Big Data Analytics on Social Media Data: A Social Media Data Acquisition Tool

Faculty Advisors: **Dr. Cyril S. Ku (Computer Science), Director of Data Science Research Lab, WPU**
Dr. Jin-A Choi (Communication), Co-Director of Data Science Research Lab, WPU

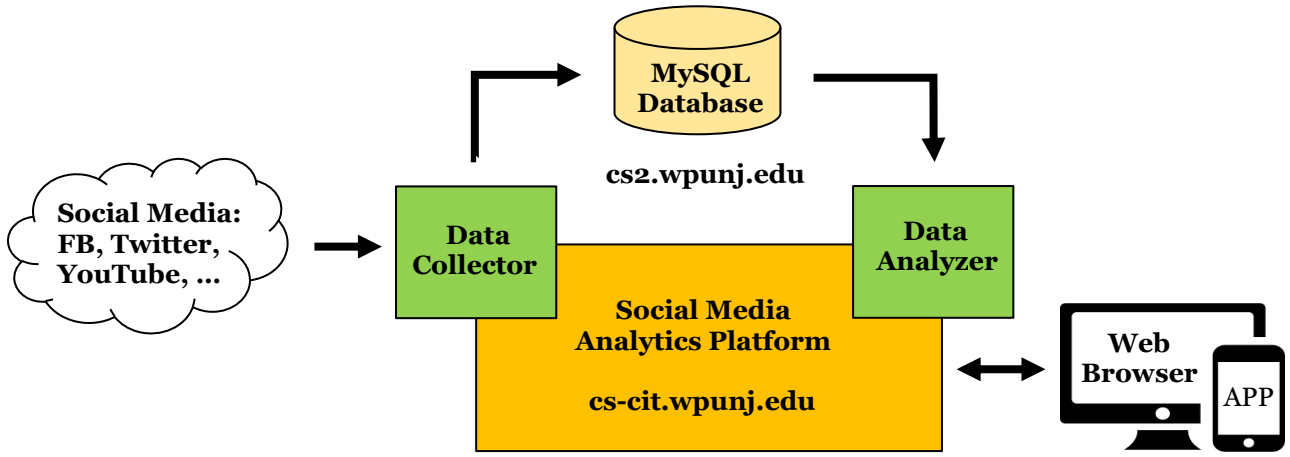
Research Assistant: **Anna Renee Chieco (Computer Science), MaCS Scholar, WPU**

Research Period: **Summer (June-August) 2021**

Project Background:

Dr. Cyril S. Ku is the Director of the Data Science Research Lab in the Department of Computer Science at William Paterson University. He has an on-going research collaboration with Dr. Jin-A Choi in the Department of Communication. She is the Co-Director of the Data Science Research Lab and her expertise is in big data analytics on textual data. Currently Dr. Choi and Dr. Ku are investigating various text mining techniques such as opinion/sentiment analysis, social networking, information or misinformation propagation; and other interactions, relationships, as well as issues in social, cultural, and health fields within and across different social media platforms.

The data science research team has established a hardware and software environment for this big data analytics research on social media data. The architecture of this environment is shown below. This project is divided into three phases: Data Acquisition (Phase 1), Data Analysis (Phase 2), and Social Media Analytics Platform construction (Phase 3). This summer research only addressed Phase 1 of the project.



Summer Research Objective:

Construct the Data Collector – a data acquisition tool to collect relevant data in real-time, then clean, transform, normalize, integrate, and organize them into the MySQL database.

Research Progress:

Anna has completed the following tasks for Phase 1 of the project during the summer:

- Set up the programming environment, installed Anaconda Navigator on her PC for the design, implementation, and testing of a prototype system;
- Studied and learned Python (Jupyter Notebook), using the Python book we ordered through the grant;
- Gathered APIs information for five social media platforms (Instagram, YouTube, Facebook, Twitter, TikTok);
- Investigated the manual process of collecting data (e.g., YouTube transcripts), in preparation to convert the manual process into an automated process.

Conclusion and Further Research:

The Data Collector is only part of the Phase 1 of the project. Anna will continue to work with Dr. Ku and Dr. Choi on the Data Collector in the coming academic year. She planned to have a prototype of the Data Collector and finish Phase 1 of the project in Spring 2022. She intended to present her research internally in the annual Exploration event at WPU and possibly present at an external conference.

This research was supported by the National Science Foundation under NSF Grant # 2028011.

FUNCTIONAL PROGRAMMING IN PYTHON

August 12, 2021

By: Kwasi Asare-Bediako

Advisor: Dr. David Nacin.

I found the summer research program quite enlightening because Dr. Nacin introduced me to a very powerful style of programming that programmers use to reduce bugs in their code while making it more readable and reusable. This style of programming is called functional programming, where programmers try to bind everything in pure mathematical function style. Another style of programming in python is the object-oriented style of programming where programmers write a step-by-step code to achieve a particular goal. Object-oriented is an imperative style of programing and variables are treated as containers where values are stored in them. The problem that arises is that the variables in these programs keep changing and as the program gets large, bugs become hard to find. This is what functional programming seeks to avoid because in the python language, one of the core concepts in functional programming is **immutability**. In other words, there is no mutation going on in a code written in a functional style. All variables are defined and treated as constants. Functional programming is a declarative style of programming.

Python as a language is a very powerful tool which has these built-in functions to support functional programming. Map() allows us to apply a function to all elements of a list. Reduce allows us to repeatedly apply a process to the elements of a list two at a time. It starts with the first element in a list and builds on its initial value someway until after we've processed all our elements and we end up with a result. Lambda expressions are used to create one-line anonymous functions. Lambdas make codes shorter and more readable. Another built in function is called filter and it is used when a programmer wants to find all the elements in a list or some other type of iterable that fits a particular criterion.

Also, in python, we have list comprehensions that can do essentially the same things that these built-in support functions do. Assuming you have a list (stored in a variable num_lists) of random integers and you want to get from this list only the evens, the list comprehension [x for x in num_lists if x % 2 == 0] gives you that even numbers. This exactly what the built-in filter function does. We could also code list comprehensions to map elements in a list to another lift of a particular criterion.

There are so many advantages as to why programmers should learn how to code in a functional style. It helps programmers to debug codes easily while improving productivity of the programmer. It also helps programmers solve problems effectively in a simpler way. I am so glad I was introduced to this powerful tool in python, and I hope to learn more in future.

This research was supported by the National Science Foundation under NSF Grant # 2028011.

WPUNJ MaCS Scholars Program

Summer Research Summary – Summer 2021

Title: Re-envisioning College Algebra with an Updated Perspective

Authors: Emily Hoagland, Paul von Dohlen

The issues with the College Algebra course have long been documented, with accounts of struggles with proper formation and administration of the course dating back to at least 1910. Recently, in the MAA publication, "A Common Vision for Undergraduate Mathematical Sciences Programs in 2025", a common theme was emphasized amongst reports from five major professional associations. The common theme was that "the status quo is unacceptable." Some of the main issues with the course deal with the course being overly focused on methods, that conceptual understanding is not emphasized, a lack of applications/realism, and a lack of focus on vocabulary. We propose changes to the approach and philosophy for a College Algebra course.

The suggested change in approach shifts the emphasis from procedural manipulations and equation-solving techniques to strengthening the student's command of mathematical vocabulary, building their realization of the connected nature of the topics, and exploring the proper use of technology. The instructor will prioritize these new focal points through guided notes packets which have been created with this change in emphasis in mind. Common themes in the notes packets will include providing multiple approaches/perspectives, promoting mathematical vocabulary and establishing connections amongst the topics.

Reinforcement, through homework, will also be tailored to this approach establishing cohesion between the classroom emphasis and the homework focus. The emphasis will also be extended to assessment, as assessment tools will reflect the revised focus. Furthermore, assessments will clearly distinguish the required procedural skills from the conceptual understanding development. Students will be asked to demonstrate certain skills and basic understanding in a timed/proctored setting while tasked with showing knowledge of the proper use of technology and deeper understanding with untimed/unproctored assessments.

We understand that there are numerous challenges in taking this revised approach but, as has been clearly established, the need for change prevails. A bold new approach is demanded by the shortcomings of past attempts as well as the everchanging definition of what are required skills. We need to change the focus and motivate more change.

This material is based upon work supported by the National Science Foundation under NSF Grant # 2028011.