



# GMU Miners

By: Flavio Amurrio,  
Calvin Alvarez,  
Sergio Gutierrez



# Technologies Used

- Visual Studio Code

- Microsoft Visual Studios

- Qt Creator/Designer

- Sublime Text Editor

## Operating Systems Used

- Mac OS X

- Windows

- Tweepy

- Git/GitHub

## Languages:

- Python

- Java

- C++



# Our Approach

We split the the problem into 3 main sections:

- Web crawling (Sergio)
- Data processing(Flavio)
- Graphic Representation(Calvin)



# Web Crawling(Sergio)

- The Crawler was build with

  - PYTHON

  - TWEEPY

- The crawler takes an input user as a seed

- Search two levels down of friends

  - friends of friends of friends



# Issues encounter

-Python

Syntax

Indentation

-Tweepy

200 results within 15

Solution : `time.sleep(60)` // TOO LONG!!

```
tweepy.Cursor(api.friends, count = 20000)
```

tom\_wilso  
kheadman  
set([u' KremlinRussia\_E', u' kheadman', u' VRSoIoviev', u' MedvedevRussiaE', u' JustinWilliams', u' tom\_wilso', u' VineCreators', u' Bobina', u' DChernyshenko', u' McIlroyRory', u' IceHockeyCampCZ', u' Kr  
emlinRussia', u' beatsbydre', u' Tommy\_Chalk', u' TwitterVideo', u' PapaJohns\_DMV', u' Eminem', u' advorkovich', u' VorobievAndrey', u' CCMHockey'])  
'KremlinRussia\_E  
#Sochi: The President attended the Russia show at the 19 World Festival of Youth and Students ahead of its closing... <https://t.co/m6jIi2Wwsc>  
19th World Festival of Youth and Students: The President attended the Youth 2030: The Image of the Future session... <https://t.co/4LTNWhGcQJ>  
Telephone conversation with President of Turkey Recep Tayyip Erdogan <https://t.co/21L3mjBhk3>  
Economic Development Minister reported on the current economic situation in the country and macroeconomic trends... <https://t.co/Gu9MKnWhD9>  
Meeting with Head of Kabardino-Balkaria Yury Kokov: the socioeconomic situation in the republic... <https://t.co/qWjDdtXB5m>  
On October 24, Vladimir Putin will meet with President of Cyprus Nicos Anastasiades in Moscow <https://t.co/fMyjYl40or>  
#Sochi: Meeting of the @ValdaiClub <https://t.co/4N2zhdXH14> <https://t.co/QBSKrxDyHz>  
Telephone conversation with President of Kyrgyzstan: EAEU performance, Russian-Kyrgyz relations <https://t.co/3GKNRtxPav>  
Press statements following Russian-Croatian talks <https://t.co/vP6kYxSvE4> <https://t.co/wMxcZg37RV>  
On October 19, the President will address a meeting of the Valdai International Discussion Club <https://t.co/yGYsnePazm>  
Telephone conversation with Israeli Prime Minister Benjamin Netanyahu <https://t.co/QQ7NltgQB4>  
#Sochi: Talks with President of Croatia Kolinda Grabar-Kitarovic <https://t.co/aV1PMfrieP> <https://t.co/9vXRRyQdIM>  
Russian Eco Week has opened in Moscow. Greetings from the President: <https://t.co/2FKcsZb0LC>  
#Sochi: Meeting with Rosseti CEO Pavel Livinsky <https://t.co/F7wiKPoZJw> <https://t.co/W3e2E15vwl>  
October 18 #Sochi: Vladimir Putin will meet with Croatian President Kolinda Grabar-Kitarovic <https://t.co/vdoVQjXEIG>  
Today is opera singer and People's Artist of Russia Dmitri Hvorostovsky's birthday  
Vladimir Putin has met with Kursk Region Governor Alexander Mikhailov <https://t.co/nF1CVmEY0W> <https://t.co/BvvYTPruQK>  
Condolences to the President of Somalia on the tragic consequences of the terrorist attack in Mogadishu <https://t.co/XhktH68dt9>  
#Kremlin, October 25: Vladimir Putin will meet with Federal President of Germany Frank-Walter Steinmeier <https://t.co/wGY11C24RS>  
#Sochi: Meeting with participants of 19th World Festival of Youth and Students <https://t.co/CgX3ugmRTY> <https://t.co/gvrxIM6f2U>  
kheadman  
Not sure I'll ever get used to this view to end the evening (or at least I'm hoping I don't). #VegasMagic <https://t.co/WhCS7wVIZP>  
RT @ESPNStatsInfo: Golden Knights (6-1-0) beat Blues in OT. They are first franchise in NHL history to win six of first seven games in inau...  
RT @JesuitBaseball: Congrats to @McCullers43! What an unbelievable job! Good luck in the World Series! #TigerPride #OnceATigerAlwaysATig...  
RT @KevinWeekes: Tough news for @GoldenKnights losing Fleury and Subban to injuries.  
Oscar Dansk with the opportunity now - Welcome to the...  
RT @JesuitTampaFL: Congrats Lance McCullers Jr. '12 and the Astros on winning the ALCS w McCullers pitching 4 scoreless to close out the Ya...  
RT @wclewis\_uf: Houston needs this. And great to see a #JesuitTiger close it out #AMDG #Game7 <https://t.co/BPPa60KvNh>  
RT @TMobileArena: Nothing like the view from @HydeTmobile to cheer on the @GoldenKnights! <https://t.co/mUb59EtERd>  
#Bucketlist <https://t.co/8rjiivpGHW>  
@tylerwong\_5 These are always welcome light. Thanks always for sharing.  
RT @NHlonNBCSports: Rudy (YES, THAT RUDY) is a @GoldenKnights fan! He's signing his book at the bookstore before No. 11 takes on No. 13 Not...  
If you haven't listened in a while, give @JohnMayer #3by5 a listen @ageswell  
@DisneylandToday For how long?  
So many special things about a @JesuitTampaFL education <https://t.co/HSBgL4dod1>  
Loving the gray #Rocky ensemble (with the @Penn top of course!) <https://t.co/DzaLjuEVmx>  
Perhaps a bit late to the party, but @netflix #Mindhunters #vow  
RT @SI\_NHL: Going to a @GoldenKnights game this year? The exec chef @TMobileArena gave @MitchGoldich a rundown of the offerings <https://t.co/...>  
RT @ngreenberg: 🍷 <https://t.co/fdBU16maWr>  
RT @myqcmallards: Mallards, Quad City International Airport to give away three Vegas vacations:  
<https://t.co/URbrPdTeSj> <https://t.co/Y7zV8k...>  
RT @TMobileArena: Headed to a @GoldenKnights tonight or soon? Download our App now for the latest updates and information! <https://t.co/v5N...>  
Remember it like it was yesterday!! And @backstrom19 pitch was right on the mark. <https://t.co/WcTFqpFSy7>  
VRSoIoviev  
@Kiриллggg38Rus Это про Купертино  
@Step\_F0X @Alexey\_Pushkov Аж кушать не можете ?  
@Step\_F0X @Alexey\_Pushkov И ты это пишешь из Питера. Ты в Воронеже сам был?  
RT @rianru: #РИАрхив  
22 октября 1929 года родился Лев Яшин – легендарный советский вратарь <https://t.co/G28im4zeuI>



# Data Processing(Flavio)

- Cleandata
  - Translate(Google Translate)
  - Remove StopWords(Natural Language Toolkit)
  - Stem the words to its root(NLTK)
  - Remove special characters
- Vectorize Data
  - Allows for easier search if adding more keywords
  - Used term frequency as metric
  - If enough data is gather, we could implement a classification model
- Pitfall
  - Couldn't use synonyms(libraries gave use no related word)

# DEMO

## GMU MINERS CRAWLER PROGRAM!

Search Program

User

ovi8

Keyword

capitals hockey puck

Search

User Tree

▼ Root\_User

▼ FriendLevel1

KremlinRussia\_E

kheadman

VRSoloviev

MedvedevRussiaE

JustinWilliams

tom\_wilso

Bobina

DChernyshenko

McIlroyRory

IceHockeyCampCZ

KremlinRussia

beatsbydre

Tommy\_Chalk

TwitterVideo

PapaJohns\_DMV

Eminem

User List Sorted by Strength

9 KremlinRussia\_E

7 kheadman

4 VRSoloviev

3 JustinWilliams

1 MedvedevRussiaE

Tweets/Comments

#Sochi: The President attended the Russia show at the 19 World Festival of Youth and Students ahead of its closingâ€¦ <https://t.co/m6jII2Wwsc>

19th World Festival of Youth and Students: The President attended the Youth 2030: The Image of the Future sessionâ€¦ <https://t.co/4lTNWhGcQJ>

Telephone conversation with President of Turkey Recep Tayyip Erdogan <https://t.co/2il3mjBhk3>

Economic Development Minister reported on the current economic situation in the country and macroeconomic trendsâ€¦ <https://t.co/Gu9MKnWhD9>

Meeting with Head of Kabardino-Balkaria Yury Kokov: the socioeconomic situation in the republicâ€¦ <https://t.co/qWjdDxbB5m>

On October 24, Vladimir Putin will meet with President of Cyprus Nicos Anastasiades in Moscow <https://t.co/fMvJYl4Qor>





# What we learned

- Python
- TWEETPY
- Communication skills
- GUI development
- Agile Methods



# What would we do different

- Use Electron or other more front end environment
- Use Twitter Streaming API in order to gather more data
- Mine the data and cluster it depending on what clusters there may be
- Implement sentimental analysis of the tweets
- Take in the account the content of the links within the tweets

