Course:	**INFO3142**
Professor:	**Osam Ali**
Project:	**Project #2 – Stock Buy/Sell Estimator – Version 1.0**
Due Date:	**Friday, December 02, 2022 / ALL Sections**
Submitting:	**Please see the last page for instructions**

How will my project be graded?

- This project counts for 25% of your final grade and will be evaluated using the following grid:

Marks Available	What are the Marks Awarded For?	Mark Assigned
5	Uses spaCy and Python to perform POS and entity tagging on the sample data in the EmailLog.txt file	
9	Uses Python to create the correct console output as shown in the Addendum	
1	Proper submission	
3	Bonus – associating dollar amounts with each company where multiple requests have been made in a single email (see Addendum)	
15	**Total**	

Project #2 Requirements

Introduction:

This final project will give you an opportunity to apply some natural language processing concepts to a real-world (semi-real-world) application.

The main focus will be to extract some free-form information from a series of (pre-filtered) email messages.

The business theme behind the project involves consuming a log file consisting of multiple emails from customers, and then building a simple report based on a subset of information from the emails.

The project is somewhat simplified in order to reduce the work required.

The business in this case will be a fictional stock brokerage firm and the customer emails will all be requesting the brokerage firm to purchase stock of a specified amount for one or more designated companies.

Example:

The following is an example entry from the "EmailLog.txt" batch email input file to be used by this project (see the Addendum).

tom.brady@gmail.com
Hi Dave, I'd like to invest $10,000 with Microsoft and another $15,000 with Amazon. Thanks, Tom
<<End>>

Your solution is to determine the total amount of money being invested and which companies are receiving the investment. For example, from the above request, you should produce a textual output to the console of:

tom.brady@gmail.com: $25,000 to Microsoft and Amazon

The emails have all been pre-processed and placed into a single log file with some parts fixed and some free-form.

Fixed Parts:

- Each email in the log file begins with a single line containing the sender's email address.
- Each entry ends with a single line containing "<<End>>" as a tag.

Free-form Part:

- The email text as above in colour.

Note: The email text will have all blank lines removed although the email could consist of multiple lines.

Doing the Project:

When it comes to implementing the project, you have to complete 2 separate operations using Python and spaCy as follows:

1. Do an initial parse on the emails log file and extract the POS and entity tags for each of the emails. This information should be stored in a separate text file

2. The generated stored text file which in step 1 can subsequently be consumed **(optional)** by another python script to get the proper output in the console. Or you can use entirely spaCy script to handle all the required tasks without using the generated text file from step 1.

Since Python and spaCy are an important part of the course, all students must at least use Python/spaCy as a means to extract basic syntax information from each of the emails.

See chapter 4 in the textbook for some examples on extracting both course-grained and fine-grained parts of speech, as well as dependency labels.

With this kind of project, it's beneficial to have some degree of customization in looking for specific tokens (see chapter 8) however you can't have so much customization (hard-coding) that you'll start to fail against a large percentage of valid emails.

When your project is being graded, it will be tested using data that is similar to the EmailLog.txt that's provided.

***** End of Requirements *****

Addendum

EmailLog.txt

tom.brady@gmail.com
Hi Dave, I'd like to invest $10,000 with Microsoft and another $15,000 with Amazon. Thanks, Tom.
<<End>>

bjohnson@sympatico.ca
As we discussed, let's try $8500 on Apple. Hopefully, that's OK. -Brooke
<<End>>

andy.average@gmail.com
Edward, can you do $50 thousand to Infotech Inc. and another $10,000 with Amazon? Also, let's try just $5000 for Newcorp stock. Call me if you have any questions.
<<End>>

john.public@rogers.com
I've finally decided on $11,200 with Microsoft. Always liked Windows.
Regards, John
<<End>>

Output to the Console Window

tom.brady@gmail.com : $25,000 to Microsoft and Amazon

bjohnson@sympatico.ca : $8,500 to Apple

andy.average@gmail.com : $65,000 to Infotech, Amazon and Newcorp

john.public@rogers.com : $11,200 to Microsoft

Total Requests: $109,700.00

Bonus mark for associating dollar amounts allocated for each company:

tom.brady@gmail.com : $25,000 to Microsoft and Amazon. $10,000 to Microsoft and $15,000 to Amazon.

bjohnson@sympatico.ca : $8,500 to Apple

andy.average@gmail.com : $65,000 to Infotech, Amazon and Newcorp. $50,000 to Infotech, $10,000 to Amazon and $5,000 to Newcorp.

john.public@rogers.com : $11,200 to Microsoft

Total Requests: $109,700.00

How should I submit my project?

Electronic Submission:

One of the Group members should submit the project to the INFO3142 “Project #2” electronic submission folder in FOL. Your code should be submitted as a single “zip” file containing your **Python scripts’** solution.

Submit your project on time!

Project or essay submissions must be made on time! Late submissions will be subject to divisional policy on missed test and late projects. In accordance with this policy, no late submissions will be accepted without prior notification being received by the instructor from the student.

Submit your own work!

It is considered cheating to submit work done by another student or from another source. Helping another student cheat by sharing your work with them is also not tolerated. Students are encouraged to share ideas and to work together on practice exercises, but any code or documentation prepared for a project must be done by the individual student. Penalties for cheating or helping another student cheat may include being assigned zero on the project with even more severe penalties if you are caught cheating more than once. Just submit your own work and benefit from having made the effort on your own.