

SI 506 | Fall 2017 | Final Project

DEADLINE: December 15, 2017, 11:59 PM.

Total Points: 3600

(400 Points for Final Project Plan, submitted in advance of the Final Project — it is included with PS9 and is due Monday, November 27th at 11:59 PM.)

Point allocations for specific portions of the project requirements may change from the approximate totals listed below, but not dramatically. The total points for the final project submission will remain 3600.

You cannot use late days for the final project. We have to turn the grades in 72 hours after the final.

To Submit

Submit to the Final Project assignment on Canvas the following files:

- **SI506_finalproject.py, a code file containing all the code for your technical requirements** *(Approx 2800 total points)*
 - You may include other files and *import* them into this one, if you like. But you may also write all the code you are writing inside this one file.
- **SI506finalproject_cache.json**, containing the cached data from running your project. *(Approx 100 total points)*
- **README.txt**, a text file containing information that follows the sample **README.txt** file we have provided, including filenames and approximate line numbers for each of the project requirements *(Approx 500 total points)*
 - It's very important that you include all of that information — it helps us grade!
- **An example of what your project output looks like.** This could be a sample file (make sure to save it as a *different* name than what your project will produce, so it does not get over written!), or a screenshot. We should understand what we are supposed to see once we've run the project, if it works. *(Approx 150 total points)*
- **ANY OTHER FILES WE WILL NEED TO RUN YOUR CODE.** Making sure your code runs should be your first priority, and points will be deducted for all code that does not work.

Overview & Recommendations

We have made two suggestions, “options,” for the final project.

Each one has the same technical requirements, plus some guidance about which APIs to access for data, and which classes to define.

In each of our two suggestions, you will still need to make a few decisions and exercise some creativity. You can pick either one, complete it, and turn it in!

There is also a third option: if you wish to propose your own project, you can do that. **It must fulfill every single one of the technical requirements below.** We will not make exceptions: there are too many of you.

We strongly suggest that you choose one of our two options. It will be more difficult to come up with your own ideas AND execute them all in the amount of time we have — designing a project is a substantial process! If you feel confident doing this, you may, but if you do not, wait and do that project after class with the skills you have learned here.

We will *not* grant extensions because you have decided to change your project idea. Take advantage of the support we offer for you to plan your project and get started in advance.

You *may* change your project idea or any specifics about your project at any time without checking with an instructor. (I always get about 25 emails asking specifically this question, so I’ve bolded it this time.) **But be cautious about why you’ve chosen to do this!** It’s often better, easier, and ultimately less stressful, to sit down and solve your problem with a friend or an instructor than it is to totally switch what you’re doing.

Project Requirements

- Get data from 2 different REST APIs
- Cache all data from REST APIs in one file (unless it’s FB data)
- Define at least 2 classes
 - Each class must have
 - A constructor

- A string method
 - At least one additional method
 - At least 3 instance variables (or 3 possible instance variables)
- You must create at least 1 instance of each class you define
- You must *use* each instance of the classes you create:
 - You must invoke the string method and the additional method for each
 - This could include e.g. invoking the additional method *inside* the `__str__` method, as you saw in some in-class examples!
- You must perform at least one sort with a key parameter on data.
 - This must be used in the process of your program (e.g. if you perform a sorting operation in your program that makes no difference to how your program works, that does not count. Planning this is part of the project)
- You must define at least 2 functions outside the class definitions that have more than one line inside the function body.
 - These could be functions to get data from APIs and/or other functions.
- Your code must create a file that is clearly structured for output.
 - Could be a .CSV file
 - Could be a .txt text file
 - Must be readable by a human being
 - And when a human being reads it, even if they have no idea what Python is or what your project was, they should have a *basic* idea of what that CSV file or text file means
 - That means you should have clear headers, a clear title, explanatory text, whatever you need
 - **Note:** You should *not* worry about making “discoveries” or ANY statistically significant result(s). This is not a statistics course. You have not learned enough scientific computing in this course to do that in Python without additional knowledge. Use the suggested options as guidelines. You are simply looking to process some data and output your results in a clear way — like a more complicated version of PS9, PS7, or PS6.

Option 1: Facebook & iTunes

- Get the last 25-50 posts from your Facebook feed (or the feed of a public group you like), using the Facebook graph API, similar to how you did in PS9. You are not required to cache the data from Facebook, but you can! See [Lecture 22 notes](#) for tips on how to do it.
- Define a class Post, but make it different from the class you defined in PS9.
 - You should have a `__str__` method that represents information about the post clearly

- Your additional method should return a list of all the words in the post that are not “stopwords” (e.g. “and”, “the”, “is”, etc — super-common words like that are often called “stopwords” in computational linguistics).
- You may have other additional methods, if you like.
- Create a list of instances of class Post.
- Using those instances and their capabilities, find the most common word that is not a “stopword” among all of those posts.
- Search on that word in the iTunes Search API — for songs that come up when you search with that as a search keyword. Make sure you are caching your data in the cache file.
- Define a class Song, but it should be different from the class you defined in section and/or the one you see in the textbook.
 - It should have a `__str__` method that represents information about the song clearly, including the length of the song.
- Create a list of instances of class Song.
- Sort the song instances you get back from the iTunes API by the songs' *length*, longest to shortest.
- Create a CSV file with information about all of those songs:
 - The song title
 - The artist of the song
 - The length of the song
 - The album the song is on
 - And any other column(s) of information you want

Option 2: NYT & Flickr

- Make a search on the New York Times ArticleSearch API for anything you want — anything you're interested in. Make sure you are caching your data in the cache file.
- Define a class Article. It should have appropriate instance variables, including a list of article keywords (available in the data), a string method that displays clear information about the article, and an additional method that finds the longest word in the abstract. You could certainly have more!
- Create a list of Article instances.
- Sort that list of Article instances by the article's number of keywords, most → least. Break ties alphabetically by article title. Find the 5 with the most keywords
- Get the longest word of each abstract of the 5 articles with the most keywords.
- For each of those words, make a search on Flickr. Make sure you are caching your data in the cache file.
 - It could be for photos *tagged* with that word
 - It could be a general search for that word on the Flickr Photos search API... anything you want.

- Define a class Photo. It should be different from the one you saw in class, and different from the one on the 506 Waiver, if you have ever seen that code.
 - It should have appropriate instance variables
 - A `__str__` method that displays clear info about the photo
 - An additional method that returns the total number of tags on the photo.
 - Could have any other methods you decide want to include as well.
- Make another set of requests, this time to get data about each of at least 20 photos that resulted from the Photo Search to Flickr. Make sure you're caching your data in the cache file.
- Create a list of Photo instances, and sort that list of instances by the *number of tags* each photo has.
- Write a .CSV file about those photos, with clear headers:
 - Photo title, if any
 - Username of person who posted the photo
 - All the tags (careful about commas, if you want tags in the same cell in the spreadsheet!)
 - The number of tags the photo has
 - Any other column(s) of information you want

Option 3: Create Your Own

You may also design your own project. **HOWEVER, some things to note if you do that:**

- YOU are responsible for ensuring that your project fulfills all of the technical requirements. You're welcome to talk it over with an instructor, but that responsibility is still yours. You will know the project better than anybody else.
- You should be extremely clear in your README — even more so than those who choose Option 1 or Option 2, so we know what to look for when grading!
- You should be *very* careful to make sure that you *can* get the data you plan on using. Check that you can get all the data you need in advance. As we've discussed, not all APIs are as good as you hope they will be for your end goals, and not all documentation is ideal.
- One recommended option for creating your own is taking Option 1 or Option 2 as a starting point, and making some changes to make the result something you are more interested in trying, or making some changes to make a more complex search, more complex class definition(s), more complex visual/readable output.
 - Maybe you're more interested in adding information about NYT articles to a CSV. Maybe you're interested in specifically searching for Music Videos, not songs.

Whatever you want. You do *not* need to check with us to do this! You simply need to turn in a project that fulfills all of the technical requirements in order to get full credit. Any small alterations that work are definitely fine.

Additional Important Notes — Academic Honesty & Rules for the Project

- Even though many people will be working on the same project outline as many others, everyone's project should be a little different because it is their own. *You* should type 100% of the code you turn in. You should *not* directly copy anyone else's work in the process of this project in any way. You are, of course, more than welcome to seek or give help to/from others! Just make sure you aren't giving away any specific answers, and make sure you don't feel like you are seeking or giving too much. You should feel like you can recreate your own project, and you should be able to explain it to someone else (especially since you have to do so to write the readme!)
- Especially if you choose Option 3, you *may* use any code from previous lectures or problem sets, but no code that is exactly the same as that that was given in lecture, section, homework, or from the textbook, may count to fulfill a project requirement.
- You may use external libraries that help you get data from APIs, but you still must cache your data, so you'll need to set up a caching pattern that allows you to do that. (This is not very dissimilar to what we've done in class, but it is a tiny bit different than the exact pattern we looked at in class, and that you used e.g. in PS 7.)
- These options provide you with general guidelines, but you'll still need to make decisions about:
 - What functions and setup to write to get and cache your data (similar to how you've done so before!)
 - What classes to define
 - How to parse your data and create instances of your classes
 - How to use your class instances to process data and create a resulting file
 - How to make decisions about things you see from real data and adapt to what happens when you make API requests and see what the data is going to look like (check out Lecture 22 for more!)

Asking Questions about your project (IMPORTANT)

- Welcomed and encouraged! Please do so often, if you have questions!
- **But as you do so, you should prepare the following things, when you want to ask a question. Have several of them ready, if you are asking in person, or write them out, if you are asking on Piazza. That will help others, including instructors, answer your question easily, and help you debug.**
- We know it sounds like a lot of work — but it's a really important part of the process. And it will either help you answer your own question to go through these things systematically, or it will make it very easy for an instructor to help you get the answer!

- **What is a sentence describing what you want to happen/what you want your code to do?**
 - e.g. "I want my code here to loop over a list of dictionaries, get a photo id out of each one, and append it to a list called my_photo_ids"
- **What have you tried so far to solve the problem?**
 - e.g. "I wrote a for loop to try to iterate over all the photos, but I'm getting an error when I try to run that code, and it looks like that for loop is what is causing the error. Then I tried printing out the sequence in the for loop, but it looks fine, so I am not sure what to try printing out next."
- **What specific lines of code, if you're sharing code, are you confused about?**
 - e.g. "I thought lines 34-38 would get the ids of all the photos in a list my_photo_ids, but they don't seem to be working. The list my_photo_ids is always empty, or I get an error message that says... <whatever the error message says> and I don't know what that means."
- **What is different between what *is* happening when you run your code and what you *expect* to happen?**
 - e.g. "I expect to be iterating over a series of dictionaries, getting a string out of each one, and adding that string to a list. But after I run the code, there's nothing in the list, it's still empty!"
- **If you are getting an error message, what do you think it means in English?**
 - You do not have to be right — we humans are often wrong about these things! — but you DO have to make an attempt to figure it out and translate it from Python-error to an English sentence.
 - e.g. "I think the error message means I'm looking for a key that doesn't exist in a dictionary, but I can't figure out where that's happening."
- **What information have you searched for or re-read to help you find answers?**

What section exercises have you looked back at? What textbook chapters' examples have you looked at? What lecture code or code samples on Canvas or problem sets have you checked out?

 - If you tell us this, we can recommend specific examples to look at to help you solve the problem, and we can point out things that might be confusing.

- If you haven't done any of this, the first thing we'll do is point you to examples that can be helpful! We can also help suggest specific questions to try to answer about the examples — we do this because it will help you analyze the code in a more useful way, and you'll be able to keep solving problems.