ICT 2202 - Digital Forensics Project

School of Information Communication Technology

Team Name:	**NoPreference**
Nathan Lepori	**1903980**
Tok Boon Jin	**1801589**
Gary Koh Zi Wei	**1802005**
Zhuo GuoBao	**1801762**
Kwan Zheng Yi	**1800744**
Charlyn Low	**1802126**

Table of Contents

User Manual

Installation

Windows

Step No.	Instructions
1.	Download and install Python 3 for Windows.
2.	Download or clone the sit-no-preference repository from GitHub.com if not already did that and place it at any path.
3.	Run the install script located in the main directory: ● Cmd: `.\install.cmd` ● Powershell: `.\install.ps1`
4.	Optional: add `no_preference.cmd` to PATH to call it from anywhere.

Debian-based Linux (Debian, Ubuntu, Kali)

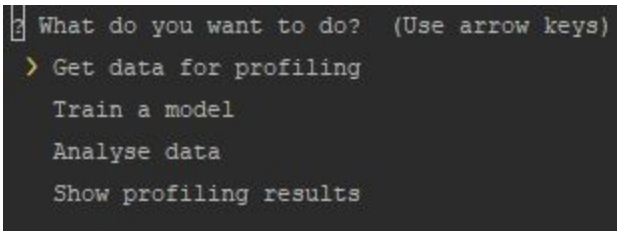
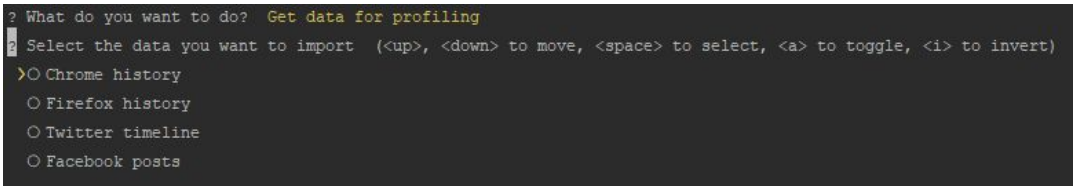
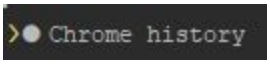
Step No.	Instructions
1.	Run `# apt-get install python3`.
2.	Download or clone the sit-no-preference repository from GitHub.com if not already did that and place it at any path.
3.	Run `# apt-get install python-dev python3-dev build-essential` to install needed dependencies.
4.	Run the install script `install.sh` (no root needed) located in the main directory.
5.	Optional: add `no_preference.sh` to PATH to call it from anywhere.

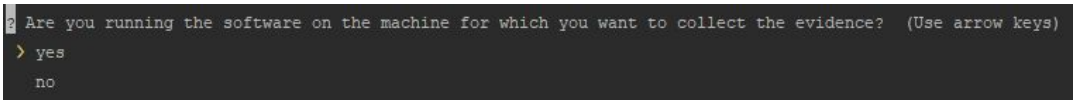
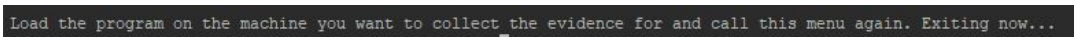
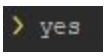
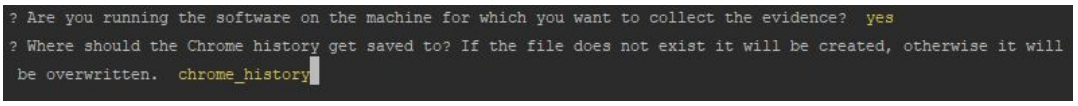
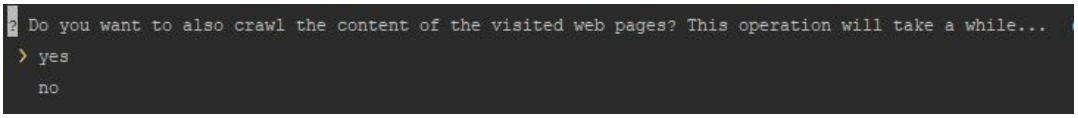
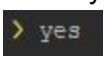
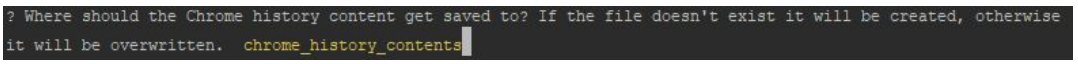
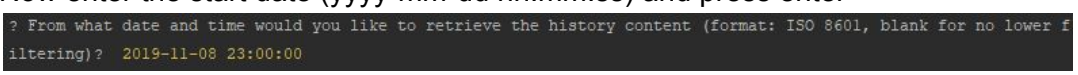
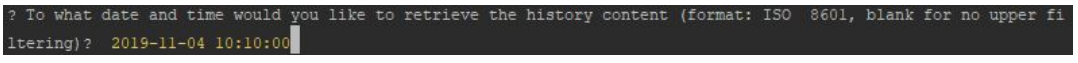
Other systems

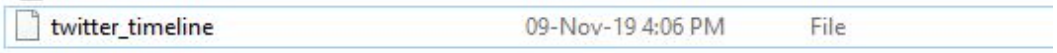
No support for installing or running on other systems is provided, but the software may still work by installing all needed dependencies and using a supported version of Python.

Usage

Getting data for profiling

Step No.	Instructions
1.	Open terminal
2.	Launch the program by using the scripts provided in the main directory: ● Windows: ○ Cmd: `.\no_preference.cmd` ○ Powershell: `.\no_preference.ps1` ● Linux/MacOS: `./no_preference.sh` ⚠ We currently only support Debian-based distributions. No support is provided for MacOS or other Unix systems, although the software may still work.
3.	The menu option will appear. It will appear as shown in the diagram below. 
4.	Select the Get data for profiling and press enter. 
5.	The profiling menu will appear and you can select the data you want to import. 
6.	Select Chrome history is you need to import the data from Chrome and press space to select it. It should look something like this after selecting Chrome history. 
7.	Now press enter and you should see this appear. If the software is running on the machine that needs evidence to be collected on that machine, proceed to step 10 else go to step 8.

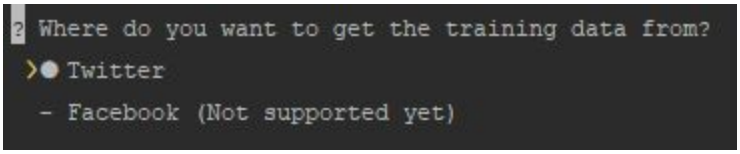
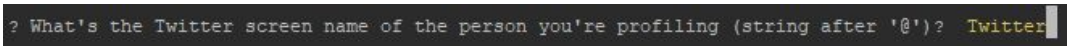
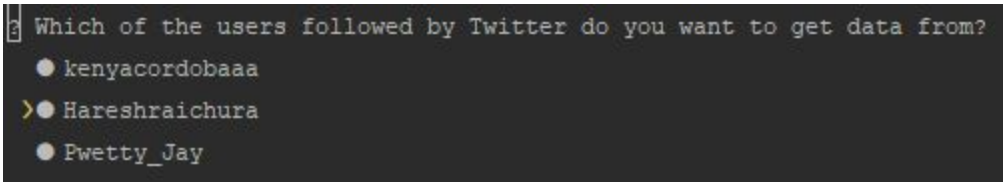
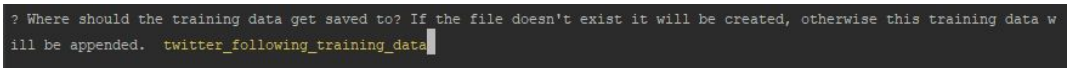
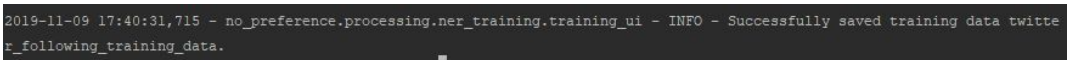
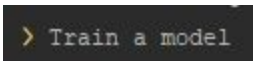
	
8.	If the software is not running on the machine with the evidence, select no and press enter 
9.	This message will appear 
10.	If the software is running on the machine that needs evidence to be collected, select the yes option and press enter. 
11.	It will prompt you to enter the name of the file which you want to save the chrome history to and press enter. I choose to save it chrome_history. 
12.	Now it will ask if you wish to get the contents of the web pages. 
13.	Select yes if you wish to get the contents of the web pages and press enter. 
14.	Now you will ask to enter the file name that you wish to save the chrome history contents to. I have chosen to name the file as chrome_history_content. Press enter. 
15.	Now enter the start date (yyyy-mm-dd hh:mm:ss) and press enter 
16.	Now enter the end date (yyyy-mm-dd hh:mm:ss) and press enter and wait for it to be done. 
17.	Once it is done getting all the data, this will appear letting you know that the crawling is done.

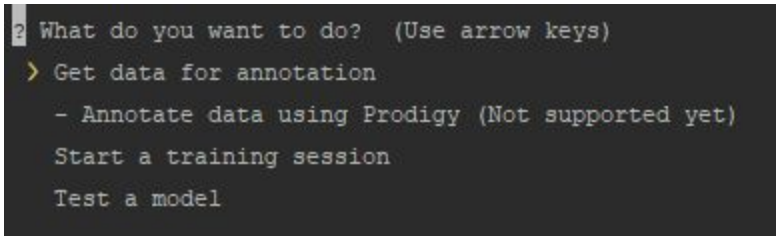
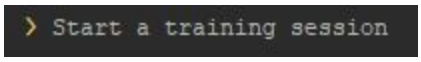
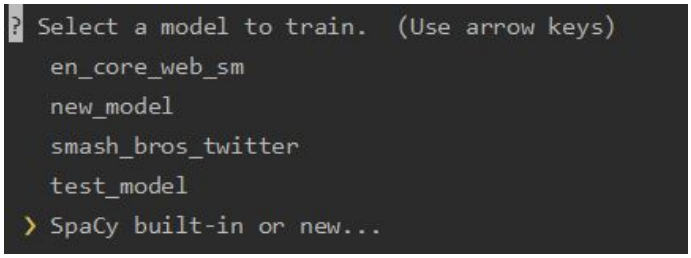
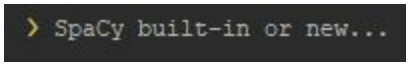
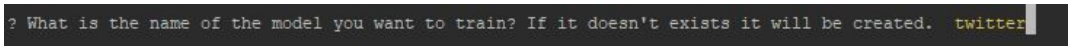
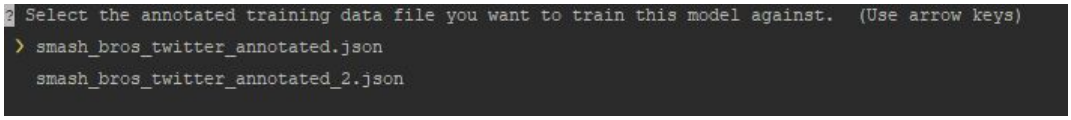
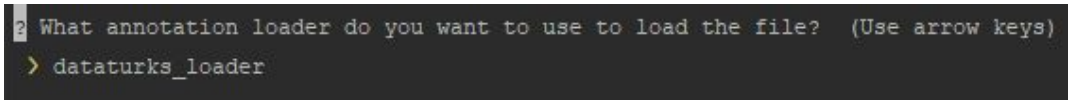
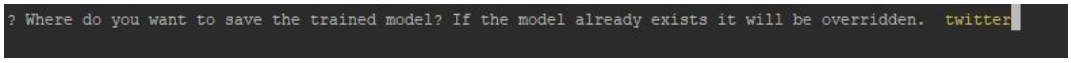
	`2019-11-09 15:39:42,189 - no_preference.datasets.datasets_ui - INFO - Saved dataset to data\datasets\history_data\chrome_history_contents`
18.	Going to the folder specified in the previous step, you will see that 2 files have been created and saved. 
19.	To get firefox history, follow from step 1 to step 4. Now choose Firefox history by pressing space and then press enter. `› ● Firefox history`
20.	Now follow through from step 7 onwards to retrieve your Firefox history.
21.	Run the program again and follow step 1 to step 4. To get the Twitter timeline, select Twitter timeline by pressing space and then press enter. `› ● Twitter timeline`
22.	Enter the username after the @ symbol and press enter. So for example, I wish to get Twitter's twitter account timeline so I enter Twitter. `? What's the Twitter screen name of the person you're profiling (string after '@')? Twitter`
23.	Now it will ask you for the file name and then press enter and wait for it to be done. I decided to save it as twitter_timeline. `? Where should the Twitter timeline get saved to? If the file doesn't exist it will be created, otherwise it will be overwritten. twitter_timeline`
24.	Once the program is done running, this message will appear signaling that it i done. `2019-11-09 16:06:00,772 - no_preference.datasets.datasets_ui - INFO - Saved dataset to data\datasets\social_data\twitter_timeline`
25.	Going to the folder specified in the step shown above, we can see that the file has now been created and saved. 
26.	Run the program again and follow step 1 to step 4. To get the Facebook posts, select Facebook posts by pressing space and then press enter. `› ● Facebook posts`
27.	This message will appear and when you have downloaded the data from the desired Facebook account, press enter.

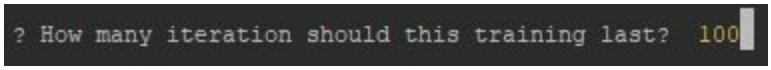
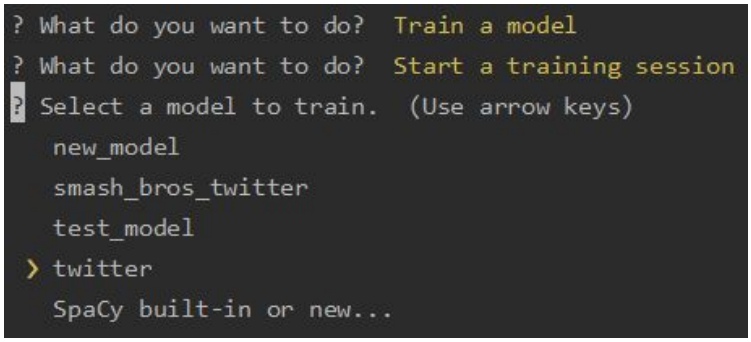
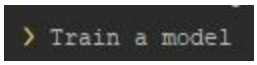
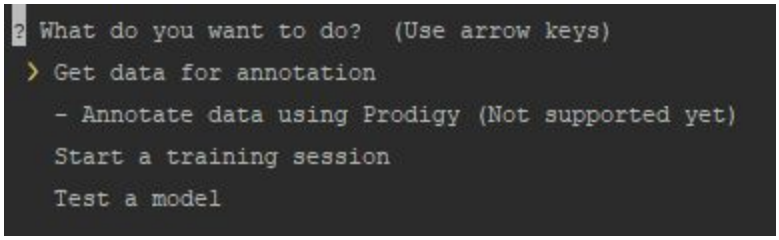
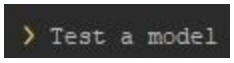
	Refer to https://www.facebook.com/help/212802592074644 on how to download data from a Facebook account. You will need the login credentials of the person you are profiling. Press enter when you are ready to proceed...
28.	Now you will need to enter the full path of the folder where the folder name is called posts as shown after extracting the zip file. ? Where is your Facebook data directory located? C:\Users\Charlyn Low\Downloads\facebook-charlynlowjialing\posts
29.	Now you will be asked to enter the file name and press enter. I choose to save it as facebook_data. ? Where should the Facebook data get saved to? If the file doesn't exist it will be created, otherwise it will be overwritten. facebook_data
30.	When it is done, a message similar to this will appear. 2019-11-09 16:36:46,548 - no_preference.datasets.facebook - INFO - Getting posts from C:\Users\Charlyn Low\Downloads\facebook json\posts\your_posts_1.json. 2019-11-09 16:36:47,096 - no_preference.datasets.facebook - INFO - Getting posts from C:\Users\Charlyn Low\Downloads\facebook json\posts\your_posts_2.json. 2019-11-09 16:36:47,273 - no_preference.datasets.datasets_ui - INFO - Saved dataset to data\datasets\social_data\facebook_data
31.	Going to the folder specified in the step shown above, we can see that the file has now been created and saved. facebook_data 09-Nov-19 4:36 PM File

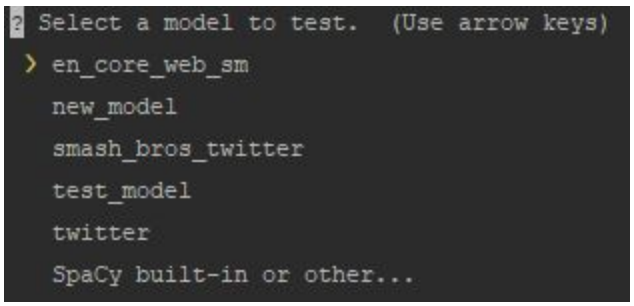
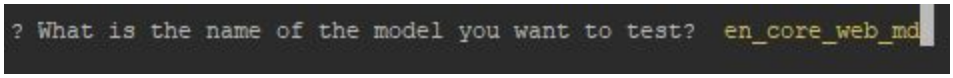
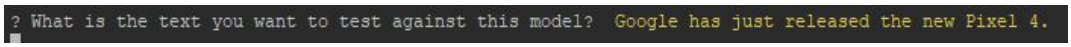
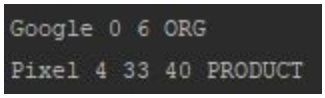
Train a model

Step No.	Instructions
32.	Run the program and navigate to Train a model and press enter. › Train a model
33.	This option menu will appear. ? What do you want to do? (Use arrow keys) › Get data for annotation - Annotate data using Prodigy (Not supported yet) Start a training session Test a model
34.	To get data, select Get data for annotation and press enter › Get data for annotation
35.	Now choose where you want to get the training data from. Right now only Twitter

	is supported so select Twitter by pressing space and then press enter. 
36.	You will be asked to enter the username of the Twitter account you wish to profile and then press enter. 
37.	Specify how many tweets you wish to download for the training data and then press enter. Default is 60 but here I chose to have 30 tweets downloaded instead. 
38.	You will be asked to select which users that is being followed by the account that you have entered previously to get the data from. Make sure to select data that is suitable for the labels that you want to recognise. Press space to select and then press enter and wait for it to end. I have decided to get data from these 3 users 
39.	You will be asked to enter the name of the file and then press enter. Here I chose to enter twitter_following_training_data. 
40.	When it is done, a message similar to this will appear. 
41.	Going to the folder called training_data, we can see that the file has now been created and saved. 
42.	Use any annotation tool to create annotated training data to feed the NER trainer in the next step. A recommended and supported free tool is Dataturks.
43.	Run the program and navigate to Train a model and press enter. 
44.	This option menu will appear.

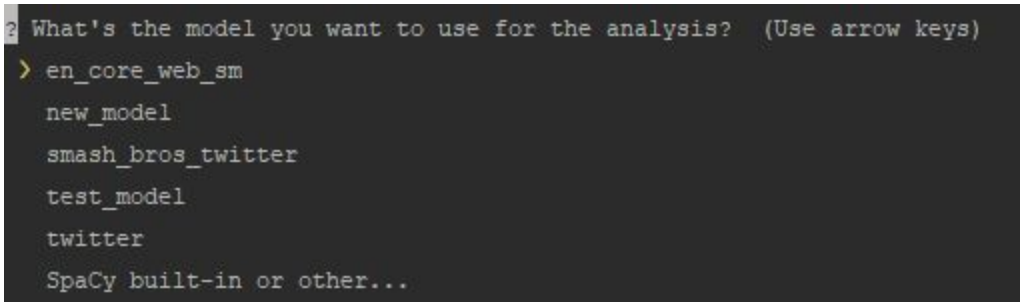
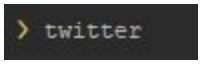
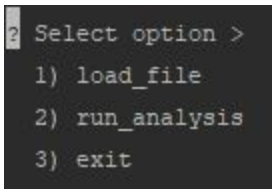
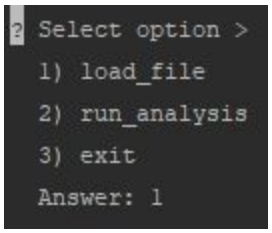
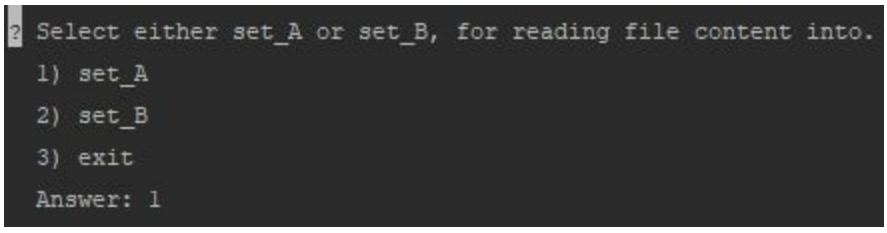
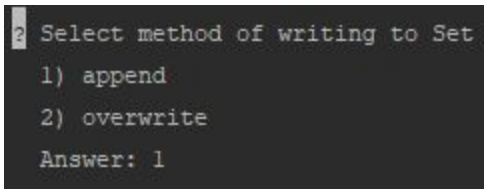
	
45.	To start a training session, select Start a training session and press enter. 
46.	This option menu will be displayed. 
47.	To create a new model, select SpaCy built-in or new… and press enter. 
48.	You will be asked to enter the name of the model you want to train and then press enter. 
49.	Choose the annotated training data file you wish to train the model against and press enter. I have decided to use smash_bros_twitter_annotated.json to train this new model. 
50.	Choose the annotation loader to load the file and press enter. 
51.	You will be asked where you want to save the trained model to. If you wish to keep a copy of the old model before the training, you will need to specify a new name, else just press enter. I will choose to just overwrite it instead. 

52.	Now you will be asked how my iteration the training should last, default value is given as 100 but you can choose any number of iterations you wish the model to train for and then press enter and wait for it to be done. 
53.	When it is done, a message similar to this will appear. 
54.	To train an existing model, choose the model you want to train and press enter. I have decided to train the twitter model again. 
55.	Follow step 49 to step 53 to train the model.
56.	Run the program and navigate to Train a model and press enter. 
57.	This option menu will appear. 
58.	Navigate to Test a model to test it and press enter. 
59.	A list of models will appear.

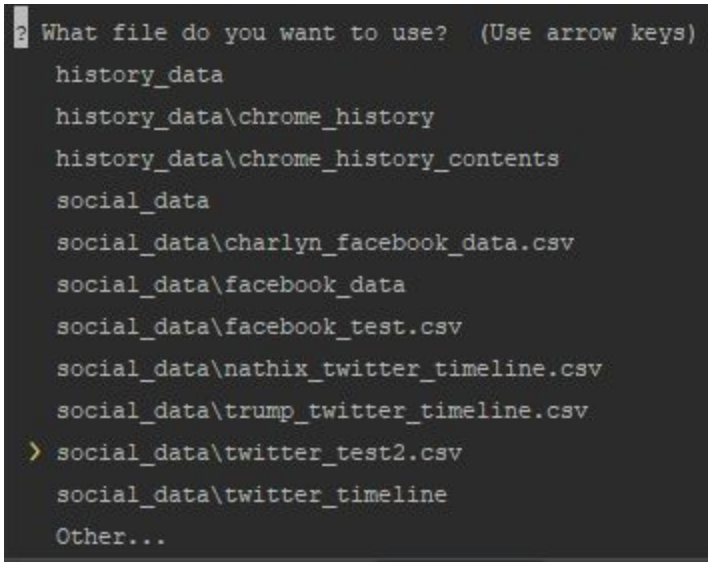
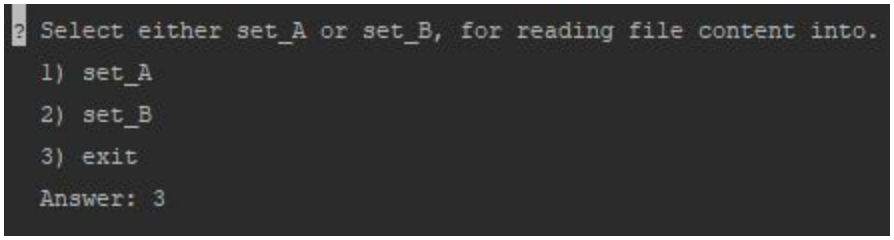
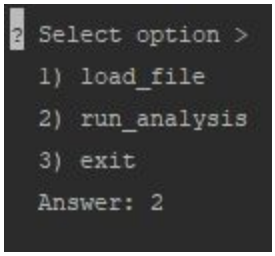
	
60.	Navigate to one of them and press enter. I have decided to test a Spacy built-in model. 
61.	You will be asked to enter the name of the model you want to test and then press enter. I have decided to test the en_core_web_md model. 
62.	Enter a text or sentence you wish to test the model against and press enter and wait. 
63.	The results will be shown. It shows that Google is an organization and Pixel 4 is a product. 

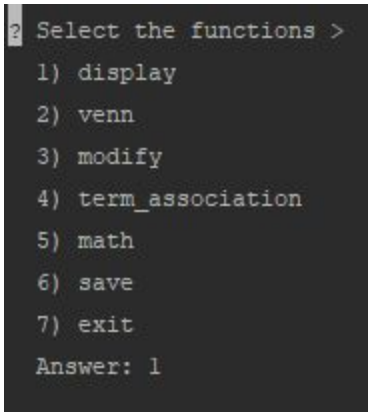
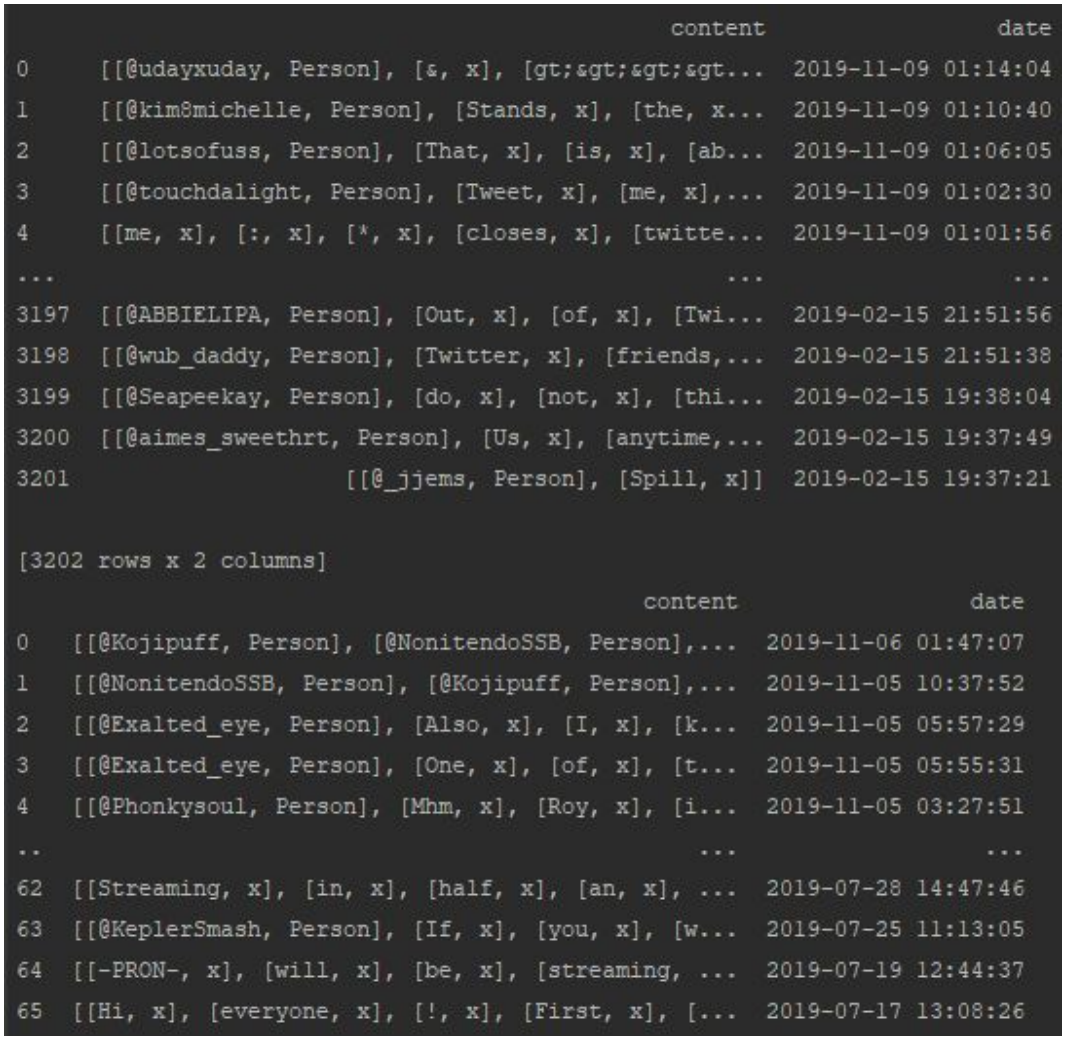
Analyse Data

Step No.	Instructions
63.	Run the program and select Analyse data and press enter. 
64.	The option menu will appear as shown below.

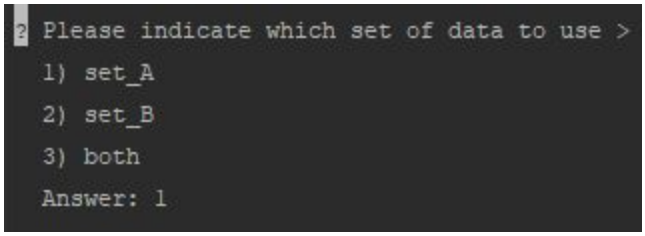
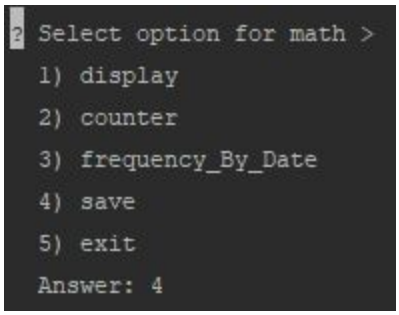
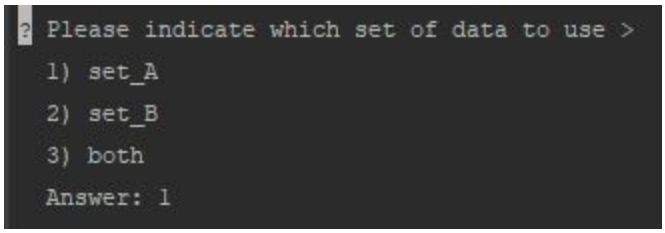
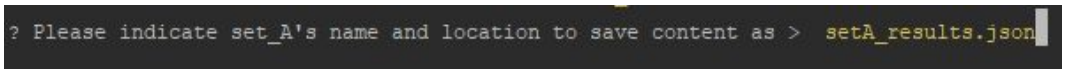
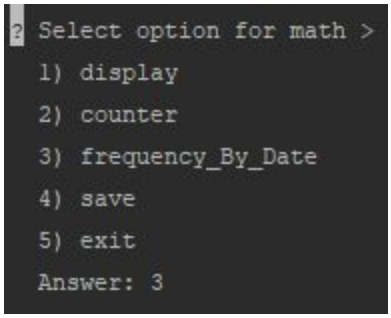
	
65.	Choose the model you want to use for the analysis and press enter. I have chosen to use the twitter model for the analysis. 
66.	An option menu will appear as shown below. 
67.	Enter option number 1 to load file and press enter. 
68.	Next choose a set for reading the file content into and press enter. For this I choose set A. 
69	Select the method of writing to the set chosen and press enter. I have chosen to append the data. 

70.	Choose a file you want to use and press enter and wait for it to finish. I have decided to use twitter_timeline `? What file do you want to use? (Use arrow keys)` `history_data` `history_data\chrome_history` `history_data\chrome_history_contents` `social_data` `social_data\charlyn_facebook_data.csv` `social_data\facebook_data` `social_data\facebook_test.csv` `social_data\nathix_twitter_timeline.csv` `social_data\trump_twitter_timeline.csv` `social_data\twitter_test2.csv` `› social_data\twitter_timeline`
71.	Choose a set to read the file contents into and press enter. Since I have chosen set_A previously, I will choose set_B this time. `? Select either set_A or set_B, for reading file content into.` `1) set_A` `2) set_B` `3) exit` `Answer: 2`
72.	You will be asked to select the method of writing to the set and press enter. I choose to append it. `? Select method of writing to Set` `1) append` `2) overwrite` `Answer: 1`
73.	Choose a file you want to use and press enter and wait for it to finish. I have decided to use twitter_test2.csv

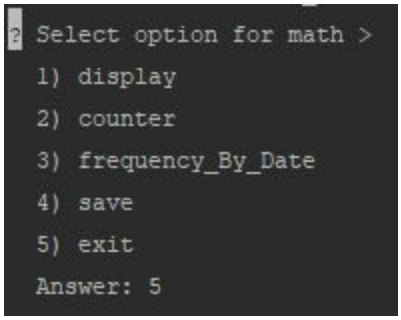
	
74.	Once you have loaded the files into both set A and set B, you may now exit by pressing 3 and press enter. 
75.	Now we can start to run analysis. Press 2 and then press enter. 
76.	To display, press 1 and then press enter.

	
77.	Something similar to this will appear. 
78.	Now press 5 and press enter to select the math option.

	`` ? Select the functions > 1) display 2) venn 3) modify 4) term_association 5) math 6) save 7) exit Answer: 5 ```
79.	The option menu will appear. ``` ? Select option for math > 1) display 2) counter 3) frequency_By_Date 4) save 5) exit ```
80.	To do the counter, press 2 and then press enter. ``` ? Select option for math > 1) display 2) counter 3) frequency_By_Date 4) save 5) exit Answer: 2 ```
81.	You will be asked if you want text or tag to be used and then press enter. I have decided to use tag here. ``` ? Please indicate either text or tag to use > 1) text 2) tag Answer: 2 ```
82.	Now you need to indicate which set of data you want to use and then press enter I have chosen to use set A.

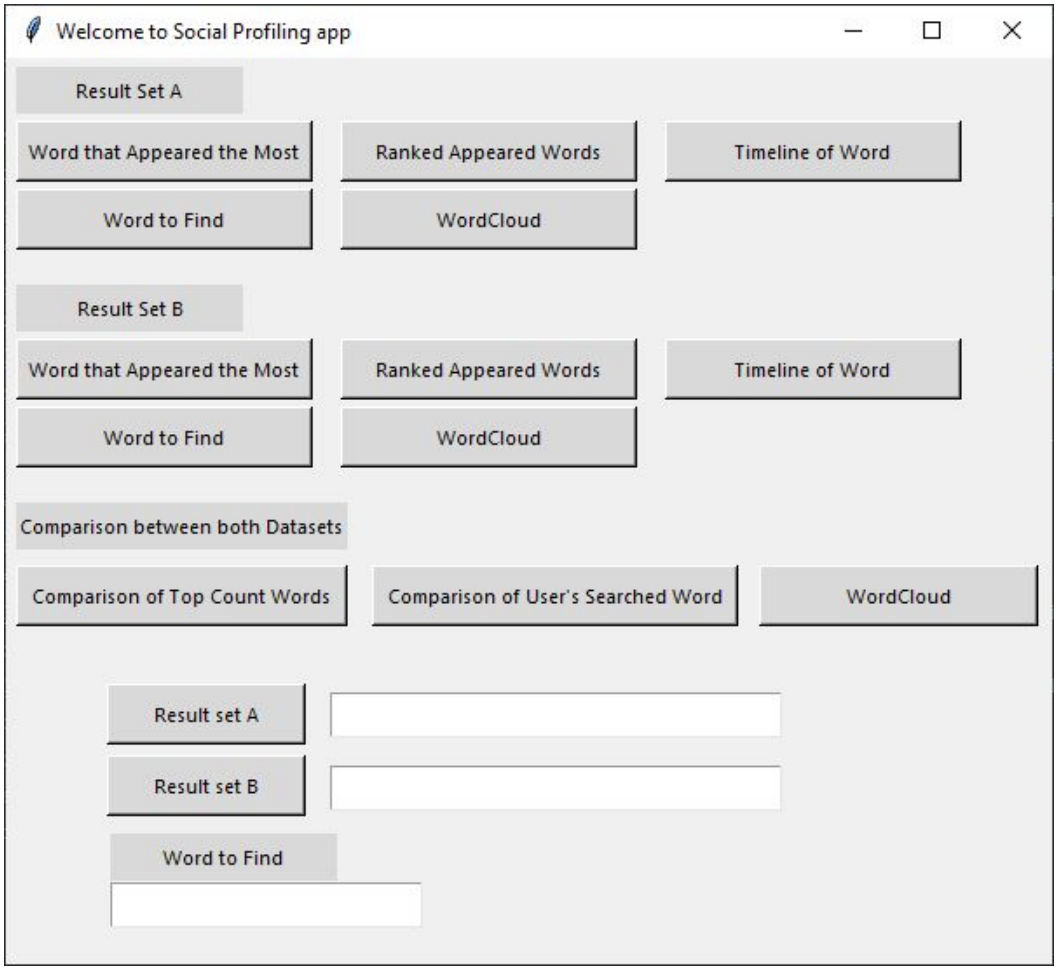
	
83.	Now you can choose to save your data by pressing 4 and then press enter. 
84.	Indicate which set of data to use and press enter. 
85.	You will be asked to enter the name of the file that you want to save the results as and then press enter to save it. 
86.	In the results folder, you can see that the file has been successfully saved. 
87.	You can also analyse it by the frequency which is by date by entering 3 and then press enter . 

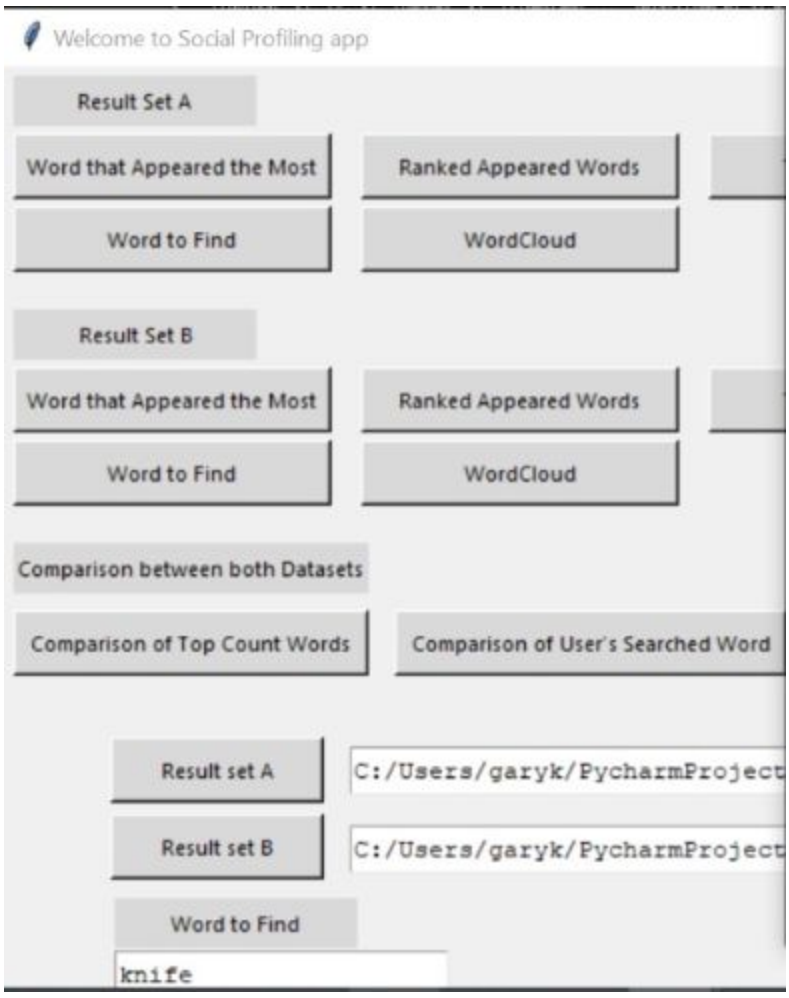
88.	You will be asked if you want text or tag to be used and then press enter. I have decided to use tag here. `? Please indicate either text or tag to use >` `1) text` `2) tag` `Answer: 2`
89.	You can choose which set of data you want to use or both but I have chosen to use set A here and then press enter. `? Please indicate which set of data to use >` `1) set_A` `2) set_B` `3) both` `Answer: 1`
90.	Here you will enter the word that you want to count the frequency. `? Please enter value to determine frequency by date >  Joker`
91.	Now you can save your data by entering 4 and then press enter. `? Select option for math >` `1) display` `2) counter` `3) frequency_By_Date` `4) save` `5) exit` `Answer: 4`
92.	Enter which set of data you want to use or both and press enter. `? Please indicate which set of data to use >` `1) set_A` `2) set_B` `3) both` `Answer: 1`
93.	Now enter the name of the file that you wish to save the data as and press enter. `? Please indicate set_A's name and location to save content as >  setA_frequency.json`
94.	Now if you go to the results folder, you can see that the file has been successfully saved there.

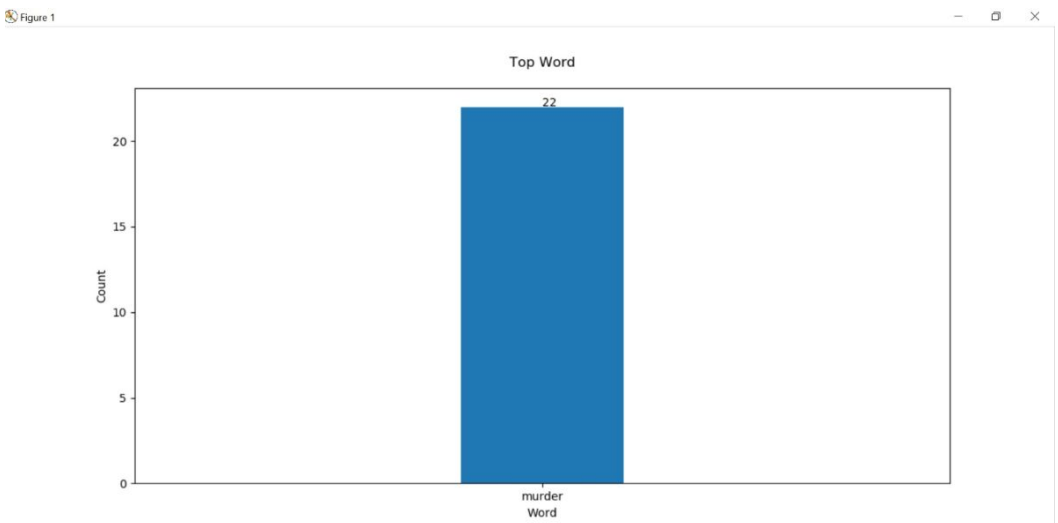
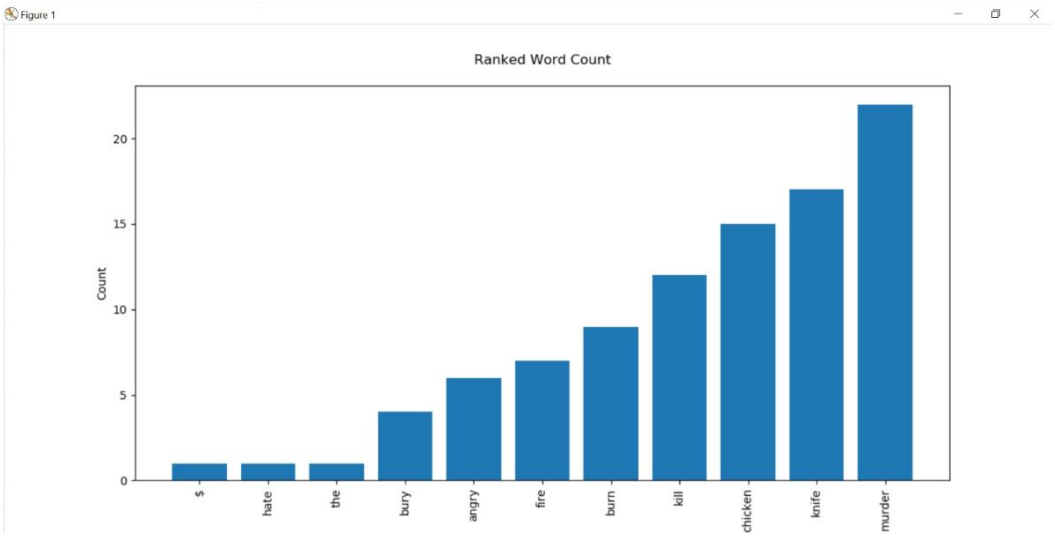
	
95.	If you are done analysing the data, you press 5 and then press enter to exit. Keep entering the exit option till the program ends. 

Show profiling results

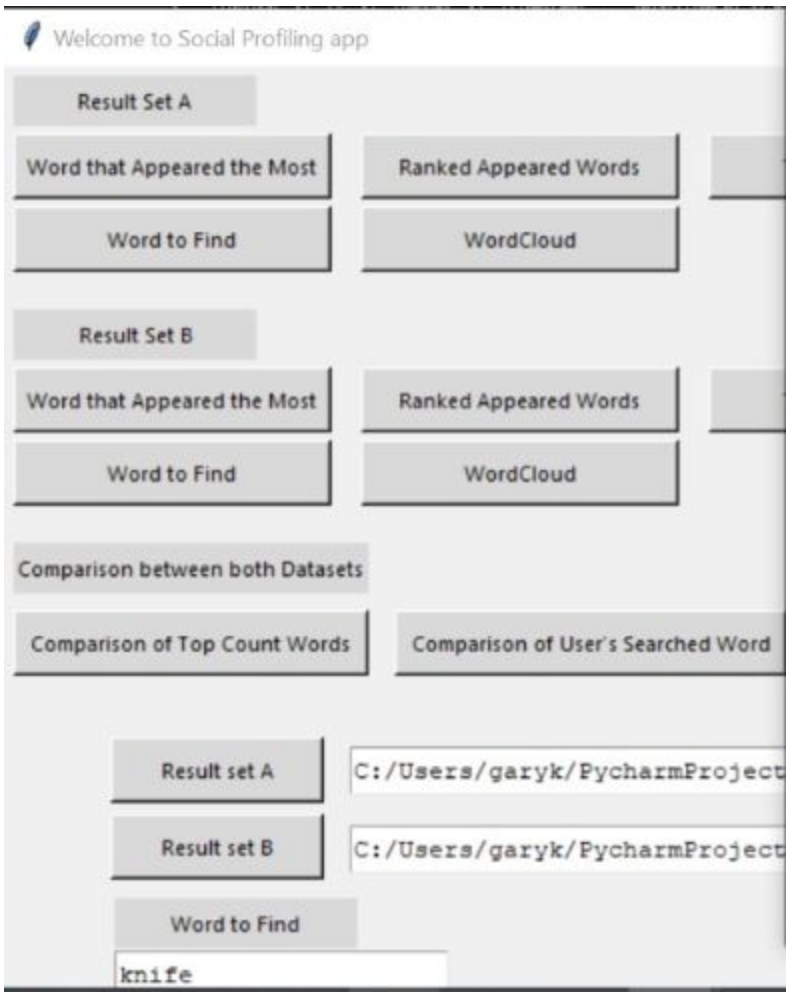
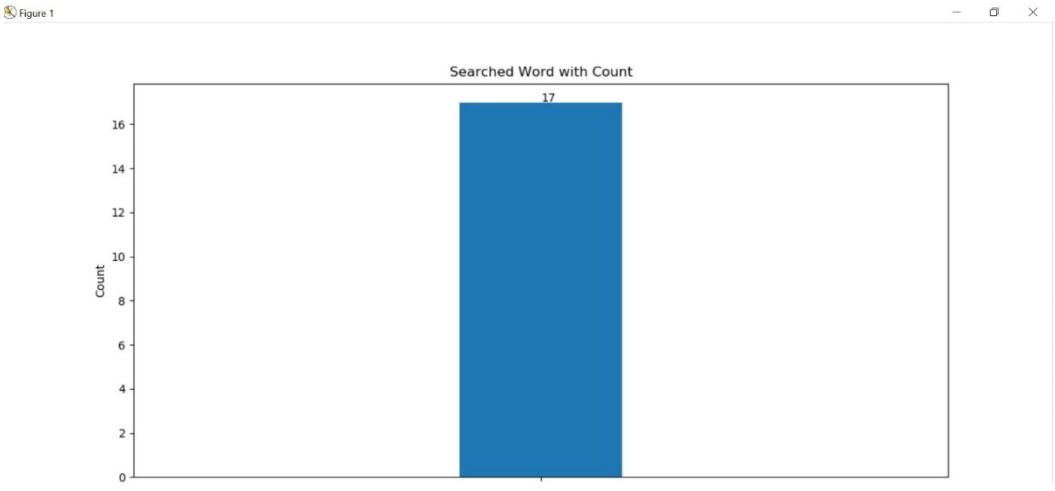
Step No.	Instructions
96.	Run the program and select Show profiling results and press enter. 
97.	This window will appear.

	
98.	User uploads the result set A and result set B. The extension of the file must be json.

	
99.	User can click on the button “Word that appeared the most” and is able to view the top word at the respective dataset.

	
100.	User can click on the button “ranked appeared words” and is able to see the total occurence of each word that appears in each respective dataset. 
101.	User can click on the button “Wordcloud” and is able to see the various words that appear in the respective dataset. The more the word appears in the dataset , the bigger the word will be.

	
102.	Below the “Word to find”, user can input any word on the text box and click on button “word to find”

	 
103.	User can click on the button “comparison of top count words” and it will display a bar graph that shows the top word count for each dataset.

	
104.	Below the “Word to find”, user can input any word on the text box and click on the button “comparison of user’s searched word”.

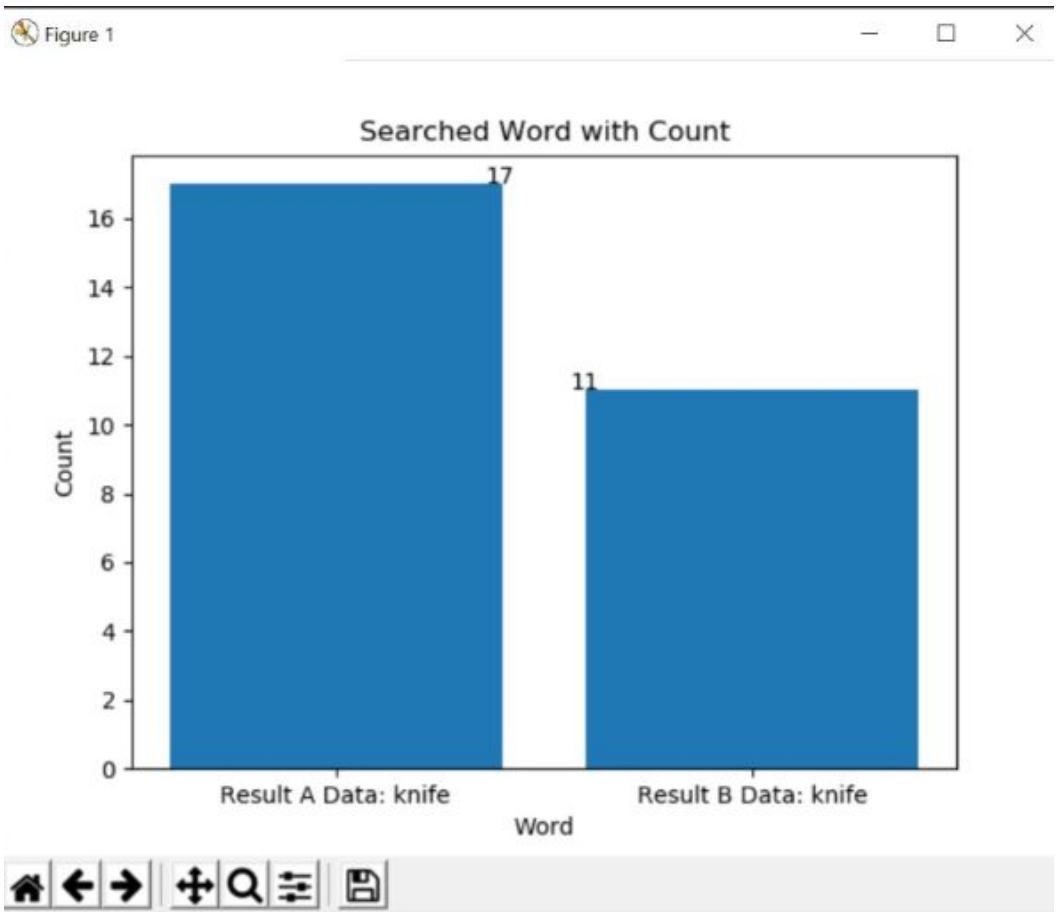
	
105.	User can click on the button “Word cloud” and it will display the result as below.The more the word count in the combined dataset , A and B , the bigger the word will be.


Figure 1
murder
burn
kill
serial
angry
sad
cook
poison
love
bury
hate
knife
fire
chicken
Francisco

Link for Video
https://www.youtube.com/watch?v=LbX5zVY_d_w