Take-Home Code Challenge

Overview: A quick description of the task

One of our current tasks at Prolegis is to build tools to help understand how lobbying impacts congress. Oftentimes, this means a lobbyist or a legal firm will have an interest in specific bills.

For example, a lobbying firm that is hired to represent an oil company will be interested in bills related to environmental regulations, but they probably aren't interested in bills related to immigration or education.

Lobbyists are legally required to register with Congress and file **reports** on anything they are lobbying on. These reports are freely available and are a key part of our work on this feature. However, these reports don't contain data points for what they are doing. They just contain a "**description**" field. So lobbyists will fill these fields in differently, and there are no data constraints on what the "description" field has to contain.

Here is an example of a description field: "Issues related to government civil fraud enforcement; HR 2515; IG issues"

This task is based around a real problem we've had to solve at Prolegis: **Taking a "description" field full of text, figuring out if it mentions a bill, and if so, documenting that relationship.**

There are many ways to potentially solve this problem. Your challenge is to come up with a solution of your own.

Background: Some context on how bills get their IDs

Every bill in congress is assigned a unique ID, based on where it came from and what it does. Congress has two chambers, the House and Senate: House bills start with 'hr', and senate bills start with 's'. We're ignoring senate bills for this exercise and instead focussing on house bills, aka 'hr' bills.

Here are some examples of what bill IDs look like: "hr182-117", "hr1-116", "hr3597-113"

The end of a bill ID, after the hyphen, is the congress number. This number increments every 2 years, after elections. The current congress is 117, which started on January 1, 2021.

Looking backwards, the 116th congress started in 2019, the 115th started in 2017, and so on.

At Prolegis, we only look at data going back to the 113th congress, which started on Jan 1, 2013.

Task: Instructions on what you'll need to create

Attached is a CSV that contains three columns:
- **id**: The ID of the lobbying report
- **description**: A 'description' field in the lobbying report
- **year**: The year the report was submitted

Your task is to write a ruby script that takes this CSV and converts it to a list of URLs pointing to bills referenced by each report. Here are the requirements:

1. Descriptions should be parsed for any and all references to a bill.
   a. Bill IDs in descriptions are commonly written to be human-readable. Here are some common variations, all of which are valid:
      ```
      H.R. 100
      H.R 100
      HR 100
      H.R.100
      H.R100
      HR100
      ```
2. "year" needs to be converted to congress number, and appended to the bill references found above to build a full bill ID.
3. For each report ID, Build a list of links to the bills referenced in their filings
4. Write this list to a JSON file with the structure shown:

```json
[
  {
    "id":7,
    "urls": [
      "https://www.prolegis.com/cards/v1/bills/hr42-117",
      "https://www.prolegis.com/cards/v1/bills/hr143-117",
      "https://www.prolegis.com/cards/v1/bills/hr23-116"
    ],
  },
  {
    "id":14,
    "urls": [
      "https://www.prolegis.com/cards/v1/bills/hr843-115",
      "https://www.prolegis.com/cards/v1/bills/hr97-117",
      "https://www.prolegis.com/cards/v1/bills/hr1-117"
```

```
    ],
  }
]
```

In this folder, you'll find two other files: activities.csv, and main.rb. main.rb should be runnable directly:

```
$ ruby main.rb
```

The main.rb file contains some boilerplate code to help you get started.