICO

Napisz program w języku Python, parsujący automatycznie dane ze strony internetowej ECACC (European Collection of Authenticated Cell Cultures) jednego z większych dostawców liniach komórkowych.

Program na podstawie adresu URL odczyta dane w sekcji „ECACC General Cell Collection" w następujący sposób:

1. Wynikiem przetwarzania danych ma być słownik (typ danych języka Python).

2. Kluczem w słowniku ma być każde słowo (wyrazy) wytłuszczone na stronie ECACC (lewa kolumna tabeli w której prezentowane są dane). Klucz musi zawierać wyłączenie małe litery (duże zamienić małymi), spacje muszą być zastąpione podkreśleniem "_", znaki interpunkcyjne typu dwukropek należy usunąć z klucza.

3. Wartością klucza powinny być odpowiadające mu dane (prawa kolumna tabeli danych).

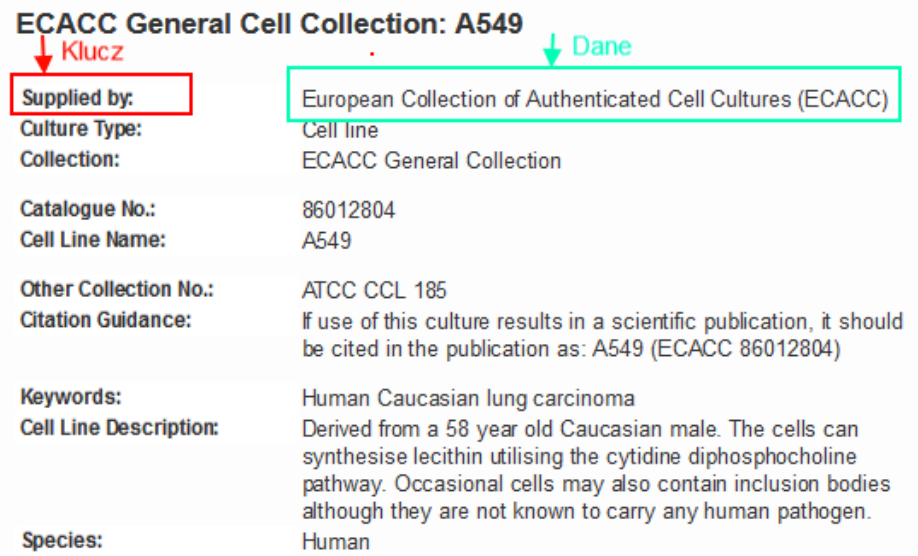
Zadanie dodatkowe. Dodaj opcję pozwalającą na przetwarzanie tekstowych plików wsadowych zawierających w każdej linii inny adres URL który powinien być przez program parsowany. Efektem powinno być wypisane na ekranie dla każdego adresu URL wynikowego słownika o którym była mowa w punkcie 1.

Wskazówka: Wykorzystaj bibliotekę **„***Beautiful Soup*".

Przykładowe adresy URL:
https://www.phe-culturecollections.org.uk/products/celllines/generalcell/detail.jsp?refId=85120602&collection=ecacc_gc

https://www.phe-culturecollections.org.uk/products/celllines/generalcell/detail.jsp?refId=93120835&collection=ecacc_gc

https://www.phe-culturecollections.org.uk/products/celllines/generalcell/detail.jsp?refId=86012804&collection=ecacc_gc



Rys. 1. Screen ze strony ECACC pokazujący dane które należy parsować wraz ze sposobem oznaczenia dla klucza i danych.