CCSC*NE* '22

**The Twenty-Sixth Annual
Consortium for Computing Sciences in Colleges
Northeastern Conference**

April 1 – 2, 2022

**Hosted by Pace University
Pleasantville, New York**

Undergraduate Student Poster Abstracts

CCSC

CONSORTIUM FOR
COMPUTING SCIENCES
IN COLLEGES

ccsc.org

Contents

Once Upon a Data Visualization: Visual Datasets for SimpleChartsRI

Sean Khang
Computer Science
Rhode Island College, Providence, RI

Faculty Advisor
Sally Hamouda
Rhode Island College, Providence, RI

We present our work on our data visualization web tool, SimpleChartsRI, which is an HTML, CSS, and JavaScript written tool that is free to all users. This tool creates visualizations from CSV files, created data through our embedded table, or through text in three simple steps. The tool was created in 2020 and it was presented and written about several times before our work was done on it. We continued to expand on our tool by adding additional features that we thought would benefit our targeted users which are high school STEM teachers and students. What we worked on was the new sample selection page which includes 60 samples, the metadata, CSV file, and the XLSX file of the data. Other tools lack samples, include very complicated samples, or have samples that aren't relatable to the user. Our tool has both simple and complex samples that have been hand-picked for our alias, RI residents. These samples, which are mostly based on data of RI or New England, have already been cleaned up and can be accessed with a simple click. They will not only help teachers create quick visualizations in the classroom but also resonate with the students. By doing so, students would be more inapt to paying attention and being engaged. They can also learn to create narratives by picking apart data, choosing which data are displayed, and creating their own data visualization story. High school STEM teachers who have used the new version of our tool admit that it's a simple tool that is both easy to use and helpful to use in the classroom.

Picklr – A Job Scraping tool for Data Analytics of Programming Related Jobs

Michael Bonilla
Computer Science
Eastern Connecticut State University, Willimantic, CT

Faculty Advisor
Garret Dancik
Eastern Connecticut State University, Willimantic, CT

According to the U.S. Bureau of Labor Statistics, job demand for developers and programmers is expected to grow, on average, by 20% from 2020 to 2030. Moreover, a job search of tech jobs on Indeed.com, an online job search and talent recruitment tool, reveals that there are several hundreds of developer positions available, each requiring knowledge of varying programming languages and tools. The returned job results provide informative data, such as what specific technologies employers are looking for. A tool that automates the data-harvesting and analysis of the search results would prove useful for job seekers, educators, and others. Web scrapers are tools that enable the automation of processing and analyzing website data, to the specifications of a programmer. In this research, we describe the development of a web scraper for data mining and analytics against job result data from a commonly used job search website. This research is aimed at providing key metrics regarding modern job requirements for developers, for interested Computer Science or other faculty that teach programming to students, and for data scientists who would like to compare key metrics across different job search queries. Such key metrics include, but are not limited to, a count of programming languages, developer frameworks and libraries, and developer tools that appear within a specified set of job postings. The tool is available as an online web-app that takes a job search query that is specified by user input and is optionally filtered by a geographical location. The web-app extracts key metrics and efficiently displays the results from the data analysis via bar graphs and pertinent diagrams that intuitively summarize the results. The web-app is available online via the following URL address:
https://www.michaelbonilla.dev/scraper/

Interactive Linux BASH Script for the Aircraft Power and Thermal Toolkit (APTT) Sensitivity Analysis Tool: The BASH SA Tool

Tye Leckinger
Department of Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisors
E. K. Iskrenova-Ekiert
SUNY Brockport, Brockport, NY

Philip Abolmoali, Soumya S. Patnaik
USAF Research Laboratory, Wright-Patterson AFB, OH

As a student at SUNY Brockport, majoring in Computer Science, I will be presenting my experience working in the Department of Defense High-Performance Computing (HPC) Internship Program. APTT supports the development of digital representations and virtual prototyping and evaluation of aircraft Power & Thermal (P&T) subsystem performance through multi-fidelity numerical analysis, supplementary data processing and simulation tools, flexible co-simulation environments, and advanced computing. APTT Sensitivity Analysis (SA) Tool is a MATLAB GUI-based tool, part of APTT, providing an interface between Dakota and P&T system models to efficiently prepare, carry out, and post-process SA simulations. Real-world large-scale simulations are not feasible on consumer machines and necessitate the use of HPC resources. Therefore, my task was to develop a stand-alone interactive BASH representation of the SA Tool, the BASH SA Tool, capable of running in an HPC environment. The major achievements of the project are as follows. The BASH SA tool has been successfully implemented, tested, and documented. This user-friendly tool provides multiple modes of operation for different levels of user proficiency. A new feature, extensive error-handling, has been implemented and tested. The user is guided through the process of setting up and executing the Dakota SA study through command-line input. For analysis and visualization of the simulation results, we implemented post-processing by making extensive use of Awk, Bash, MATLAB, Perl,

and Python scripting. The implementation of the interactive BASH SA Tool allows large-scale studies to be carried out within feasible timescales while making the tool more user-friendly and accessible, thus expanding the tool's user-base by allowing non-expert Linux users access to HPC resources. While working on this project, I learned the development process of applications, advanced features of BASH shell scripting, working with HPC resources, and how to effectively collaborate in a multidisciplinary team of scientists and engineers.

Editing DEM Files to Represent Overhanging Structures and Arches

Hannah Lord
Computer Science
Wheaton College, Norton, MA

Faculty Advisor
Michael Gousie
Wheaton College, Norton, MA

GIS platforms such as Google Earth rely on mass amounts of elevation data to render natural terrain. One type of elevation data file is a Digital Elevation Model, or DEM, which can represent the Earth's terrain in a raster format. These ASCII text files contain a grid of elevation values that map onto a bare-earth surface at specified intervals. Currently, these files are unequipped to represent overhanging structures, such as natural arches or bridges. The goal of this research was to edit these files to allow for arch functionality and develop an implementation, using the OpenGL C++ library, to successfully render the structure.

The resulting edited DEM file has an additional "multi" line in the header. The first value corresponds to the number of rows containing the elevations of the top of the arch structure, and the second is the row location within the file where the structure is first detected. The elevations of the inner top of the opening followed by the elevations of the ground beneath the arch in a specified number of rows are included at the end of the file.

The rendering program separately triangulates the top, inside, and ground portions of the structure. The Y values of the triangulation points are based on the point's position relative to the arch, and allow for the accommodation of the opening. The current software and DEM file format can successfully support structures that fall directly on a horizontal axis. Further research will explore ways of representing more natural occurrences.

The resolution of the original terrain was enhanced by interpolating three additional elevations between each real point using a Catmull Rom

spline. This allowed for smoother transitions between the ground and the arch structure. Additional interpolation was used to create a more cylindrical lintel of the arch.

Experiences in Developing a Customized Software Application using Agile Methodologies with Extensive Customer Involvement

Kyle D. Adams, Matthew E. Morgan
SUNY Brockport, Brockport, NY

Faculty Advisor
Sandeep Mitra
SUNY Brockport, Brockport, NY

We developed a software application that would keep track of student performance data related to institutional student learning outcomes (ISLO) assessment required by the Middle States accreditation process. The app is intended to be used by personnel in the ISLO assessment committee. To accomplish our goals with the resources and time frame available for development (i.e., one academic year) we decided to use an Agile approach, whose hallmarks include having a continuously working prototype of the system, and extensive customer interaction to ensure appropriateness and usability of the desired features. At the start of the project, we interacted extensively with the customer in order to ensure that we captured the initial set of requirements precisely and unambiguously. Thereafter, we discovered that although creating and documenting design models are not an integral part of an Agile approach, our unfamiliarity with the "business domain" of the desired app required us to create and document models of the back-end database and the system itself. During this period, the (non-technical) customer was not involved. The next customer interaction took place when we presented mockups of, and the workflow among, the GUI screens to them. This proved to be a valuable exercise as customer feedback received on these mockups enabled us to develop the initial prototype to the customer's satisfaction. Following this, we took an Agile approach to adding new features to the system, and demonstrated the functionality to the customer at regular intervals. Through this approach, we found that while we had to carefully guard against "scope creep", we were able to either accommodate the customer's requests, or communicate with the customer about alternatives, such that the product being developed

continues to satisfy the customer. We hope to complete the development of the full first version by the end of the academic year.

Learning ARM Assembly Through a Binary Game and *Dive Into Systems*

D'Angelo M. Gourdine
Department of Electrical Engineering and Computer Science
United States Military Academy, West Point, NY

Faculty Advisor
Suzanne J. Matthews
United States Military Academy, West Point, NY

Motivation and Background: Learning reverse engineering gives students a deeper understanding of how their programs work at a lower level. The use of "binary games" to teach reverse engineering is not a new concept, with one of the earliest examples being the "Binary Bomb", developed at Carnegie Mellon University. This poster explores the effectivness of teaching the ARM assembly architecture on a Raspberry Pi by having students reverse engineer a novel "binary game", guided by the textbook *Dive into Systems*. Unlike previous work, our project is open-source and contain hints that directs students to areas of *Dive into Systems* when they get stuck.

Methods: Students involved in the study played the game in the context of a 2-hour lab. These students had prior knowledge of x64 assembly, but no prior knowledge of ARM assembly. During the lab, students were provided a Rasbperry Pi with the game pre-loaded, and were briefed on how to access and start the game, and basic ARM instructions.

Results and Current Work: A Likert-scale pre-survey and post-survey were used to measure student confidence and motivation on reverse engineering an unfamiliar ISA and ARM assembly. Our results suggest that using the binary game increased student confidence learning an unfamiliar assembly language (2.8 average on pre-survey to 4.0 on post-survey) and reverse engineering programs in ARM assembly (2.2 to 3.56). Furthermore, student motivation to learn reverse engineering increased (3.9 to 4.14). Work to enhance user experience is continuing this semester, along with collecting additional assessment data.

Are COVID-19 and H1N1-09 Pandemics covered differently?

Nathaniel Lichauco, Ziran Fei, Rachael Tovar, Kimberly Ekstrand
Computer Science
Wheaton College, Norton, MA

Faculty Advisor
Zhou Tong
Wheaton College, Norton, MA

The COVID-19 pandemic continues to wreak havoc on our society, affecting the health of billions of people around the globe. Since the beginning of the pandemic in 2019, news outlets have been continuously reporting statistics of hospitalizations and fatalities. The media coverage of the pandemic allows people to remain up to date on the pandemic, circulating through people's feeds and helping shape our conceptualization of the pandemic. In this research, we analyze the news coverage of the ongoing COVID-19 pandemic and the H1N1 pandemic from 2009 to 2010 (H1N1-09). By studying the public sentiments through thousands of news articles covering these two pandemics, we examine whether these two health crises are covered differently two decades apart.

We collected approximately twenty thousand news articles from The New York Times (NYT) and The Guardian (GRD). Then, we identified seven basic tones of the news articles using the IBM Tone analyzer. The results indicate that *Sad*, *Analytical,* and *Tentative* are the most commonly observed tones for both COVID-19 and H1N1-09 pandemics. Whereas no *Confident* tone is detected potentially due to the high level of uncertainty about the trajectory of these two pandemics. The overarching tone seen in the articles though was *Sad*, which is most likely in reaction to the lives lost during these two public health crises. Public sentiments peaked during the first wave of national lockdowns in April of 2020 for COVID-19 and subsided in the subsequent summer. For H1N1-09, the public sentiments peaked months before the first wave of infections was reported. Overall, the news coverage of H1N1-09 is significantly less but

with higher fluctuations compared to COVID-19 because of its scale, severity, socio-economic impact, and the boom in social media for the past two decades.

Text Summarization Tool Evaluation: A Study on Automatic Summarizing News Articles

Michael Foley, Andrei Gerashchenko, Rachael Tovar
Department of Computer Science
Wheaton College, Norton, MA

Faculty Advisor
Zhou Tong
Wheaton College, Norton, MA

Automatic text summarization is the process of shortening a piece of text with a program while retaining the main topics and events. We surveyed a wide range of text summarization tools that are publicly available online and present our preliminary results using 80 published news articles. A shortlist of 8 text summarization tools was selected for this study, with four extractive-based and four abstractive-based. We used the selected summarization tools to summarize the news articles. Such machine-generated summaries are compared with four different user-generated summaries including the headline, the placeline, and the first sentence of an article. We use the state-of-the-art metric, ROUGE-1, and an additional binary indicator (On-Track) by human readers to judge the grammar and whether inaccuracies were missed by ROUGE-1. The preliminary results indicate that Bart, Pagusus, and SMMRY produce the most accurate and concise summaries. Bart performs the best with an average Rouge-1 score of .244 when the placeline of an article is used as the target and an On-Track score of 61.66%. Bart and Pagusus outperform other abstractive tools and SMMRY leads the extractive tools. We observe that the abstractive tools produce better summaries in general. Some limitations that we ran into with the tools were the accepted character length and problems dealing with certain special non-English characters. In addition, Bart tends to pick the first sentence of the article as the summary with high possibility and DistillBart does produce garbled summaries from time to time. We plan to support more text summarization tools, conduct a more comprehensive study with a larger number of news articles, and improve the web application in the future.

Creating An Interactive Dashboard for Industry-Based Employment Data

Britney Bourassa
Manning College of Information and Computer Science
University of Massachusetts, Amherst, MA

Faculty Advisor
Neena Thota
University of Massachusetts, Amherst, MA

I present my research on the development of an interactive dashboard application utilizing industry-based data from the US Bureau of Labor Statistics. The goals of this research are a) to develop an interactive dashboard for industry-based employment data for users and b) to adhere to usability and accessibility guidelines during the development of the dashboard. Creation of the interactive dashboard will add to the current research on the effectiveness of data visualization and will help determine the effectiveness of data visualization to help interpret a dataset. The motivation for this project comes from the need to make industry-based data more available to late high-school and early-college career students as they make important decisions regarding their careers. There is a need from the industry to have employment-based data easily accessible to users in a manner that the Bureau of Labor Statistics does not provide. The methodology for this project follows the structure of full-stack development and utilizes Agile/Scrum practices to implement the end product. The dashboard will utilize line graphs for time-based data and radar graphs to give an overview of data on an industry. To evaluate accessibility and useability, the Web Content Accessibility Guidelines will be referenced to help achieve Conformance Level "Double-A" priority guidelines. My poster will discuss the development of the product and the challenges encountered by the developer using Agile/Scrum practices. The full description and implementation features of the dashboard will be presented in detail.

InfinAI: NPCs Get Smarter

Brandon Olah, Anthony Radziunas, Joshua Bartholomew
Computer Science Department
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

In general, the programming of computer opponents in video games is insufficient and rigid. Inevitably, this leads to predictable and similar actions being taken to solve immediate goals. Our research seeks to discover new implementations of machine learning to enhance the competence of computer opponents and make the experience more "fun" for the player. This will allow developers to create games that are more dynamic, unpredictable, and evolutionary.

The gaming industry is only growing, and gamers expect new games to have different dynamics. One way this can be accomplished is by implementing enemies who learn from past experiences with the player and can become more challenging. We hope to use machine-learning to make innovative game entities that will act in a way that prevents the creation of dominant strategies and ways to break the illusion of the perceived "intelligence" of the system.

Python & Private Variables

Joshua Bartholomew
Computer Science Department
University of New Haven, West Haven, CT

Faculty Advisor
Liberty Page
University of New Haven, West Haven, CT

This project will look at the relative security of Python regarding private variables and functions used in most other programming languages. Python has only grown in popularity due to its simple syntax and developing capabilities. However, little research has been published about how secure Python code and programs compiled from Python code are without private variables. This research will seek to expose vulnerabilities in Python code, determine what must be done for these vulnerabilities to be exploited, and how, if possible, programmers can protect against exploiting possible vulnerabilities in Python code.

The proposed methodology includes examining the private variable concept in other programming languages and conducting experiments to determine whether Python has any vulnerabilities specific to a lack of private variables and functions. Based on the findings of these experiments, further research may be needed to explore the range of vulnerabilities in Python code and how to protect against exploiting these vulnerabilities.

Gesture Recognition and Lip Reading Through Computer Vision

Niko Severino
Computer Science
Connecticut College, New London, CT

Faculty Advisor
Ozgur Izmirli
Connecticut College, New London, CT

This research project focused on developing a lip reading and gesture recognition program based on computer vision. To achieve this, a recurrent neural network called an LSTM (Long Short Term Memory) was trained on a video data set the author created that captures the subject's hand gesture and lip movements. The context for the dataset is simple communication in a noisy environment such as a bar, where communication through speech is generally not possible. The dataset consists of six word classes with three different speakers: Shot, Pitcher, Beer, One, Two, and Three. Each spoken word is accompanied by a unique gesture in the video. Words for this dataset were chosen both for their applicability to the situation, but also because these words were fairly distinguishable from one another. That is to say, there is a lesser chance of confusing the words "Shot" and "One" than to "One" and "No". Although training on raw video would be optimal for capturing subtleties in movement, the video clips were processed using a program called OpenPose. The program recognizes the body and face, and returns a set of pixel locations for facial landmarks and joint locations of the skeleton representing the body for each given frame of the video. These data points are then fed into our LSTM as sequence batches of frames. The neural network was trained on 110 clips and tested on 50 clips. Each modality (gesture and lip movements) was tested separately. For the lip reading test, the neural network achieved a 73.9% accuracy. As for the gesture recognition, it achieved a 93.4% accuracy. Future work will focus on joint modality inputs to the neural network, experimenting with other architectures to improve performance, augmenting the current vocabulary, and adapting the system to different contexts that could benefit from the described system.

SmartPark

Ben Greenfield, Ben Placzek, Steven Atilho
Computer Science
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

Parking is a nationwide issue that millions of people face each day. With the production of more than ninety million new cars each year, this issue will only continue to become worse. The creation of new parking lots and parking garages is prohibitively expensive and require large areas to develop. Additionally, the environmental impacts of parking cannot be ignored. As much as three times the amount of carbon dioxide is emitted from vehicles looking for a parking space compared to the amount produced while driving to and from destinations. Solutions proposed by others introduce reservation style systems where users can reserve a parking space. However, this introduces additional issues of the need to add more hardware and people to enforce reservations which leads to a solution which adds extra expenses and is difficult to scale. Because of this, we propose our solution, SmartPark. SmartPark utilizes pre-existing infrastructure of camera surveillance systems within parking lots. These cameras can be used for image processing which will allow the detection and analysis of available parking spots in a given area. This information can be reported to end users in a mobile UI. The data collected will also be used to formulate predictions of parking availability. This will be able to inform users what parking lots at what time will have the best availability for parking. This solution will reduce the time needed for cars to circle parking lots which will help reduce emissions. Since the solution uses the existing infrastructure, no additional hardware or resources are needed to incorporate the solution.

AutoBot

Krystina Lorch
Computer Science/Cybersecurity
Felician University, Rutherford, NJ

Faculty Advisor
Sofya Poger
Felician University, Rutherford, NJ

AutoBot is a virtual mechanic application that gives drivers insight to their vehicle when a problem occurs. This app will let drivers have immediate access to a problem of a vehicle as well as step-by-step instructions on how to solve the problem. The simple process goes as follows: download AutoBot, set up a profile, add your vehicle to the vehicle list, input any problems when the car was purchases/obtained, and use the AutoBot scanner to scan the cars data for a baseline. The motivation behind this project comes from personal experience. My home residence is in California, but I attend school in New Jersey. I've driven my personal vehicle across the country without my parents twice and in those two times I've had to deal with problems with my vehicle. When a light comes on in a vehicle, many drivers, like me, panic due to not knowing what is wrong with their vehicle and not knowing how to fix it. AutoBot lets drivers know exact issues with their vehicles to clear any worry when it comes to fixing it themselves if there is no service close by or even eliminating the risk of being scammed for unnecessary fixes when the vehicle is taken to a mechanic. The methods for data collection revolve solely using the AutoBot app. AutoBot will have access to the user's cloud to record data of the vehicle when an incident occurs by screening daily data. The daily data will read and when AutoBot decides there are no issues at the end of the day, unnecessary data will delete to give room for future data. If an incident is found, the data saves for future reference whether it may be the user wanting to sell the vehicle or the vehicle has a similar issue.

Gargoyle Guard: Real-Time User Activity Fingerprinting

Charles Barone, Orion Duffy, Alexander Russello
ECECS Department
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

Limitations in current authentication implementations is an emerging problem in today's cybersecurity landscape. A significant point of failure in the currently deployed authentication systems is the lack of a means to validate that the active user present post successful authentication remains consistent for the entirety of the authenticated session duration. This work aims to propose an extended definition of continuous authentication within existing taxonomies to allow for an emphasis to be placed onto the fingerprinting of user activity on technological assets in addition to defining a new component of the authentication paradigm, what you do.

PharmBERT: a Pre-trained Language Model for Pharmaceutical Error Prediction

Dustin Doctor
Computer Information Systems
SUNY Buffalo State College, Buffalo, NY

Faculty Advisors
Gang Hu
SUNY Buffalo State College, Buffalo, NY

Bo Yu
Dalhousie University, Halifax, NS, Canada

Total number of retail prescriptions filled annually in the USA has reached 4.69 billion in 2021. However, the tracking over the service quality of the dispensation process is still very limited. In an effort to address factors that lead to quality-related events, some healthcare organizations and governments adopt error-reporting systems. Such reporting systems have collected pharmaceutical errors that either reach patients (incident events), such as incorrect drug, dose or quantity, or are intercepted at pharmacies (near miss events).

To discover common contributing factors that may have led to quality-related events, large-scale analysis of these event is crucial. Many common factors in retail pharmacies that resulted in an incident may not be obvious to the human eye and traditional data-mining solutions. With the progress of deep learning in natural language processing (NLP), the development of effective mining has been boosted, including the field of extracting valuable latent information from pharmaceutical documents.

In this research, Bidirectional Encoder Representations from Transformers (BERT), is utilized to make predictions on the pharmaceutical transaction data (collected by a Canadian error-reported system). To fit pharmaceutical data with the BERT model, we formatted event information into Natural Language tokens, and fine-tuned on the pre-trained BERT model. The trained pharmBERT model is able to achieve an accuracy of ~84% when predicting whether an event would

result in a near miss (caught beforehand), or an incident (caught afterwards). We are also working on using this model to further predict other aspects of the event, such as what stage of the events (prescribing, transcribing, dispensing, administration, storage, and monitoring) the incident occurs or what category of issues the event falls under. We believe that the findings from this study could lead to solutions to reduce pharmaceutical incidents and provide improvements in patient safety.

Neural Network Learning Based on Visual Data

Evan Patterson
Computer Science and Innovation
Champlain College, Burlington, VT

Faculty Advisor
Jim Hoag
Champlain College, Burlington, VT

Learning in neural networks has historically been something that doesn't typically happen with much sample efficiency. This is in contrast to what biological brains are capable of. In this paper, an attempt is made at replicating the behavior and some of the capabilities of biological brains, including the ability of a biological brain to learn the structure of its environment through observation alone. Ultimately this attempt results in an original algorithm that can be used to set the weights of a neural network based on visual data collected from a synthetic environment consisting of a simple simulated organism and a maze. While the neural network is used to condense and store data about the environment, the learning algorithm uses a Markov clustering algorithm to discover relationships in the collected data, and may be an example of a way to train neural networks without needing to initialize the weights with random values, or even necessarily use gradient descent. The overall application of the neural network in this case may illustrate how neural networks can be used to understand and navigate novel environments.

Investigating Practices in Long Term Student Development

Gabriel Bick, Michael Lanners
Computer Science
University of the Pacific, Stockton, CA

Faculty Advisor
Osvaldo Jiménez
University of the Pacific, Stockton, CA

We present our experience working on an existing educational video game in the Computer Science Department at University of the Pacific. Currently named AlgoGame, it aims to help computer science students understand recursive backtracking and other programming concepts in a puzzle format in the spirit of LightBot. The game has been developed using Unity, JetBrains Rider and is currently hosted on GitHub. The project has been in development since 2019, with 32 student contributors over the course of development. The development of AlgoGame presents an opportunity for more experienced student developers to work on an existing codebase and handle legacy code. Despite extensive documentation, it is difficult for student developers to understand how code changes have affected functionality. Therefore, we studied existing version control revisions and developed practices to help future developers of AlgoGame understand changes better. We created and documented these practices by analyzing different versions of the game that have been stored into version control and examining and integrating those features into the official version of the game. Our primary objective was to understand all of the functionality that was implemented by different student developers, and to understand intentional versus accidental or superfluous changes to files. This becomes even more challenging when dealing with proprietary files which may obfuscate modifications. Due to the difficulty of collaborative student development over time, we have developed rules and practices for future developers to follow on this project which could apply to similar game development projects. We have also implemented various testing practices into Unity to facilitate in implementing future changes and identifying potential bugs. These practices include Unity-specific guidelines to aid in

integrating changes between branches and to minimize technical debt. We hope our findings will make future developer transitions more efficient for student-led projects.

Intelligent Question Selection in Tutoring Software

Emerson Bieck
Computer Science
Ithaca College, Ithaca, NY

Faculty Advisor
Toby Dragon
Ithaca College, Ithaca, NY

I present my contribution to Artificial Intelligence in Education as part of a summer research internship offered by the Computer Science Department at Ithaca College. The goal of the internship is to gain experience conducting research and collaborating with professors. Unlike Kahoot! or Quizlet, which have students answer a list of questions in a set order that provides one overall assessment of performance, the PAR (Practice, Assessment, and Review) project helps students master course material by assessing their ongoing performance and adapting the questions asked. My work involves the intelligent agent which uses question data organized by tiered concepts. I want to ensure a student can progress through the concepts at their own pace, so that they receive questions of appropriate difficulty and variety. Thus, students with mastery of the material would answer fewer questions and progress through concepts faster than those with a more basic understanding. To start, I worked with Dr. Dragon to encode the concepts and order in which they should be taught for part of an introductory data structures course, and then authored appropriate questions for those concepts. We also experimented with model behavior based on each performance measure. For example, we looked at when a student's progress in a concept would be considered "Unprepared" versus "Developing" and how that concept would be treated by the system. By the end of the internship, I successfully created a fully integrated AI component ready for user tests to validate the relationship between prior mastery and performance using the system. A Java codebase which entailed research into Spring Boot (web framework) and Mockito for stress testing. My poster will describe in detail the pedagogical model informing the intelligent decision-making, the challenges in controlling question difficulty and variety, the preparation of the intelligence for user testing, and future project plans.

What Do Patients Value? Using Natural Language Processing to Define a Good Doctor's Visit

Courtney Saqueton, Yash Karandikar
Computer Science
Occidental College, Los Angeles, CA

Faculty Advisor
Celia Chen
Occidental College, Los Angeles, CA

Doctors' reviews are becoming increasingly common - not only are more people anonymously reviewing their doctors' visits, but more people are consulting these online reviews to make their own healthcare decisions. Since online reviews are anonymous, they provide a platform for patients to share their honest and unfiltered feedback. Such feedback is highly informative and contains unexplored insights into the healthcare system, but it is difficult to utilize effectively due to its unstructured nature. Therefore, we propose an automated approach using natural language processing techniques to extract features from doctors' reviews that indicate which qualities patients value in their doctors. These doctors' reviews are extracted from a public Yelp dataset from businesses containing keywords such as "Doctor" or "M.D.". Our approach consists of extracting the highest intensity patient values (e.g., "caring" and "horrible") using logistic regression, and remapping these patient values back to their context using dependency parsing (e.g., "caring doctor" and "horrible customer service"). The logistic regression model uses a train set (N=1029 reviews) and is tested on a test set (N=257 reviews). The effectiveness of our approach is evaluated by using dependency parsing on an untrained evaluation set (N=263 reviews) and calculating metrics on precision (93%), recall (95%), and F-1 measure (94%). We found that what patients value most in their doctors' visits are friendly staff and caring providers. Following those primary values, patients seek professionalism ("very professional team") and kindness ("a kind demeanor"). With such an approach, we can obtain timely and direct insights about what thousands of patients value in their doctors. To extend our research, we are currently running our model separately on different medical specialties in order to check for bias in reviews (e.g., gender bias).

Stance Prediction for Contemporary Controversial Issues

Mark Stowell
Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisor
Adita Kulkarni
SUNY Brockport, Brockport, NY

Confirmation bias is the tendency of humans to favor information that confirms one's previously existing beliefs. Contemporary debatable issues usually carry some political weight and controversy. Such information is widely available on social media which is one of the primary sources of news across the world. This information is persuasive and promotes conformation bias, which enables internet debates spreading discrimination, misinformation, and hate speech, all of which are crucial problems on social media that need to be addressed. Stance detection focuses on identifying a person's opinion, either in favor of or against an idea. This work aims at using machine learning for detecting stance of users on social media for contemporary controversial issues. I collect data related to issues such as vaccination (in particular, COVID19 vaccination), and legalization of abortion, from social media platforms such as YouTube, Twitter, and Reddit. I aim to design machine learning models that leverage network features and sentiments in user posts/comments to automatically detect stance of users regarding these issues. This reseach will help understand the origin of misinformation and enable spread of unbiased information.

ScreenAware: an iOS App to Manage Problematic Smartphone Use

Zoe Beals
Computer Science
Skidmore College, Saratoga Springs, NY

Faculty Advisor
Aarathi Prasad
Skidmore College, Saratoga Springs, NY

Current smartphone management tools, like Apple's ScreenTime, focus on time-based restrictions, and have been shown to cause negative emotions like feelings of shame or stress from how much time is spent on an individual's smartphone. Our initial solution for an alternate approach for smartphone management called ScreenSnooze (presented at CCSCNE 2020) focused on making a user mindful about when they unlock their phone. However, prior research shows that it is the nature and content of use and not the time spent on the phone that determines problematic smartphone use. We conducted focus groups with 38 college students to better understand how to help users build better phone habits. Based on our findings, we hypothesize that a tool to help people reduce problematic smartphone behavior should account for their goals for using their phone and may best help by allowing the user to create personal, measurable goals for their phone use based on their own values. I developed an iOS app called ScreenAware that allows people to set daily goals for their smartphone use, mark goals as complete, and track their progress. I also added incentives such as badges and avatars to engage the user, based on feedback from the focus groups. Next month, we are planning to conduct a week-long user study to understand the effectiveness of the app in helping people manage their smartphone use, during which participants will install and use the ScreenAware app on their personal iPhones. To understand how the goal-based approach affects user emotions, we will also ask users to record their emotion via a popup when they create and complete a goal. I will present the findings from the focus group, describe and provide screenshots of the ScreenAware app and finally, report our findings from the user study.

Redesigned UI and Spatial Data Structures
for METAL Algorithm Visualizations

Bailey Cross, Luke Jennings, Spencer Moon
Computer Science
Siena College, Loudonville, NY

Faculty Advisor
James Teresco
Siena College, Loudonville, NY

We present our web application METAL (Map-Based Educational Tools for Algorithm Learning), an ongoing project aimed at helping students learn search and graph algorithms, by using real-world highway data in Algorithm Visualizations (AVs). The AV system allows students to interact with many algorithms, showing the pseudocode of the algorithm, changes in variables and data structures, and a visualization of the algorithm on the map. The use of highway data and AV allows students to run an algorithm on datasets of various sizes simulated at different speeds. As a result, students acquire a deeper understanding of a given algorithm more quickly. In the Summer of 2021, we performed a massive user-interface overhaul to make the application more visually appealing and improve the user experience. In addition, two new AVs were added, the first simulating inserting points into a quadtree, which is a spatial data structure useful in quickly accessing points in multidimensional space. Our main goal with our quadtree AV was to emphasize the recursive nature of the data structure and refinement algorithm while being easily understandable. The second AV simulates a space-filling curve traversal, which is a remapping of multiple-dimensional points to a single line. Space-filling curves are useful in that they are used in mesh-partitioning algorithms, image and video rendering, and object tracing. Our AV allows students to choose a variety of space-filling curves to determine the linear ordering of the vertices of a selected graph, from simple longitudinal scans to Hilbert and Morton curves. We hope with these changes that METAL can become a more widely adopted teaching tool for undergraduate computer science departments across the country.

Unicon UI/IVIB Integration

Jonah Schnettler, Matthew Lane
Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisors
Zakariya Qawaqneh
SUNY Brockport, Brockport, NY

Jafar Al-Gharaibeh
Architecture Technology Corporation, Eden Priarie, MN

We present our experience participating in the independent study practicum course in the Computer Science department at SUNY Brockport. The focus of this course is the continued development of the Unicon programming language, led by Dr. Jafar Al-Gharaibeh and supervised by Dr. Zakariya Qawaqneh. In this project, our scope was to integrate two pre-existing development tools into one integrated development environment (IDE) to create a more user-friendly interface. The first developer tool being Unicon's improved visual interface builder program "IVIB" and the second being "Unicon IDE," Unicon's integrated development environment. While both tools function as intended independently, there was no way to build a GUI using IVIB and edit the code at the same time. To start with, we had to figure out the best way to integrate 100+ source files with 30k+ lines of code between them to be built within the same interface. We faced a series of tribulations, but by the end of the semester, we had achieved our goal. The significance of this project is that newcomers to the language will typically use Unicon's IDE and by integrating these developer tools, we are creating a more beneficial learning tool for these developers. With a new semester, we are given more time to continue development to fix bugs and continue with UI enhancements. As undergraduates on this project, we gain insight into working on projects with little documentation and no way to simply google the answer. We face problems that we have to figure out for ourselves through creative solutions or suggestions from mentors that require us to continuously

adapt and overcome. It is projects like this that give students some insight into the unknown and force us to look outside of the realm of academia to reach new innovative levels.

Clustering for Community

Autumn Smith
Computer Science
Wellesley College, Wellesley, MA

Faculty Advisor
Christine Bassem
Wellesley College, Wellesley, MA

I present my experience working with Professor Bassem in the Computer Science department at Wellesley College, inspired by a real problem that the Crew team was facing due to COVID-19. After developing a framework for solving the initial problem, our focus expanded such that our proposed solution would fit a broader audience. Traditionally, the crew team maintains a well-knit community by creating mentorship groups with a balanced mix of new and senior members. However, due to the pandemic and campus safety protocols, these groups could not be easily determined. We started by modeling our problem as a network graph, where the goal was to cluster the Crew team into small, similarly sized groups of people who had not met in the safest way possible, given a safety metric built off of dorm location, dining hall preference, and practice schedule. After researching how different clustering algorithms worked, we decided to adapt the K-Means algorithm that uses Euclidean distance (L2 Norm) as a distance metric to produce clusters within a pre-set group size, which is something not typically considered in K-Means. To build a solution that could benefit others, I built a website that accepts any data that needs clustering. My design includes data cleaning techniques and One-Hot encoding, in order to meaningfully work with categorical data, and an intuitive project workflow that meets all current goals while still maintaining flexibility for new ideas. My poster will discuss the project workflow–including data cleaning, processing, and clustering–and the final version of the product. The full description and implementation of the project will be presented in detail.

Unicon HTTP Server

Stryder R. Coleman, Kyle Knopp, Daniel Cona
Computing Sciences
SUNY Brockport, Brockport, NY

Faculty Advisors
Zakariya Qawaqneh
SUNY Brockport, Brockport, NY

Jafar Al-Gharaibeh
Architecture Technology Corporation, Eden Priarie, MN

We present our experience participating in the independent study practicum course in the Computer Science department at SUNY Brockport. Dr. Jafar Al-Gharaibeh, one of the creators of the Unicon programming language, led this project, while Dr. Zakariya Qawaqneh supervised it. The objective of this project was design and develop an HTTP server class library for Unicon so that end users can implement their own custom http server solutions on top of the class library. The HTTP server class was built to support the common HTTP protocol specifications, including request methods such as, Head, Get, Delete, Put, and Post. The server uses TCP socket connections and launches a thread to handle each new client. Using threads allowed us to 1) build a server that can scale to large number of clients, and 2) each client session state is maintained in the context of that thread improving data isolation and security of the server. Client requests are parsed and handled according to HTTP protocol 1.1 as defined by RFC 2616. We carried out testing using an HTTP client that we built as part of the project, as well as using popular web browsers. At the end of the semester, we met all the project requirements and delivered a class library with a functional HTTP server. Some Unicon users already started using our HTTP server implementation. The modular HTTP class library we built will allow future Unicon developers to enhance the server and create new applications. The HTTP class is an ongoing project. Future features and improvements are in production such as configurable files, improved logging, client authentication, and more. The source code is available at https://github.com/coleman3616/HTTP_Server.

Parallelizing the RNA Secondary Structure Dynamic Programming Problem

Matilda Ferguson, Alice Huang, Katie Knox, Sojin Lim
Computer Science Department
Swarthmore College, Swarthmore, PA

Faculty Advisor
Tia Newhall
Swarthmore College, Swarthmore, PA

We explore tradeoffs in two parallel computing approaches for solving the RNA secondary structure problem from biology. Predicting the RNA folding pattern, or the maximum number of pairings formed in a sequence of RNA bases, is essential to understanding the structure's biological function. This problem is both computationally intensive and data-intensive, motivating the need for efficient parallel solutions.

There have been numerous attempts to formulate and solve the RNA secondary structure problem computationally. Our approaches are based on Nussinov's dynamic programming algorithm. Dynamic programming is a natural solution for this type of problem that can be solved by breaking the problem down into smaller subproblems whose results are reused and combined. Solving these numerous subproblems lends itself to parallelization. However, Nussinov's algorithm, which iteratively fills the upper triangular of a 2D table, results in an irregular computation pattern, complicating efficient parallelization.

Our solution involves designing and implementing a shared memory CUDA parallelization and a distributed memory MPI parallelization. Our parallelization is based on existing approaches using CUDA, and we additionally draw from Liu and Schmidt's block-cyclic based wavefront pattern in our MPI algorithm. We then compared the performance of our parallel approaches and a sequential solution, for a range of RNA sequence problem sizes, and investigated different partitioning approaches for the MPI version (column, block, and block-cyclic).

Our results show that the CUDA shared memory parallelization produces greater speedup than the MPI implementation, though this speedup tapers

off due to the synchronization high overhead in the implementation. We also find that the MPI column-based partitioning of the computation table proves to be more scalable than the CUDA implementation, as it achieves much greater speed-up when the number of processors increases. Additionally, both parallelizations outperform the sequential version by a factor of up to thirteen.

"Thinking" About the Next Generation of Gaming

Chris Howard, Jason Smith, Jim Durante
Computer Science
University of New Haven, West Haven, CT

Faculty Advisor
Mehdi Mekni
University of New Haven, West Haven, CT

The goal of a virtual environment is to facilitate an immersive user experience. In 2021, consumer grade virtual reality (VR) headsets have increased in functionality and demand. We intend to develop an innovative immersive experience that complements a VR environment and increases the overall accessibility of the virtual landscape. We intend to meet these goals by merging the immersion of a VR headset with the intuition of a brain computer interface (BCI), the BCI intrepets brain signals to enable users to play hands free using the Emotiv Epoch headset. Currently, we have moved an in-game 3D model with EEG data.

The research questions we are trying to answer with our prototype are, "Are pure BCI controls, traditional controls(hand held controls), or a mix of the two paradigms the most engaging?", "Can we give real-time insight on user engagement?", and "Will BCI controls increase user accessibility for users who may not have access to traditional controls?" To answer these questions, we must compare the quantitative data observed during the testing of our prototype. This includes analyzing brain waves like alpha and beta waves in order to see when the brain is being engaged. Along with looking at brain waves we will also note the scores of the user and see if they correlate with the engagement the user is having with the game using the BCI model or the traditional model.

We intend to validate our data using a combination of EEG power bands and user-surveys. By expanding our core concept, we can then seek to monetize our product by demonstrating that our system could be used to categorize user engagement beyond just our project. Game developers could gain quantifiable and actionable data on the level of engagement of a new product is one of many applications for our project.

BERT-based Negotiation Chatbot

Matthew Clifford
Computer Information Systems
SUNY Buffalo State College, Buffalo, NY

Faculty Advisors
Gang Hu
SUNY Buffalo State College, Buffalo, NY

Bo Yu
Dalhousie University, Halifax, NS, Canada

Business negotiations are often hard due to conflicts of involved parties. Some negotiations can be not only time-consuming but also negative, resulting in the damages of business relationships when unexpected negative emotions grow. A solution to these problems is to automate negotiations with a robot. We propose a project to create a chatbot that can conduct business negotiations.

The current project will utilize BERT (Bidirectional Encoder Representations from Transformers), which is a deep learning natural language representation model that has a powerful bidirectional prediction and contextual understanding feature. The BERT model is pre-trained on an enormous amount of unlabeled data. The model allows high performance when it is fine-tuned to a specific task through additional training. The BERT model trains data by performing 2 tasks: MLM (Masked Language Model) which is used to predict the missing word(s) in close vicinity within a sentence, and NSP (Next Sentence Prediction) which is effective for the question/answer task.

The first step in this research will be fine-tuning the model to our negotiation task, using more than a thousand bilateral negotiations experimentally conducted globally. Next the MLM and NSP tasks will be extended to predict the binary result of a negotiation (successful or unsuccessful) and to generate an automated response to a human counterpart. This BERT-based negotiation chatbot could be utilized as a business representative who can sense and foresee the positive and negative atmosphere during the negotiation, and strategically react to the

opponent. The goal is to help two parties reach an agreement that both parties are satisfied with.

Topic Modeling of Dark Websites on Tor Using Deep Neural Learning

Christopher Schmitt
Computer Science
Central Connecticut State University, New Britain, CT

Faculty Advisor
Mahdieh Zabihimayvan
Central Connecticut State University, New Britain, CT

Tor is the most well-known dark network that protects the identity of both content providers and their clients against any tracking on the Internet. It has noble uses, including as a platform for free speech and information dissemination, but may be culturally notorious to be a platform for criminal activity and illegal marketplaces. The previous research on the information provided on Tor either conducted exclusive study on a particular topic or utilized exhaustive supervised investigation to label Web pages on Tor. In this work, we model topics of Tor's Web pages using RoBERTa that works based on deep neural learning. Our study is the product of a corpus of 5000 dark Web pages that we collected on Tor in February 2020. To encode semantic meaning of pages, the method transforms token streams into dense vectors so that documents with similar semantics can be located at similar positions in the embedded space. We further apply two unsupervised classification methods of K-Means and DBSCAN to group documents into semantically similar groups. RoBERTa's result recovers 5 topics of captchas, casinos, bitcoin, guns, and credit cards. We also compare RoBERTa's performance with two other topic modeling methods, BERT, ALBERT, using silhouette score across all clusters. Based on the experimental results, RoBERTa outperforms both BERT and ALBERT in creating semantically similar and distinct clusters for both K-Means and DBSCAN. The RoBERTa embeddings produce both fewer outlier documents and have a higher average silhouette score across all clusters, which implies that documents of each cluster are more similar to other documents in that same cluster, and more distinct from documents in other clusters.

Multimodal Framework for Computational Musical Performance Analysis

Nikesh Ghimire
Computer Science
Connecticut College, New London, CT

Faculty Advisor
Ozgur Izmirli
Connecticut College, New London, CT

This research project is focused on constructing a multimodal framework capable of tracking audio-visual musical features via computer vision and digital signal processing techniques for use in Music Information Retrieval (MIR). State-of-the-art MIR techniques utilize deep learning with various model configurations, thus shaping the motivation for this project: Improving on traditional monomodal MIR techniques via machine learning that incorporates multimodal features. The framework's design prioritizes end-to-end efficiency, from data preparation to training and evaluating machine learning models, and was evaluated with the controlled and well-annotated University of Rochester Multi-Modal Music Performance (URMP) Dataset. OpenPose, an open-source platform capable of detecting locations of joints of multiple human bodies in videos was utilized to extract skeletons (11 key points) of the upper body of musicians. This allows the framework to automatically crop videos of musicians with configurable Regions of Interest (ROI) and extract coarse movement information relevant to musical performance. Optical flow, a technique utilized to compute the apparent movement of pixels between consecutive frames, was applied to cropped videos of individual musicians to extract finer movements such as a violinist's bow movements and fingers on the fingerboard. The framework also extracts audio features such as spectral flux and chroma with traditional MIR techniques for both mono and polyphonic pieces. These streams of audio-visual features were combined in various configurations for machine learning, pyTorch, to tackle MIR problems: Instrument detection and recognition (94.5% accuracy with 13 classes), single instrument onset detection (95.2% accuracy for violins), and others such as bowing direction and reversal detection. Future work will focus on adding more multimodal extraction techniques and data

streams, implementing new deep architectures, such as Temporal Convolutional Networks (TCN), and evaluating models with other datasets. Beyond music performance analysis, the framework is easily applicable to a wide range of audio-visual data from other contexts.

Automatic Sanitization of Computer Workstations using Computer Vision and UV-light

Yurock Heo, Zachary Beucler
Computer Science
Connecticut College, New London, CT

Faculty Advisor
William Tarimo
Connecticut College, New London, CT

As students return to in-person classes after a hiatus due to the COVID-19 pandemic, shared spaces in schools like computer labs and classrooms have seen a rise in usage which demands a more rigorous cleaning schedule. Sanitizing such frequented electronics and spaces manually using traditional disinfectants is tedious and unsafe to the electronics and to the personnel cleaning. We have designed and developed an autonomous system for the sanitization of computer workstations using UV-C LED lights, IoT devices, and computer vision. Our approach involves installing a structure at each computer workstation that overlooks the keyboard and mouse. On the underside of the structure is a UV-C LED strip that is connected to an IoT outlet. A Raspberry Pi microcomputer uses a video stream to monitor the computer workstations and track usage/occupation using computer vision. Once the Raspberry Pi detects that a human has used a computer workstation, it marks that workstation as dirty. Once the Raspberry Pi has determined that it is safe to sanitize the workstation, it will send an HTTP request to the IoT outlet that is responsible for the marked workstation. Once the request has been received, the IoT outlet turns the LED strip ON for a period of time, sanitizing the keyboard, table surface, and mouse. This system can be applied to any situation where there are multiple locations that need to be observed and action has to be taken on them when a person or object has had close proximity to it.

Sensing the walking velocity of a person by using mobile devices

Anthony Smith
Computer Science
Winston-Salem State University, Winston-Salem, NC

Faculty Advisor
Muztaba Fuad
Winston-Salem State University, Winston-Salem, NC

In an indoor space, determining a person's speed of mobility has a lot of research significance and applicability in real-world scenarios. This research has developed a mobile application to investigate how to determine a person's walking speed. The goal was to determine a person's walking speed by using the number of steps. There has been similar work to test the accelerometer sensor in detecting steps. However, the accuracy of using the steps to calculate the velocity was not studied. This application uses the accelerometer sensor in the mobile device to detect steps and then compute the velocity. The accelerometer provides information about the user's motion and acceleration, and an algorithm was developed to use that data to determine the steps. Once steps are determined, the person's speed is calculated by using the change of location within a pre-determined space and time. Therefore, accurately measuring the number of steps was vital and it was determined that the position of the mobile device in the body plays a significant role in that accuracy. The experiment used three device positions: the pants front pocket, the right hand, and the backpack. While walking, the number of steps was manually counted and recorded. A comparison was made between the recorded number of steps to the application's measured steps. The experiment was conducted multiple times for each device position. The placement of the mobile devices in the front pants pocket gives the most accurate results, whereas the other two device positions gave reasonably accurate results. The position of the device played an important part in the research and had a significant impact on the accuracy of the results. In the future, testing can include additional device positions. Additionally, other mobile device sensors could be included in the testing and can be compared with this approach.

Thank You CCSC National Partners

Platinum Level Partners

Gold Level Partner