CSEN100_Project_2.pdf Project Journal

1. Reading references:

- Read CSEN100_Project_2.pdf https://canvas.instructure.com/courses/12318939/files/303388742/download?download_frd=1
- Read through PRAW 7.8.2.dev0 documentation
- Follow the instruction here and create a Reddit APP and api credentials Authenticating via OAuth - PRAW 7.8.2.dev0 documentation
- In order to hide my credential of Reddit API, I used a separate file to store my api_key, and named it api_key.json
- I will started with the sample topics: [”christmas”, ”fruitcake”, ”meatloaf”,”new year’s”, ”pie”]
- The search method is well documented here and I just follow it. Subreddit - PRAW 7.8.2.dev0 documentation
  It also provided an example to search within the results:

```
for submission in reddit.subreddit("all").search("praw"):
      print(submission.title)
```

- The properties of a submission can also be found here Submission - PRAW 7.8.2.dev0 documentation
- One of the Project 2 specifications also recommend us using 'flair' to tell whether the post is a recipe or not.
- Specifications also require us to
  - get the username of the person who posted the recipe and output it to the console as well.
  - save the recipe post image as a local file.

2. Code implementation

At this point, I have everything I need to implement the code.
The code-flow is simple, I search one topic at a time in the r/recipes subreddit.
Get the top 50 posts and find the one with max upvotes.

When I use flair to filter posts with "recipe" flairtext, I found many recipes don't use "Recipe" as their flair, and they use other flair-text, like "Dessert", "Beef", "Seafood", "Pork", "Poultry", etc. So I ended up filtering the posts with a larger list of "okay-items".

The recipe is almost 100% the top pinned comment.

And I just need to search the comment with the highest "scores" (actually upvotes) for the contents of the recipe.

It took me some web search to realize how to download the image.
If a post has an image, the image url is just the post url.
However, the post url sometimes may not be an image.
So I need to check whether the file extension belongs to image formats.

But still, it's not as trivial because sometimes the image url does not show the proper image format extension as it is.
After more web searching, it turned out I needed to use "PRAW preview" in this situation.
It really took me some time to figure it out, because even the PRAW document didn't mention it.
[Submission - PRAW 7.7.1 documentation]
And in this case I have to treat the grabbed url dictionary in a different way.

The code works actually pretty well and I can add some more topics.
My current topic list is as follows and it works well.

```
TOPIC_LIST = [
    "christmas"
    "fruitcake",
    "meatloaf",
    "new year's",
    "Pie",
    "fried rice",
    "beef noodles",
    "ramen",
    "alfredo chicken pasta",
    "xiaolongbao",
    "egg salad sandwich",
    "oyakodon",
    "bibimbap",
    "Kung Pao Chicken",
    "Nasi Goreng",
    "Pad Thai",
    "Bulgogi",
    "Samosas",
    "Mapo Tofu"
]
```

3. Project 2 Challenge

I read the document first.
https://canvas.instructure.com/courses/12318939/files/303388743/download?download_frd=1

But I am stuck at Part 1.
I have used pandas before but am not very familiar with it.
In the document it mentions there is a file “recipe info.csv” but I cannot find it anywhere from the files of this course. 文檔

As for Part 2.
Since we can automatically extract information from Reddit, I think there are several applications or projects we can do, for example:

- Personalized News Aggregator/Newsletter
- Niche Problem/Pain Point Finder
- Trend and Hype Detector
- Q&A/FAQ Generator
- etc