Tunisia Twitter Data

Ryan Gniadek, Victoria Hardy, Sraavya Gudavalli, Steven Ruckert

CS 4624: Multimedia, Hypertext, and Information Access Capstone
Dr. Edward Fox

Virginia Tech, Blacksburg VA 24061
May 11, 2023
Outline

- Motivation
- Project Plan
- OKRs and Approach
- Data Filtering and Sentiment Analysis
- Visualizations
- Acknowledgements
- References
Motivation

After the 2011 Arab Spring, Tunisia underwent a democratic revolution.

Political Twitter usage increased as the population exercised freedom of expression.

If we believe in democracy and human rights, we should be loudly decrying the current assault on democracy in Tunisia—the heart of the Arab Spring.

If Tunisia’s president does not reverse course, we should suspend all security aid, as is required by law.
Project Plan

- Subtasks based on each team member’s skill set/background
- Collaboration on all aspects of the project
- One team member heads one aspect

Ryan Gniadek
Back-End & Data Management

Sraavya Gudavalli
Documentation & Visualizations

Victoria Hardy
Documentation & Visualizations

Steven Ruckert
Back-End & Data Management
OKRs

● Archive of Previous Twitter Data
  ○ Consolidate data into one library
  ○ Coordination with previous researchers/contributors

● Import of Live Twitter Data *
  ○ Utilize SFM framework for incoming data
  ○ Standardize format of imported twitter data

● Improvement of Data Filtering Methods
  ○ Reduce collection of short, uninformative tweets
  ○ Reduce collection of spam Twitter data
OKRs

● Analysis of Tunisia Twitter Data
  ○ Produce new data trends
  ○ Add trends to existing research

● Visualization of Analysis
  ○ Single platform to view data trends
  ○ Compare the old and new data

● Development and Demonstration of Professional Work Skills
  ○ High attendance
  ○ Fair work distribution
OKR Overview: Progress

- Archive of Previous Twitter Data: 100%
- Improvement of Data Filtering Methods: 100%
- Analysis of Tunisia Twitter Data: 100%
- Real-Time Import of New Twitter Data: 100%
- Visualization of Analysis: 100%
OKR Project Approach

Collect Twitter Data (Archived and Real-Time)
Social Feed Manager
February 17, 2023

Filter Twitter Data
Sentiment, Keyword, and Language
March 3, 2023

Analyze Twitter Data
Political Events and Protests
March 24, 2023

Visualize Data
Trends in keyword frequency and sentiment
April 12, 2023
Approach

- Using the Social Feed Manager system, Twitter data was collected
  - Due to uncertainty around Twitter API access, ongoing collection was not possible. Tweets were added through February 2023
- Format of existing tweet archive needs to be standardized and consolidated
- Language and sentiment analysis determined for tweet contents
- Interactive visualizations of Twitter data using modern web technologies
If we believe in democracy and human rights, we should be loudly decrying the current assault on democracy in Tunisia—the heart of the Arab Spring.

If Tunisia’s president does not reverse course, we should suspend all security aid, as is required by law.
Data Filtering and Sentiment Analysis

**Preprocessing**
- sentiment.py
- JSONL format

**Filtering**
- filtering.py

**Format for Visualizations (Postprocessing)**
- create_tsv.py
- classify_tweet.py
- client specified TSV format

**JSONL format**
- group_by_date.py
FOR each line of tweets in JSONL file

TRY:
   FOR each tweet in the line
      Get the language of the tweet as 'lang'
      IF: Lang is 'en':
         TRY:
            Calculate sentiment of tweet's text & add to 'sentiment' attribute
            Write tweet as JSON string to output_filename
         EXCEPT: Print "Skipped tweet - No Sentiment Score"
      ELSE: Print "Skipped tweet - Not in English"
   EXCEPT: Print "Skipped line"
Visualizations

- Created for tweet data filtered by keyword
  - Keyword-in-tweet volume
  - Tweet sentiment trend
- Analysed with respect to Tunisian political events between 2019-2023
“Tnelec2019” Keyword Visualization

September 15, 2019
Presidential elections

October 6, 2019
Parliamentary elections

October 13, 2019
Presidential run-off

“Tunisiavotes”, “tunisiadecides”, & “tnelec2019” have the most tweets in 2019.
July 25, 2021
President Saied coup

July 25, 2022
Tunisians go to polls to vote on a referendum on a new constitution.
July 25, 2021
President Saied coup
July 25, 2021
President Saied coup
October 15, 2022
Islamist Ennahda and Free Constitutional Parties held parallel rallies against Saied

December 17, 2022
Lower Parliament elections are held
Visualization Analysis Takeaways

- Tweet sentiment was always more negative than positive
- Spikes frequently occurred around major elections and protest dates
  - September 2019: Presidential elections
  - October 2019: Parliamentary elections and Saied’s presidential run-offs
  - July 2021: President Saied’s coup, dismissal of parliament, enforced curfew
  - July 2022: Tunisians vote for a referendum
  - October 2022: Major anti-Saied protests
  - December 2022: Lower Parliament elections
Project Takeaways

**Challenges**
- Bias from English-only tweets

**Successes**
- Use of sentiment analysis
- Learned Tableau for visualizations
- Added as authors to research paper

**Future Work**
- Create sentiment algorithm for non-English tweets
- Create visualizations for non-English tweets
We’d like to thank Dr. Fox and the CS 4624 Teaching Team for supporting us throughout our Capstone project.

We’d also like to thank our clients Dr. Kavanaugh, Dr. Sheets, the Data Collection team, and VT University Libraries for advising our Tunisia Twitter Data project.
References


