

**Consortium for  
Computing Sciences in Colleges**  
in cooperation with

**The Association for  
Computing Machinery  
Special Interest Group  
on Computer Science Education**  
presents



**CCSCNE-2016**  
**The Twenty-First Annual**  
**Consortium for Computing**  
**Sciences in Colleges**  
**Northeastern Conference**

**Student Poster Abstracts**



April 29-30, 2016   Clinton, NY



## Table of Contents

Poster	Title	Page
5001	Comparing Statistical Approaches to Anomaly-Based Intrusion Detection Systems	6
5003	Converting Cyberlearners' Behaviors into Insights	7
5004	Correlation Between Twitter Moods and Market Performance	8
5005	PolRank	9
5007	Cyber Security Issues with Autonomous and Semi-Autonomous Robots	10
5008	R Language Across Disciplines	11
5009	From AppInventor to Android Studio: A Student's Perspective in Mobile Application Development	12
5010	uSafeNH: A mobile app for helping sexual assault victims on college campuses in New Hampshire	13
5012	Computer Game Industry as an Educational Path: From Serious Games to E-sports	14
5013	Efficient Placement of Photovoltaic Arrays using Genetic Algorithms	15
5014	Personal Cloud Storage	16
5015	Engaging Middle School Girls Using CS Games	17
5016	Keep Calm and Carry On: A Voice-Based Medical Prediction Engine	18
5017	Comparison of Heuristics in the Bin Packing Problem	19
5018	A Comparison and Analysis of Graph Storage Scenarios	20
5020	RFID Controlled Robot Game	21
5021	Slicing the Pie: Big Data Processing with Raspberry Pi	22

Poster	Title	Page
5023	Communication Therapies with the NAO Robot	23
5024	Computational Model of Surprise Using Narratives	24
5025	A Case Study on Modeling Social Network Privacy Policies Using Event-B	25
5026	Lowerbounds on the Minimum Metric Online Matching Problem on the Line	26
5027	Designing of Inventory Management System Using Open Source Object-Oriented Languages	27
5028	Securing Common Software Vulnerabilities	28
5029	Trade Efficiency Problem	29
5030	Business Intelligence Dashboard	30
5031	Seeking the Foundations of Artificial Neural Networks for Image Recognition	31
5032	AI Bracketology	32
5033	A Location-Based Algorithm for Encrypting and Decrypting Text	33
5034	Visualizing Trends in Top Ten Lists	34
5035	Automated Analysis of Eligibility Criteria in Clinical Trials	35
5037	Attentional Cost of Different Notification Types	36

Poster	Title	Page
5038	Increasing STEM Engagement with Mobile Applications	37
5039	Cooperative and Uncooperative Voting in Doodle Polls	38
5040	Real Time Virtual Window Simulation	39
5041	Development of a Formally Verified Android Application Using EventB2SQL	40
5042	Simulating Blackjack for Maximization of Success Rate through Card Counting	41
5043	Digital Dirt: Novel Terrain Error Metrics	42
5046	Space Invaders Software Engineering Learning Experience	43
5047	An Android Application Suite to Supplement Elementary Math Education	44
5048	Pruning Neural Networks in the Neural Evolution with Augmented Topologies (NEAT) Algorithm	45
5049	MiddGuard: A Framework for Collaborative and Extensible Visual Analytics	46
5050	Engaging Students with Adaptive Learning	47
5051	Bitcoin: Advancing Acceptance and Usability	48
5052	Improving Navigation And Searchability in App Inventor	49
5053	Predictive Link Following for Accessible Web Browsing	50
5055	A Model and Cross Country Analysis of Multitasking with Technology in Academic Settings	51
5056	Applying Novelty search to the construction of Ensemble systems	52

Poster	Title	Page
5057	Style and Sentiment Analysis in Movie Reviews	53
5058	Profiling Synchronization Patterns in Multithreaded Programs	54
5059	International Union for Conservation of Nature Ecosystem Visualizer	55
5061	Whiteboard Scanning using Super-Resolution	56
5062	Punctuated Anytime Learning to Evolve Autonomous Wheeled Agent Using Genetic Algorithms	57
5064	Optimizing Technical Indicator Parameters	58

**STUDENT**

Alrawi, Danya 38  
 Bender, Maximillian 26  
 Bielinski, Nicholas 12  
 Bisch, Joseph 16  
 Blithe, Michael 19  
 Boland, Bridget 23  
 Bond, Cody 32  
 Brown, Kevin 24  
 Caban, Adrian 16  
 Casal, Carla 17  
 Chakraborty, Maitry 17  
 Chao, Ella 7  
 Cleary, Jonathan 12  
 Cleary, Tucker 13  
 Colla, James 32  
 Combs, Kyle 16  
 Conlin, Tom 48  
 Coughlan, Matthew 6  
 Cox, Gunther 24  
 Crosby, Mark 23  
 da Silva, Iago 11  
 Diaz, Cristina 35  
 Doktor, Adam 22  
 Dong, Yangyan 51  
 Duval, Jared 24  
 Emerson, Maryjean 13  
 Eordanidis, Stephanie 17  
 Fahnbulleh, Whitney 7  
 Finta, Peter 45  
 Foun, Amanda 7  
 Franca, Brenda 43  
 Germinara, Brock 15  
 Giri, Christopher 39  
 Godard, Sean 31  
 Gonzalez, Kevin 30  
 Grajeda, Cinthya 30  
 Grannis, Justin 21  
 Grohe, Brain 24  
 Grupen, Nicholas 18  
 Howells, Alexandra 39  
 Iori, Melissa 20  
 Jones, Stephen 53  
 King, Alexis 35  
 Klein, Alyssa 36

**ABSTRACT**

LaMontagne, Alexa 37  
 Lawson-Keister, Elizabeth 41  
 Le, Hieu 52  
 Le, Son 55  
 Le, Trish 43  
 Lewis, Michael 27  
 Lu, Michelle 9  
 Lu, Yifan 41  
 Miller, Andrew 45  
 Ndiaye, Moustafa 57  
 Ng, Skyler 28  
 Nguyen Tra, My 25  
 Ngwenya, Mengezi 58  
 Ni, Wode 56  
 Nobrega, Christopher 12  
 Notarino, Louis 30  
 Oliveira, Victoria 30  
 Ozan Sivri, Ahmet  
 O'Donnell, James 32  
 Palmer, Tim 29  
 Panagopoulos, George 33  
 Passantino, Larry 44  
 Paustian, William 9  
 Peralta, Nikolai 9  
 Pianka, Adam 32  
 Pinder, Kyle 30  
 Rahn, Zachary 18  
 Ramos, Christian 43  
 Roznovjak, Jiri 50  
 Rubin, Zevi 34  
 Sarmanho, Kaê 14  
 Seggerman, Katie 21  
 Shen, Kasey 54  
 Silver, Dana 46  
 Sivri, Ahmet 11  
 Smith, Bonnie 13  
 Snider, Thomas 10  
 Tan, Zhuoming 55  
 Tinkham, Gregory 12  
 Tosca, Diana 7  
 Tran, Tony 32  
 Tsui, Cecillia 51  
 Tucker, John, 6  
 Ulinski, Grace 34  
 Valan, Connor 8

**INDEX**

Vurgun, Mert 11  
 Wentzel, Alexandra 8  
 Wereszczak, Nicholas 16  
 Wilkins, Jacob 42  
 Williams, Graham 40  
 Winchester, Phil 47  
 Young, Jacqueline 49  
 Young, Joshua 13

## Comparing Statistical Approaches to Anomaly-Based Intrusion Detection Systems

### Authors

**Matthew Coughlan,  
John Tucker**

Department of Mathematical Sciences  
United States Military Academy  
West Point, NY

### Advisors

**MAJ Thomas Nelson,  
MAJ Benjamin Klimkowski,  
MAJ Donald Koban**

United States Military Academy  
West Point, NY

Intrusion Detection Systems (IDSs) are systems used to defend a network against cyber-attacks. Specifically, anomaly-based IDSs are systems that detect malicious activity on a network by identifying departures in network traffic from a previously established norm. For this project, we will first create a data set of network activity, starting with a pre-established dataset of activity. We will then inject malicious traffic and intrusions in order to mimic a realistic network attack. The end goal of the project is to assess the validity and effectiveness of various types of statistical methods used to designate certain external traffic as malicious.

The first step in analyzing the statistical IDS methods will be to establish a baseline of external activity across a very specific portion of the network; we may do this by collecting traffic data from a few select ports that allow external packets to enter the network. We will then inject data that is indicative of an intrusion in order to easily test the statistical methods against familiar, simulated traffic.

The statistical methods that we will implement and compare are the cluster analysis methods of k-means, k-means++<sup>1</sup>, and Y-Means<sup>2</sup>. We will delve into the applicability of Bayesian statistics in identifying and classifying cyber threats as well. Given our results, it may be of interest to combine multiple methods in order to create a redundant and more robust method of identifying intrusions.

We will evaluate the chosen methods by looking at certain benchmarks such as success/failure rates, false positive rates, and consistency across varying data.

---

<sup>1</sup> David Arthur and Sergei Vassilvitskij, "k-means++: The advantages of careful seeding," *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete Algorithms*, (Society for Industrial and Applied Mathematics: 2007).

<sup>2</sup> Yu Guan, Ali-Akbar Ghorbani, and Nabil Belacel, "Y-means: A clustering method for intrusion detection," (2003).

## Converting Cyberlearners' Behaviors into Insights

### Authors

**Ella Chao, Amanda Foun,  
Whitney Fahnbulleh, Diana Tosca**  
Department of Computer Science  
Wellesley College  
Wellesley, MA

### Advisor

**Eni Mustafaraj**  
Wellesley College  
Wellesley, MA

Learning Sciences is a multidisciplinary field, bridging the gap between research and practice to improve learning experiences. Scientists in this field are interested in understanding the cognitive processes involved in effective learning to gain insight on design decisions of learning environments. Recently, we've seen a rise in the digitization of education through platforms offering Massive Open Online Courses (MOOCs). Our research focuses on students' video learning experiences in order to improve video and course content.

We analyzed the video watching behavior of approximately 7,000 students on 146 videos from an introductory programming course in edX with a dataset containing students' click-level interactions. Using this data, we aggregated video-watching events such as pausing, playing, and re-watching to analyze peaks in activity. A peak occurs when many students are interacting with a video at a particular segment. We adopted a metric, called the video engagement factor, to measure student engagement with the course videos at these peaks. To explore the cause of high video engagement factors at peaks, we retrieved video transcripts at these points of interest with the YouTube API and categorized them using the Latent Dirichlet Allocation. Our implementation of topic modeling allowed us to automatically categorize video transcript snippets into topic models to determine the content that was presented at the given peak in the video.

We discovered the main causes of peaks were transitions in videos, such as switching from the whiteboard to the instructor, and explanations and definitions of new terms. These causes may be attributed to students pausing videos to take notes or transitions occurring too quickly, leading students to re-watch segments. Peak levels increased when topics such as loops, lists, and algorithms were discussed. The points of interest from our analysis can inform instructors of student confusion and improve the video-learning experience for students.

## **Correlation Between Twitter Moods and Market Performance**

### Authors

**Connor Valan, Alexandra Wentzel**

Computer Science Department

Widener University

Chester, PA

### Advisors

**Yana Kortsarts, Suk-Chung Yoon**

Computer Science Department

Widener University

Chester, PA

At its core, the stock market is a place of educated gambling. Investors may spend years perfecting their strategy and they could still not make a profit. However, by collecting a large amount of social media data, it is theoretically possible to determine the general trend of thought across a large population. By understanding this trend of consumer thought, it is possible to apply that information to make more accurate predictions of company performance in the stock market. Twitter provides the perfect social media outlet to collect this data for numerous reasons. The social media site provides a unique relationship between users, simple communication (limited to 140 characters), a high volume of text information, and a very accessible API. Our overall goal for this project is to use the compiled mood data in order to determine a correlation between any changes in mood and the performance of the Dow Jones Industrial Average. In order to accomplish this goal, we must collect tweets from Twitter, analyze them for an indicated mood, and compile the mood data over time. In order to filter through the collected data, we implemented a profile for mood states. This profile categorized our collected tweets based upon a dictionary which we created to gauge the value of each mood. Lastly, we focused on using this data to analyze the trends in mood states and compare those trends to the DJIA.

## **PoIRANK**

### Authors

**William Paustian, Nikolai Peralta,  
Michelle Lu**  
Mathematics & Computer Science  
Ursinus College  
Collegetown, PA

### Advisor

**Akshay Dhawan**  
Ursinus College  
Collegetown, PA

The emergence of large-scale social networks has led to research in approaches to classify similar users on a network. While many such approaches use data mining techniques, recent efforts have focused on measuring the similarity of users using structural properties of the underlying graph representing the network. We identified the Twitter followers of the 2016 presidential primary candidates by using the Twitter API Twython and classified them as Democrat, Republican or Independents. This was done by designing a new, domain specific approach to measuring structural similarity called PoIRANK. PoIRANK computes the similarity of a pair of users by accounting for both the number of candidates they follow from each party and the specific candidates they follow. We were influenced to do this by looking at three other structural similarity measures and how they used in-neighbors (nodes that point in towards the node being calculated in a directed graph) and out-neighbors (nodes that the node being calculated are pointing towards) to calculate their similarity. The algorithm was created by developing a similarity measure and then implementing it using SNAP (Stanford Network Analysis Platform). In order to test our algorithm, we crawled a data set of all followers of every presidential candidate in June 2015 and then ran experiments on a random subset of 10% of that data. We tested our measure against similar algorithms from the literature including SimRank, P-Rank and Cosine-Similarity. PoIRANK outperforms these algorithms since it is more efficient when used on large data sets. This efficiency is due to PoIRANK's ability to calculate similarity of a pair of nodes independent of other nodes. Additionally, the time complexity of PoIRANK is  $O(n^3)$  which is lower than that of P-Rank ( $O(n^4)$ ).

## Cyber Security Issues with Autonomous and Semi-Autonomous Robots

### Author

**Thomas Snider**

Computer Science & Information Technology Department  
Chestnut Hill College  
Philadelphia, PA

### Advisors

**Pamela King, M.S.**

**Susan Ceklosky, M.S.**  
Chestnut Hill College  
Philadelphia, PA

As the autonomization of robotics platforms becomes more prevalent, priority needs to be taken to address security vulnerabilities. Many ethical, legal, and technical issues are directly connected to these security vulnerabilities and it is essential to apply additional cyber security measures to counteract them. In order to analyze the extent and impact of these issues, I conducted an experiment to determine whether the program known as *SkyJack* would permit the successful hacking of a commercially available drone. The *SkyJack* program was used along with a myriad of supporting devices to attempt to hack a Parrot AR Drone 2.0 and fully take control of it. *SkyJack* takes advantage of the Parrot AR Drone's use of Wi-Fi in order to take control away from the user and take over. *SkyJack* works in conjunction with two other assisting programs known as *Aircrack-ng* and *node-ar-drone* to deauthenticate the true user of the drone and gain full access to the controls. The hack was successful and allowed for the full control of all aspects of the drone as if it was still being flown by the original authorized owner.

This experiment showcased just one of the myriad of vulnerabilities plaguing different robotics platforms across the world. The implications of this experiment convey a need for heightened security for autonomous and semi-autonomous platforms by both the user and the manufacturer to protect against breaches of privacy and control.

Within the poster I discuss the widespread nature of robotics security issues and details the steps taken to setup and execute the *SkyJack* drone hacking experiment. The implications of the vulnerabilities with an emphasis on the relation to security problems rampant in the field of robotics are discussed. Based on background research and the results of the experiment, some potential solutions to the problem are discussed.

## **R Language Across Disciplines**

### Authors

**Ahmet O. Sivri, Iago da Silva, Mert Vurgun**

Computer and Information Sci.

SUNY Fredonia

Fredonia, NY

### Advisor

**Reneta P. Barneva**

SUNY Fredonia

Fredonia, NY

Data analytics is very important for decision making and knowledge extraction in all fields of human activities. Unfortunately, not all professionals, especially from non-technical disciplines, have been trained to write elaborate computer programs that would process their data. However, if they were able to present the data visually in an appropriate way, they could use their intuition to draw conclusions and make decisions.

We investigated various software packages for data visualization to propose them to be used in the classroom for student training in non-technical disciplines. This way, the students would acquire knowledge of processing data in their specific field.

In this investigation we took into consideration: the simplicity of use by non-professionals; the cost – possibly free software; the potential of visualizing data in various ways so that it can be used for various purposes; the possibility to run on multiple platforms.

We came to the conclusion that R language and environment are the most appropriate tool for non-specialists. They provide not only statistical calculations, but a variety of graphics techniques. It is open source software, easy-to-use, and supported by multiple platforms.

We developed a number of examples from sociology, history, psychology, sports, business, health, education, and political sciences, where it can be used. For instance, in presidential elections, we can observe a correlation between the age of the poll and the preferences for candidate; in health sciences, we can observe the spreading of an infection on a map; in education, we can compare and assess different educational methods from the test results.

The proposed examples could either be demonstrated in classes of non-technical disciplines, such as political science, health, or education, or be combined in a team-taught general education course, accessible to all interested majors.

## **From AppInventor to Android Studio: A Student's Perspective in Mobile Application Development**

### Authors

**Jonathan Cleary, Nicholas Bielinski,  
Gregory Tinkham, Christopher Nobrega**  
Computing Technology Program  
University of New Hampshire  
Manchester, NH

### Advisor

**Karen Jin**  
Computing Technology Program  
University of New Hampshire  
Manchester, NH

In Computing Technology program at the University of New Hampshire, two mobile development classes are offered that leverage the program AppInventor [1] and the Android Studio [2] respectively. First, we present our experiences and struggles bridging the gap between the lower level class that uses AppInventor and the more advanced class that utilizes Android Studio. AppInventor uses visual blocks to form statements and functions, which makes the syntax of the software. Android Studio is an industrial strength IDE that uses Java to develop native android applications. AppInventor aims at getting students interested in programming and teaching them basic programming structure and logic. However, moving from AppInventor to Android Studio, we experienced the big hurdle in learning the Android framework as well as programming concepts, particularly the correct syntax and use of delimiters. AppInventor strays from certain terminology and conventions, such as not using zero-based numbering for lists and not mentioning classes/objects. It also hides Android framework constructs such as activities and fragments. Therefore, we propose our solution for bridging the knowledge gap between the two courses. We plan to build an application that translates the visual blocks in AppInventor into pseudo codes based on Java. This functionality would help students better understand the language concepts and syntax that are obfuscated using a GUI tool like that of AppInventor. Ultimately, the significance of this approach is to better transition to a high level programming language, as well as to achieve a deeper understanding of the Android framework. We will present the proposed design and idea of such a translator. Our project is open source and we plan to submit to Google Summer of Code project [3].

[1] [appinventor.mit.edu/explore/](http://appinventor.mit.edu/explore/)

[2] [developer.android.com/sdk/](http://developer.android.com/sdk/)

[3] [developers.google.com/open-source/gsoc/](http://developers.google.com/open-source/gsoc/)

**uSafeNH: A mobile app for helping sexual assault victims on college campuses in New Hampshire**

Authors

**Maryjean Emerson, Joshua Young,  
Tucker Cleary, Bonnie Smith**  
Computing Technology Program  
University of New Hampshire  
Manchester, NH

Advisor

**Karen Jin**  
Computing Technology Program  
University of New Hampshire  
Manchester, NH

The percentage of victims that report sexual assault is very low compared to actual assaults. In particular, sexual assault victimizations of college students are likely to go unreported [1]. We want to make it easier for the victims to report their assaults and search for resources. Our solution is to design a mobile app because the use of smartphone apps has increased dramatically in recent years, particularly among college students.

In cooperation with Prevention Innovations Research Center and the Attorney General's Office of New Hampshire, we have developed the uSafeNH mobile app. It is designed to help victims of sexual assault or their friends, family and community members to get help immediately. The app includes resources provided by college campuses, local town, state and national levels. Moreover, uSafeNH app will help the victims to report their sexual assaults.

uSafeNH is designed and developed by us, a team of University of New Hampshire (UNH) Computing Technology undergraduate students. We started the project in the spring of 2015. Through meetings and focus groups on four New Hampshire college campuses (Keene State, UNH, St. Anselm, and White Mountain Community College) we collected data and feedbacks and completed several development iterations. We have implemented total eight releases through iterative design and development process. Our official launch of the uSafeNH app is scheduled in May 2016. In the future we hope to market this app out to other states and offer it to their college campus as a free app. We are also looking to create an iOS version in addition to our current Android version, so that this app will be accessible to more college students.

[1] Rape And Sexual Assault Among College-Age Females, 1995-2013, by U.S. Bureau of Justice Statistics, Department of Justice <http://www.bjs.gov/index.cfm?ty=pbdetail&iid=5176>

## Computer Game Industry as an Educational Path: From Serious Games to E-sports

Author

**Kaê Sarmanho**

Computer and Information Sci.  
SUNY Fredonia  
Fredonia, NY

Advisor

**Reneta P. Barneva**

SUNY Fredonia  
Fredonia, NY

Nowadays, computer games are the most popular entertainment among the youth. Many computer science programs include game development courses. Several reputable universities pioneered with stand-alone programs in game simulation, design, and development at certificate/minor, bachelor, and master's level [1,2,3].

However, serious games, which have as a primary goal educational objectives are not featured in the academic curriculum. One of the reasons for this is that they require knowledge from the subject field for which they are developed and students and instructors in non-technical fields do not have the skills to develop such games.

Another aspect of the computer games is related to playing. Game playing has been considered as incompatible with studying, but recent research indicates that it can actually be beneficial for the learning process in many ways [4].

Finally, e-sports gained popularity to the level that in some countries official e-sport organizations have been founded [5]. There are already universities that propose even scholarships to e-sport athletes [6]. In this work, we propose elements of game design, development and playing to be included in various disciplines of the college curriculum. We give examples in: Adolescence education, Music, Arts, Natural sciences, Business, Medicine, Sport management, and others. We will also explain how data analytics could be used to improve e-athletes performance.

**References:**

1. BS Degree in Games and Simulation Arts and Sciences, Rensselaer Polytechnic Institute, <http://www.hass.rpi.edu/pl/gaming>
2. MS in Engineering in Computer Graphics and Game Technology, University of Pennsylvania, <http://www.cis.upenn.edu/prospective-students/graduate/cggt.php>
3. Minor in Game Design & Development, RIT, <https://www.rit.edu/gccis/igm/minor-game-design-development-gamedd-mn>
4. 9 Ways Video Games Can Actually Be Good For You, HuffTech, [http://www.huffingtonpost.com/2013/11/07/video-games-good-for-us\\_n\\_4164723.html](http://www.huffingtonpost.com/2013/11/07/video-games-good-for-us_n_4164723.html)
5. Korean e-Sports Association, Affiliated with the Korean Olympic Committee, <http://e-sports.or.kr/>
6. Meet America's First Video Game Varsity Athletes, Time, <http://time.com/3756140/video-games-varsity/>, 2015

## Efficient Placement of Photovoltaic Arrays using Genetic Algorithms

Author

**Brock Germinara**

Computer Science

Wentworth Institute of Technology

Boston, MA

Advisor

**Karina Assiter**

Wentworth Institute of

Technology

Boston, MA

The purpose of this project is to solve an optimization problem involving a solar power plant and its location compared to an area requiring power in Boston, MA.

This project involves the use of a genetic algorithm (GA) to optimize the placement of a 1 megawatt (MW) photovoltaic (PV) system. This “solar farm” will be a simulation of a theoretical solar power plant.

Data on solar conditions were calculated within a large portion of New England based on estimates from the Solar and Wind Energy Resource Assessment (SWERA) tool provided by the National Renewable Energy Laboratory (NREL). The solar condition data was comprised of Latitude/Longitude coordinates and their accompanying Direct Normal Irradiance (DNI). Calculations were made using these solar conditions and certain assumptions about electricity transportation to determine the net energy per day (kWh/day) sent to the location in Boston.

The use of this GA resulted in an ideal location for the placement of a 1MW solar farm. In comparison to the starting location within Boston, MA, the GA determined that the optimal location was at a specific Latitude and Longitude coordinate pair which happened to be in the town of North Attleborough, MA.

## **Personal Cloud Storage**

### Author

**Nicholas Wereszczak, Kyle Combs,  
Joseph Bisch, Adrian Caban**  
Computer Science  
University of New Haven  
West Haven, CT

### Advisor

**Frank Breitinger, Bijan Karimi**  
University of New Haven  
West Haven, CT

Security and privacy are important issues nowadays and more and more people are starting to protect themselves. The easiest steps that can be taken are encryption of data, encrypted email traffic, and providing less access to private data / information.

Our research originated from the idea of a private cloud that allows users to control their own data. We allow our users to store their data in a privately owned environment instead of sharing it online, e.g., with Google. The idea is that it works like iCloud or Google's backup service, but with the added security of knowing you are the only one who has access to your data. To meet these goals, we researched specific encryption methods and the means to move data wirelessly in the most secure way possible.

Our project uses a server-client architecture. We utilized a raspberry pi to act as our server/cloud and an android device represents the client. However any computer could be used as the server. As of now, the system only works within a local home/office network to protect the data from being sniffed. Initially, the server and client set up a connection through the android application. Next, the system no longer needs the users interaction. At a defined interval, the information is sync'd between server and client. This transfer of information is protected by the Secure Sockets Layer (SSL) and utilizes the HTTPS protocol. For added security the information is also encrypted on the client side before transmission.

## Engaging Middle School Girls Using CS Games

Author

**Stephanie Eordanidis, Carla Casal,**

**Maitry Chakraborty**

Computer Science, Math Ed

Kean University

Union, NJ

Advisor

**Carolee Stewart-Gardiner**

Kean University

Union, NJ

Technology is all around us. It's everywhere from the vehicles we drive to the classrooms where our children learn. With the augmentation of technological use and innovation, also comes the increasing need for more computer scientists and engineers alike. To keep up with this demand we at Kean University explored the use of games teaching computer science concepts and principles to potentially educate, interest, and engage the youth of today. This study is a sub part of the main research venture funded by the NSF-DRL-AISL research grant (Award Number: 1421806). That main study focuses on educating middle school girls through the use of story in analog and digital games. Our team created 3 analog games, each of which teaches a CS principle and has a story and abstract version. In addition, we took one of those games and generated it into digital format as a standalone PC game. These games were then tested in an eight week after school program we held and facilitated in order to gather statistics and other pertinent data to see whether games teaching CS principles and concepts are more useful with story or without it, to engage, interest, and educate middle school aged females. We utilized pre and post-game assessments, video/photography, facilitator observational notes, and attitudinal, demographic, and engagement surveys to collect our data during the after school program testing. After analyzing the data gathered, we were able to conclude that the games we created, of both analog and digital format, had appealed, interested, and engaged a large portion of the participants. All information and findings gathered will be used to make modifications and fine tune the games and facilitator plans the team had produced. Summer 2016 is set to prepare for presenting and holding workshops showcasing these games with educators.

## **Keep Calm and Carry On: A Voice-Based Medical Prediction Engine**

Author

**Nicholas Grupen, Zachary Rahn**  
Department of Computer Science  
Villanova University  
Villanova, PA

Advisor

**Edward Kim**  
Villanova University  
Villanova, PA

We present a machine learning based mechanism for predicting physiological and behavioral statistics using analyzed voice samples. Our algorithm not only expands on previously researched correlations between the voice and the human body, but it uncovers novel correlations that involve subjective measures such as stress. For students, working adults, and elderly citizens alike, it is very difficult to get daily or even weekly personal health updates. The average person visits their doctor only once or twice a year, which means they go through a routine vitals check – to make sure his/her blood pressure, body temperature, and heart rate is at acceptable levels – rather infrequently. In addition, such tests do not expose potential underlying health issues such as stress, anxiety, or sleep deprivation. Thus, we were motivated by the possibility of providing users a system for predicting health and emotional statistics, at any time, from a voice recording alone. In regard to data collection, each member of the team recorded routine voice samples and labeled them to a set of several health measurements. Voice samples were taken using a script of imitated, elicited, and spontaneous speech levels, which involve the production of elongated vowel sounds, repetitive phrases, and randomized sentences. Health statistics include heart rate, body temperature, blood pressure, and subjective measurements include stress, sleep deprivation, sickness, and voice use. Results from real experiments of various sizes indicate prediction accuracies of 88% for stress, 89% for sleep deprivation, 94% for heart rate, 99% for body temperature, 94% for sickness, 89% for voice use, and 96% and 93% for the systolic and diastolic components of blood pressure, respectively. The full description and implementation features of our project will be presented in detail.

## Comparison of Heuristics in the Bin Packing Problem

Author

**Michael Blithe**

Computer Science Department

Widener University

Chester, PA

Advisor

**Yana Kortsarts**

Widener University

Chester, PA

How to pack bins most efficiently, using as little time and as few bins as possible, is a question frequently asked in many fields, including Environmental Science and Logistics Management where maximizing usage of space is key. The way the items are placed in the bins can be based on a variety of different heuristics. Some of these heuristics include: worst fit, first fit, best fit, next fit, first fit decreasing, and best fit decreasing. Depending on which of these heuristics is implemented, bin packing can either be more space efficient or time efficient.

Computational implementation for each of these heuristics was designed using Python programming language. Python was chosen because of its ability to handle large data sets, while still being efficient. A set of data was randomly generated, and each heuristic was applied to the data. The computational time and the number of bins that were required was logged and compared.

The average speed and the number of bins used was calculated for each heuristic over one thousand trials and compared to the brute force solution. The poster will present the results of the comparison analysis, indicating the best performance heuristic. The poster will also discuss the possibilities to explore more advanced methods to solve bin-packing problem, including genetic algorithms and parallel processing.

## A Comparison and Analysis of Graph Storage Scenarios

Author

**Melissa Iori**

Department of Computer  
Science and Mathematics  
Marist College  
Poughkeepsie, NY

Advisor

**Alan Labouseur**

Marist College  
Poughkeepsie, NY

This project examined some current techniques for storing graph data and the performance tradeoffs involved, measured in query time. Centralized relational databases have been used as the data storage solution of choice for decades, as they are reliable, use intuitive tables to store data, and are based in set logic. However, current trends indicate that “big data” may be better expressed as a graph, a collection of connected nodes. As a graph, data can be analyzed to make new deductions, including evaluating the relative popularity of particular nodes in relation to the entire graph, shortest paths, commonalities, average popularity, and other interesting queries. Graphs can even be stored in relational databases. The database storage structure options tested for this research include the centralized row-based approach, commonly used by databases including PostgreSQL, the centralized column-based database, like HP’s Vertica, the distributed relational database, like IBM’s Netezza, and “native” graph databases, such as G\*. By generating power law / random attachment graphs via a graph generating module, I was able to cast graph data representations into a format suitable for each platform (like SQL, for relational), then run the same queries against the same graph on each platform. By timing these queries, a performance comparison was made, as well as an analysis of the query plan chosen by the particular database. Overall, these results shed some light on the relative performance of today’s lesser-known forms of data storage, and possible ways in which to internally store a graph relationally, which may assist in future considerations for graph storage implementation.

## **RFID Controlled Robot Game**

Author

**Katie Seggerman, Justin Grannis**

Department of Electrical and Computer Engineering  
University of New Haven  
West Haven, CT

Advisor

**Christopher Martinez**

University of New Haven  
West Haven, CT

We present our project of an RFID controlled robot game as part of the senior design course for the department of electrical and computer engineering at the University of New Haven. The project was envisioned as a means to generate interest in the field of computer science at an early age by introducing the logical thinking necessary for programming through a new interactive interface. In this game, a child user can use the physical input of a stamp pad to move a robot across a game grid based on a given storyline.

The design for this project involved three major components: a PC program, a stamp pad consisting of an Arduino, LCD touchscreen, and RFID sensor, and a robot. In order to create an interactive game, we wrote a program using C++ and OpenFrameworks libraries to design a storyline for the user to follow. This story prompts the user for inputs on the stamp pad which reads the combinational touch and RFID value key to recognize desired robot movements. If the sequence of inputs matches the required sequence of code for the robot, then the robot will move accordingly. The stamp pad and robot components communicate serially within the C++ program using USB connection and Bluetooth respectively. This game will teach children programming skills by introducing them to sequential code, routines, and subroutines all through the use of physical interactive components in place of a programming language. Our poster will discuss, in detail, the motivations for our project, its design and implementation, and its desired outcomes.

## **Slicing the Pie: Big Data Processing with Raspberry Pi**

Author

**Adam Doktor**

Computer Information Systems

SUNY Buffalo State

Buffalo, NY

Advisor

**Sarbani Banerjee**

SUNY Buffalo State

Buffalo, NY

Advanced technology has allowed engineers to push the boundaries of computer power. A modern example of this can be seen with the Raspberry Pi (a small, low cost, and palm sized computer). Today it is important to complete meaningful tasks as fast and as efficiently as possible. This research project intends to test the efficiency of Raspberry Pi computers using more data than it normally can handle, thus introducing the term 'Big Data'. Big data consists of data sets that push the limits of the technology that they are contained on. The slicing and dicing of data, commonly known as data analysis can be done with the assistance of software like Hadoop. Hadoop is a tool used to manage large sets of data across multiple computers. It is often used to identify trends and develop predictions based on data. On an industry scale, data is stored on multiple large systems which can be very expensive. This research will look into the possibility of creating a small Hadoop cluster using multiple Raspberry Pi units. This will in return be cost efficient, because it eliminates the use of more expensive and more powerful computers. The primary goal is to maintain a system of devices using distributed processing and storage. Future research will look into extending the cluster, and may experiment using Arduino computers in order to identify which devices are more compatible.

**Communication Therapies with the NAO Robot**

Authors

**Mark Crosby, Bridget Boland**

Mathematics, Computer Science, and Statistics  
SUNY College at Oneonta  
Oneonta, NY

Advisors

**James Ryder**

SUNY College at Oneonta  
Oneonta, NY

**Elizabeth Soltan, M.S. CCC-SLP**

B-T BOCES/Oak Tree Program  
Binghamton, NY

We present our findings on programming applications for the humanoid NAO robot to be used in a classroom for children with Autism. The goal of our research is to program tools to be used by educators to promote communication in their students. NAO is a humanoid robot, from Aldebaran corporation, standing at 58cm tall intended for human interaction. It has been used in areas of research and education. In education for children with Autism, it is used in a program called *Ask NAO*. The robot has thrived in this area of study. In the fall of 2015, we spent a semester learning how to program the different functionalities of the NAO robot in Python using Choregraphe. After gathering sufficient knowledge on the robot's capabilities, we partner with the Broome-Tioga BOCES Oak Tree to conduct our research. A speech-language pathologist there, Elizabeth Soltan, M.S. CCC-SLP, meets with us and shares her ideas of what applications she would like to see from the robot based on NAO's abilities. We then program these applications for her to test and use in the classroom with the students. She then tells us what worked with the students, or what improved communications with them, and what could be improved. These findings will be fully gathered by early May 2016. We will present early findings.

## Computational Model of Surprise Using Narratives

### Author

**Kevin Brown, Gunther Cox, Jared Duval, Brian Grohe**

Department of Computer Science

Western New England University

Springfield, MA

### Advisor

**Brian O’Neill**

Western New England

University

Springfield, MA

Authors craft their stories to evoke emotion from their audience. However, computer-generated narratives typically lack the ability to elicit emotion. In an effort to understand emotion in a computing environment, we develop a model of surprise using narratives. A model of surprise has many applications, such as creating training scenarios for medical fields and evoking human emotion in digital media. Our goal is to use the model to quantitatively calculate how surprising an event is in a given story.

Our model relies on knowledge about how stories typically progress. We begin by collecting stories from people using a web form. We provide a common setting, characters, start scenario, and end scenario. Volunteers fill in the sequence of events that make up the story. Through coding techniques [3], the actions present within each event are extracted. Using a combination of manual [3] and automated [1] techniques, events are used to generate STRIPS operators with sequences of events becoming scripts. These scripts are used to build plan networks [2]. This data structure best represents our model of the events in a story, allowing us to represent frequency—the number of times one event immediately follows another event in participants’ stories. This plan network is a directed graph where nodes represent events and edges contain frequencies. By traversing the graph, measuring frequency, applying STRIPS operators, and retaining a memory of previous details, we can determine how surprising an event is compared to other events.

[1] Boyang Li, Stephen Lee-Urban, George Johnston, and Mark O. Riedl. “Story generation with crowdsourced plot graphs.” In Proc. 27th AAAI Conf. Artificial Intelligence, Bellevue, WA, 2013. AAAI.

[2] Jeffrey David Orkin. “Learning Plan Networks in Conversational Video Games.” MS thesis, MIT, Cambridge, MA, 2007.

[3] Johnny Saldaña. “The Coding Manual for Qualitative Researchers.” Sage Publications, Los Angeles, CA, 2009.

**A Case Study on Modeling Social Network Privacy Policies Using Event-B**

Author

**My Nguyen Tra**

Department of Computer Science and Mathematics  
Dickinson College  
Carlisle, PA

Advisor

**Timothy Wahls**

Dickinson College  
Carlisle, PA

We present a case study on the development of a formally verified PHP web social network application. With the expansion of social media and our presence in the cyber world, the privacy and security of our content on-line have become a great concern. Our goal was to investigate whether we could formally verify privacy policies of a social network using the Event-B method. We chose to build a formally verified social network application for the following reasons: the privacy policies are readily formalized and formally verifying the model with privacy policies shows consistency between the visibility of the user's content and the privacy setting of that content. The project involved building an Event-B model for the application in the Rodin environment, verifying that the user's data is visible to the appropriate group of people in the network using Rodin's automated theorem provers, and translating the model from Event-B to PHP code using the EventB2SQL tool. We also extended the EventB2SQL tool to generate parts of the user interface, reducing the amount of code that needed to be completed by hand. The end-product of this project is an application that resembles a real-world social network. By developing a formally verified software system, we confirmed Event-B's ability to build models in the area of social network privacy policies. Our poster will discuss the motivation for formal development of a social network privacy policies, Event-B modeling and the Rodin environment, the EventB2SQL tool, and the challenges of the project.

## Lowerbounds on the Minimum Metric Online Matching Problem on the Line

Author

**Maximillian Bender**  
Computer Science  
Connecticut College  
New London, CT

Advisor

**Christine Chung**  
Connecticut College  
New London, CT

We consider the online minimal metric matching problem, where we are a priori given a set of servers in a metric space in which a request sequence of equal length will arrive in locations unknown until arrival. As each request arrives, it must be irrevocably matched to a server. Each server can only be matched to one request. The two different goals we discuss for the problem are 1) the *minimum weight* objective, which minimizes the average (or equivalently total) distance between any request and its paired server and 2) the *bottleneck* objective, which minimizes the maximal distance.

This problem has been studied most for the general metric, however significant work remains to be done even in the special case of the line metric. Despite the problem having a very 'simple' description, there is still a large gap between the effectiveness of the best known algorithms and the established lower bound for the effectiveness of any algorithm on this problem. In fact, some researchers consider the line-metric to be the most interesting, citing reasons such as the offline-version being a trivial problem, the relation to the well-studied cow-path problem, and the applications of this problem in web-markets.

We present lower bounds of the competitive ratios for both the minimum weight and bottleneck objectives. For the minimum weight objective, the gap is improved to a lower bound of 2 and a previously established upper bound of  $\log n$  from the algorithm HARMONIC. For the bottleneck objective, the gap in the competitive ratio is improved to a lower bound of  $\frac{n}{2}$  and an already established upper bound of  $2n - 1$  from the algorithm PERMUTATION. Because this shows that the lower bound of the competitive ratio of any randomized algorithm on the bottleneck objective is higher than the competitive ratio of the best known algorithm for the minimum weight objective, we have also proven that the bottleneck objective is a *harder* problem for randomized algorithms than the minimum weight objective on the line.

## **Designing of Inventory Management System Using Open Source Object-Oriented Languages**

Author

**Michael J. Lewis**

Dept. of Computer and Information Sciences  
SUNY at Fredonia  
Fredonia, NY

Advisor

**Gurmukh Singh**

SUNY at Fredonia  
Fredonia, NY

This scholarly research project involves creation of an inventory manager system that employs a remote server built in open-source PHP language to interact with a MySQL database. The client-side part of management system is designed in object-oriented language C# along with windows-based forms to insert the input data. The server-side management system verifies the identity of user with a valid username and password. Upon a matching input information of a pertinent user, server generates a random number that is used as a token. The token is hashed and stored into a table in the database along with a time stamp for initiation and expiration of a session for a client who is going to interact with a restful API. The non-hashed random number is returned to the client-side, which must be provided for the next transaction prior to the expiration time.

The expiration time can be set to any duration in hours or days from creation depending on the needs of the business. The time is all handled on the server-side to ensure that time differences of the client will not affect verification system. The user management system part of the scholarly research project also has different levels of access. This is determined by the user's rank in the users table of the database. Before any administration actions may be implemented like adding or removing users or editing other users until the access level is verified.

The server returns a JSON encoded associative array to allow it to be used with any client-side language that has the ability to work with JSON. The inventory part of the system is built in a modular structure so that it is separate from the user verification section of the server. This facilitates adaptability to allow the verification system to be employed with different inventory systems. The inventory system designed in this research project was successfully able to track down the data on communication networking equipment. This system further facilitates the ability for networking technicians to track down the amount and type of equipment they have on hand in the inventory.

## Securing Common Software Vulnerabilities

Author

**Skyler Ng**

Department of Mathematics, Computer Science, and  
Statistics

St. Lawrence University  
Canton, NY

Advisor

**Lisa Torrey**

St. Lawrence University  
Canton, NY

I present my experience studying cybersecurity for my Senior Year Experience (SYE) Project in the Department of Mathematics, Computer Science, and Statistics at St. Lawrence University. The goal of this project is to understand and prevent software vulnerabilities due to programmer error. Amongst these software vulnerabilities, the ones I focused on are SQL injection, buffer overflow, and race conditions. An SQL injection occurs when malicious SQL commands are forced into an SQL database query via a web page input. This is a concern with database systems, especially since it involves dealing with data, which is important whether we are dealing with a company's data or a user's information. A buffer overflow occurs when data is written beyond the allocated bounds of a data structure. The exploit of such a vulnerability can lead a program to execute malicious code or to crash completely. A race condition occurs when there are multiple threads working with a mutual variable, such that the sequence of execution between all threads makes a difference in the result of the variable. An exploit of this type can lead to inconsistencies in variables, such as a person's current account balance. Through creating examples of programs that suffer from these vulnerabilities, I explored how these attacks work in order to understand how to guard against them.

## Trade Efficiency Problem

### Author

**Tim Palmer**

Computer Science  
Connecticut College  
New London, CT

### Advisor

**Christine Chung**

Connecticut College  
New London, CT

We propose and consider a game with  $n$  players with  $m$  items to be distributed evenly among the players, where  $m$  is a multiple of  $n$ . Each player has a descending preference ordering over the items, which corresponds with that item's *value* to the player. Each player wishes to maximize their *score*, the sum of the values of the items they possess. The basis of this work is to examine the efficacy of and efficiency of a variety of greedy exchange algorithms in order to model real world trading scenarios. We study two natural algorithms:

- Trade algorithm. Items are randomly distributed amongst the players. All players take turns trying to make trades with other players until there are no more possible trades. A trade will only be possible if scores of both trading players have a net gain after the trade, and the player whose turn it is will take the highest payoff trade possible.
- Draft algorithm. Players start with no items. All players take turns picking their highest valued item remaining from the set of all items.

We also consider the following metrics for measuring performance.

- Utilitarian. The first metric is to find the total value of an optimal distribution, and then see what percentage of this optimal value our algorithms achieved. In order to determine OPT, managed to apply the Hungarian 1 to 1 matching algorithm to the many to one matching problem that is our game. Once the trading or drafting has been stabilized, the sum of all of the players' scores are totalled, and then divided by the sum of players scores given in the OPT distribution of items.
- Fairness. Taking ideas from fair division, one approach to fair division was comparing players' score in contrast to their competitors. Another interesting measure of fair division in this problem was to analyze how much each player values the items of another player over their own. The performance of the Trade, Draft, and OPT allocations from a utilitarian standpoint are not necessarily equivalent to their performance from a fairness standpoint.

## **Business Intelligence Dashboard**

### Author

**Kyle Pinder, Kevin C Gonzalez, Victoria Oliveira, Louis Notarino, Cinthya Grajeda**

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

### Advisor

**Frank Breitingner, Alice Fischer**

University of New Haven  
West Haven, CT

DocuWare, one of the world's largest document management software companies, offers solutions to over 125,000 users in the form of web-based Electronic Content Management (ECM) software. Their software allows users to easily store electronic documents as well as digital replications of paper documents which can later be found and opened with a simple filter. While the system is meant for searching through documents; it does not give the users any insight into the data aggregated within the system, nor does it show the data effectively on a small scale device.

For this project, we developed a dashboard to visualize the data in the system (i.e., through graphs). The final result is a web application and hence completely independent of any platform. Our implementation can be divided into a backend developed in C# that connects with the DocuWare's API; as well as a frontend that is composed of HTML5, CSS3, and JavaScript.

**Backend:** To create our application, we first had to gain access to DocuWare's system. Once we set up an account, we populated our own file cabinet with a multitude of documents. This allowed us to test how our system connected to DocuWare's servers via their API. A successful connection allowed us to see our data within the file cabinet. Before the document is uploaded, users are prompted to fill out specific input criteria. This data is linked to the document and is used to create graphs.

**Frontend:** The frontend consists of multiple dashboards which allow the user to easily navigate between filtered results. For usability, we implemented a variety of tools within each graphical dashboard such as undo/redo, copy, and save. Our implementation permits users to view a more detailed and manageable depiction of their filtered results.

## Seeking the Foundations of Artificial Neural Networks for Image Recognition

Author

**Sean Godard**

Computer Science  
St. Lawrence University  
Canton, NY

Advisor

**Lisa Torrey**

St. Lawrence University  
Canton, NY

While computers excel at solving most complex logical problems with greater speed and accuracy than their human counterparts, one area that humans have continued to dominate is within the field of image recognition. While humans perform image recognition tasks with little effort, a computer given a similar problem faces a significantly greater computational challenge. As computer vision and artificial intelligence applications have grown in numbers over the past years, with the arrival of such things as self-driving cars, this challenge has become even more crucial to solve, as humans will literally place their lives in the hands of a computer's ability to perform these and similar computations within a very small margin of error. Perhaps one of the most riveting attempts to bring computers up to speed with their human counterparts at these tasks is to mimic the elasticity of the human brain through the creation of an artificial neural network. One of the advantages to this model is that it tends to give solutions quickly while remaining flexible to "fuzzy" inputs, however, they are often quite slow to train. I sought to build an artificial neural network from the ground up in hopes of discovering where the process might be optimized. To judge the efficacy of my implementations, I chose to use the MNIST Handwritten Digits Dataset benchmark, which is a dataset containing handwritten digits, often used to compare different artificial intelligence algorithm's performance. Greater analysis of neural networks, exploration of preexisting methods of optimization, and the performance of my implementations will be discussed.

## **AI Bracketology**

### Authors

**Adam Pianka, James O'Donnell,  
James Colla, Tony Tran, Cody Bond**  
Western New England University  
Springfield, MA

### Advisor

**Brian O'Neill**  
Western New England University  
Springfield, MA

Every March the National Collegiate Athletics Association hosts the largest college basketball tournament in North America. Thousands of fans attempt to construct the perfect bracket, predicting winners post-selection for that year's tournament. Our research focuses on predicting which teams will be selected initially to compete in the tournament and which teams will be paired against each other in the first round. There are two modes of entry into the tournament: 32 teams admitted by winning their conferences, whereas the other 36 "at-large bids" are chosen from the set of eligible teams by the tournament selection committee. Only teams seeking at-large bids need to be classified as in or out of the tournament. We divided this project into three phases. First, we need a model for predicting which teams would participate in the tournament. Second, each team selected to the tournament must be assigned a seed. A seed is a team's preliminary ranking within the tournament, which is used in bracket construction to prevent highly ranked teams from meeting until later in the competition. Third, we need a model to generate the tournament bracket.

In order to predict the teams receiving at-large bids, we use a neural network framework [1]. To train our network, we gathered data about selected teams from past tournaments back to 2005. We are approaching completion of phase one of our project, classification of at-large bid selection. Our neural network has a validation error of 11%, meaning we are on our way to accurate classification of teams receiving at-large bids. In the subsequent phases, we plan to predict seed assignment and the initial bracket structure using techniques similar to the first phase.

## **References**

[1]J. Heaton, "Encog : Library of Interchangeable Machine Learning Models for Java and C#", Journal of Machine Learning Research, vol. 16, pp. 1243-1247.

## **A Location-Based Algorithm for Encrypting and Decrypting Text**

Author(s):

**George Panagopoulos**

Department of Mathematics and  
Computer Science  
Quinnipiac University  
Hamden, CT

Advisor(s):

**Mark Hoffman**

Department of Mathematics and  
Computer Science  
Quinnipiac University  
Hamden, CT

All companies nowadays spend enormous amounts of money in computer and network security. Passing large amounts of unencrypted personal information via the internet is not safe and unauthorized people can easily access your data at any time. The purpose of my project is to assure secure communication by text between users. If you encrypt your information, the odds of getting attacked successfully are very few. This is accomplished by using polyalphabetic substitution and location based encryption to encrypt the communication between the sender and recipient. For the location based encryption my project uses the Cartesian coordinate system and the Distance Formula to find the distance between the sender and recipient. For the encryption and decryption process I will be using a symmetric key that will use a single shared key between users. In most cases I prefer to use symmetric keys, because of the time efficiency. In order to generate the symmetric key, the sender will need his geographic coordinates and recipient as well. The main reason why my project uses the Geographic Coordinate System and the Cartesian coordinate system, is because it enables every location on Earth to be unique and be specific on a two-dimensional map. The key must remain secret during the entire communication process. Logan Scott and Dorothy Denning developed a location based encryption that requires a device such as a GPS receiver, satellite or radio frequency position system. My encryption and decryption process does not require a device in order to encrypt decrypt a message. By following my encryption and decryption method I can assure that parking lot based attacks are not effective. The algorithm is mostly build for short text messages, but that does not mean it cannot be used for larger text messages or any sort of applications and websites.

## Visualizing Trends in Top Ten Lists

Author

**Grace Ulinski, Zevi Rubin**

Computer Science

Wheaton College

Norton, MA

Advisor

**Michael Gousie**

Wheaton College

Norton, MA

We present our experience working with information visualization in the Computer Science Department at Wheaton College. The goal of the research was to study robust data visualization techniques for top ten ranking data over time. Previous research had been done using Flash ActionScript code, but browser incompatibility has made the language outdated and thus the app impractical. We explored alternative visualization strategies for multi-variable visualizations, but ended up sticking with the line graph format. The goal of this visualization was to display a data set for a top 10 ranking group over time. For testing purposes, we used the previously collected movie star data from 1927 to 2013; however, the app is designed to accept any data set in this format. The application was created at the request of Professor John Grady of Wheaton College's Department of Sociology, who studies social trends of various data sets over time. Using JavaScript, XML data is read in and processed for graphing. A variety of JavaScript functions are used to display the data based on chosen specifications. D3, a JavaScript library whose main purpose is to supplement data visualization, was one of the most important tools used in the project. With D3, the data lines and points could be easily plotted onto a basic graph. CSS and HTML are used in conjunction with JavaScript to display features of the graph such as color and line thickness, and to create dropdown menus. Our poster will discuss our reasoning for the reimplementation in JavaScript, specific functionality of the web app, and future work for this research. The full description and implementation features of the app will be presented in detail.

## Automated Analysis of Eligibility Criteria in Clinical Trials

### Authors

**Alexis King, Cristina Diaz**  
Division of Computer Science,  
Mathematics and Science  
St. John's University  
Queens, NY

### Advisors

**Bonnie MacKellar, Christina Schweikert**  
Division of Computer Science,  
Mathematics and Science  
St. John's University  
Queens, NY

We aim to design and implement a prototype patient-oriented clinical trial search system. This clinical trial search system will aid patients and caregivers in the process of searching for a clinical trial in which the patient may participate. Patients and providers are often concerned with interventions involved in the trial, the possible side effects, dangerous drug interactions, as well as the patient's eligibility. Clinical trials contain eligibility criteria, which are conditions that qualify or disqualify patients from the trial. Eligibility criteria are written in free text. This format makes sifting through each individual trial a challenging task. Extracting patterns and concepts from the free text enables us to create programs to analyze and process many clinical trials to find a good match, as well as identify conflicts between trials [1]. We analyzed several representative clinical trials by hand to identify the types of structural patterns and interactions that may be found. We are using this information to parse eligibility criteria in order to develop a semantic representation. We have compared the hand-analyzed trials to an existing list of patterns created to analyze the criteria in breast cancer trials [2]. We have developed rules using UIMA RUTA to automatically identify these patterns in a body of pediatric oncology clinical trials. In this analysis we can highlight the differences between the eligibility criteria in breast cancer and pediatric cancer clinical trials.

### References

- [1] B. MacKellar and C. Schweikert, Analyzing Conflicts between Clinical Trials from a Patient Perspective. Proceedings of the IEEE 17th International Conference on Health Networking, Applications and Services (Healthcom), 2015.
- [2] K. Milian, R. Hoekstra, A. Bucur, A. Ten Teije, F. van Harmelen, J. Paulissen, Enhancing reuse of structured eligibility criteria and supporting their relaxation. Journal of Biomedical Informatics 56 (2015) 205-219, 2015.

## Attentional Costs of Different Notification Types

Author

**Alyssa Klein**

Computer Science  
Connecticut College  
New London, CT

Advisor

**James Lee**

Connecticut College  
New London, CT

The study of Attentional Costs of Different Notification Types looks at text message notifications. We aim to see if a notification with a message preview is more distracting than the notification that only shows the sender's name and its effects on the user's ability to retain information. Users are motivated to change their notification settings due to their environment-turning the sound off when in class or a meeting, or for their privacy-not wanting prying eyes to peek at their latest text message that popped up on their lock screen.

Multitasking and turning to technology is deeply ingrained in our habits and gives us a sense of pleasure or reprieve. The study, *Facebook and texting made me do it: Media induced task switching while studying* by Larry D. Rosen, et. al. analyzed how different types of stimuli affect our ability to multitask. The study found that "technology breaks" satisfy students' needs for short and long term cognitive emotional rewards which increases learning opposed to transitional media driven multitasking habits. With the paradigm shifting from the utilitarian framework to the experiential quality framework in Human Computer Interaction, computer scientists are increasingly more responsible for its impact to the end user holistically rather than just in functionality.

We plan to conduct a user study to analyze distractibility, impulse control and comprehension with varied notification types. In the study, the participant goes through a series of short readings with one question after each reading. During the reading segments the participant is interrupted with text messages from unknown numbers with everyday content. Evaluation will be based on participants' pre and post experiment survey, number of correct answers and if they opened/responded to the messages. The implications of our habits being manipulated based on notification type will inform future implementation of control options.

## Increasing STEM Engagement with Mobile Applications

Author

**Alexa LaMontagne**

Mathematics and Computer Science

Ursinus College

Collegeville, PA

Advisor

**April Kontostathis**

Ursinus College

Collegeville, PA

With the rising use of technology to supplement classroom instruction and the widespread prevalence of9 cellphone usage among students, Kizzle aims to bring the two together, giving students the ability to improve their academic performance through a game-like app right on their phones. Students will be able to assess their understanding of subject materials through short bursts of questions, the content changing weekly to match pace with the classroom. Unlike existing applications, Kizzle is centered around learning goals for the subject, rather than a specific textbook, and the focus is to achieve those goals rather than finishing a set number of questions. In addition, the app will have an ability to send alerts to the instructor, giving them detailed metrics on the individual and class level, allowing them to both reach out to struggling individuals, or alter class lesson plans accordingly. The focus this year is in developing Kizzle for the iOS platform. Beginning with a database and some sample data, we are creating an app that not only fetches from and inputs to a database, but incorporating color-theory and popular game features to make a “studying application” fun and desirable to use. The app is using Apple’s new Swift language on XCode as well as standard PHP and MySQL for the backend and is being designed with the latest iOS version (iOS9) in mind. The poster will discuss the student-centered initiative behind Kizzle, the challenges in creating an app for the iOS platform, and the future plans to implement the app in Computer Science and other STEM-subject classrooms. The full description and implementation features of the app will be presented in detail.

## Cooperative and Uncooperative Voting in Doodle Polls

### Author

**Danya Alrawi**

Department of Computer Science

Connecticut College

New London, CT

### Advisors

**Christine Chung**

Connecticut College

New London, CT

**Barbara Anthony**

Southwestern University

Georgetown, TX

Doodle is an online scheduling tool that helps event organizers to decide on a time slot for a meeting with a group of people. The organizer selects a set of possible times for the meeting and sends a table of the time slots as a poll to a group of attendees. Each attendee selects the time slots in which they are available, the poll recommends the time slot with the most ‘yes’ votes to the organizer. We assume that for each time slot, each attendee has a utility for that time slot, and a default threshold  $0 \leq t \leq 1$  where whenever their utility is  $t$  or above, the attendee selects ‘yes’ for that time slot. The social welfare of a time slot is the sum of the attendees’ utilities for it. In this research, we focus on the effect of reporting availability based on a different, private threshold that is either above or below the default. In other words, what happens if a participant, perhaps in an attempt to be cooperative, says yes to a slot with valuation less than  $t$ , or instead, perhaps “selfishly” says no to a slot that is above  $t$ . We examine the effect of this behavior on social welfare. Previous research has shown that voters are likely to approve very popular and very unpopular slots. Our preliminary findings show that uncooperative behavior may improve welfare more than cooperative behavior. In other words, voting based on a higher private threshold can improve social welfare, while voting based on a lower one can harm social welfare.

## **Real Time Virtual Window Simulation**

Author

**Yifan Lu, Christopher Giri, Alexandra Howells**

Computer Science  
Connecticut College  
New London, CT

Advisor

**James Lee**

Connecticut College  
New London, CT

This project involves real-time manipulation of a live video feed with user head-tracking data to provide a viewing experience similar to that of looking out a window. While concepts related to head tracking to enhance viewer experience have been considered in the past, we seek a comprehensive approach with the added integration of live video. With this integration, we hope to not only push the technical boundaries of this problem to their logical ends, but also to engage with the ways this technology can impact and potentially heighten user experience, providing a new way of conveying information in public places. A "Virtual Window" will finally be built in the hallway of our computer science department to display latest information related to computer science. Moreover, this project can be applied and referred in many other computer science research projects and topics such as game design, computer graphics and human interaction, etc. By situating our problem within the confines of live video footage, we are forced to use more optimized solutions that minimize latency to allow for multiple levels of video manipulation in real time, such as the use of general purpose GPU programming to correct for barrel distortion in a wide angle camera lens. In our efforts to maximize realism, technologies used include a high definition network video camera and the second generation Microsoft Kinect sensor. Meanwhile, we adopted a low-pass filter algorithm to eliminate the noisy signals we get from the Kinect which enhanced the stability and accuracy of our tracking system.

## **Pain Management: Development of a Formally Verified Android Application Using EventB2SQL**

Author

**Graham Williams**

Department of Computer Science and  
Mathematics  
Dickinson College  
Carlisle, PA

Advisor

**Timothy Wahls**

Dickinson College  
Carlisle, PA

This project is a case study on formal verification of an Android application used for pain management. When it comes to healthcare applications, the consistency of the application is crucial, as it may affect the wellbeing of the user. The application will check for trends in the user reports of pain, stress, etc. and notify the user accordingly. The model that specifies these trends is defined in Event-B using an Eclipse based IDE called Rodin. We used the automated and interactive theorem provers built in to Rodin to verify that the model always gives the user consistent information. The Event-B model is translated from Event-B to Java using the EventB2SQL tool. The Java class generated from the model forms the core of our Android application.. We have recently extended the EventB2SQL tool to generate Android user interface components, further facilitating the development of our application. Since the model is based on a formally verified Event-B model, we can say with confidence that the application will not output any harmful or misleading information to the user. Our poster will further elaborate on the research we are doing in Event-B, EventB2SQL, and the advantages of formal verification for a healthcare application.

## **Simulating Blackjack for Maximization of Success Rate through Card Counting**

Author

**Elizabeth Lawson-Keister**

Mathematics

SUNY Geneseo

Geneseo, NY

Advisor

**Homma Farian**

SUNY Geneseo

Geneseo, NY

I am presenting my research for the Mathematics Department at SUNY Geneseo. The goal of this research is to analyze the game of blackjack and create a program that will find the maximal success strategy for the player by using a card counting scheme. The project has two phases. The first is to create a working blackjack simulator in Matlab. This would support one player against an automated dealer. Keeping track of the dealt cards and the changing probability of winning the game will enhance the chances of the player. The second part is to run simulations using different strategies to maximize success rate for the player. I have come to the conclusion a player is more often successful with a strategy of "high risk, high reward" on lower counts. However, at higher counts the player should avoid hitting above a 60% bust rate. Interestingly the game is designed so that the player is at a disadvantage and his or her success rate will never go above 50%.

## Digital Dirt: Novel Terrain Error Metrics

Author

**Jacob Wilkins**

Department of Computer Science

Merrimack College

North Andover, MA

Advisor

**Christopher Stuetzle**

Merrimack College

North Andover, MA

In this work, we present and visualize a series of novel metrics for measuring distances between terrain datasets. These metrics are based on the hydrology network of a terrain surface, and measure statistical quantities such as river junction flow balance, the amount of meander displayed by river segments, and the relative importance of individual pixels that are part of the network. By the nature of the collection procedures of terrain data, it is prone to error especially with regard to surface hydrology. Realizing and evaluating that error is essential to understanding the significance and reliability of the information obtained from analysis of the dataset. Traditionally, error metrics have included root means squared error (RMSE) and slope surface error (SSE), which involve overlaying two datasets and comparing the heights between pixels and the angles of the surface normal, respectively. However, these metrics suffer a great deal from small errors (such as a row or column offset of 1) and do not inform the smaller scale terrain calculations based on important statistical analyses, such as hydrology and associated watershed networks. We present the novel statistical quantities pixel load, junction balance, and river meander, and integrate them with standard error metric calculations in a way to create a novel family of terrain error measurements. This is part of a larger project by Dr. Christopher Stuetzle to investigate various novel representations of terrain surface data. Because a primary focus of this work is hydrological accuracy of terrain surface procedural generation, it requires more robust error metrics than RMSE and SSE, which this work helps provide.

## Space Invaders: Software Engineering Learning Experience

### Author

**Trish Le, Brenda Franca, Christian Ramos**

Computer Science

Clark University

Worcester, MA

### Advisor

**John Magee**

Clark University

Worcester, MA

We present our experience in learning software engineering techniques through a project-based learning approach by creating a 2D game from the ground up: a colorful and playable recreation of the beloved Space Invaders. Our project provides experience in all phases of the software development lifecycle and provides valuable group-project and communication experience. Motivation behind this idea was the novelty of being able to make a workable game not only for the learning experience but also for other to play the game and provide feedback.

Our team is constructing the game engine, the collision detection, physics simulation and animations using Java. We incorporated our previous knowledge of programming languages in choosing our tools and platform while we continue to research the work needed to make a 2D platformer game. Our experience so far has brought a new appreciation of the effort and long work it takes to bring a concept through the entire development lifecycle to make a game.

We are using github to keep track of our efforts on the project. The main software processes incorporate both agile development and iterative frameworks. Our group process uses frequent meeting in peer to peer settings and weekly group meetings to discuss the current user story and tasks for the current sprint as well as for the next sprint.

We plan to evaluate our product using User Experience (UX) methods so that we can gain further experience in another area of the practice of software engineering. Our presentation here will include quantitative and qualitative evaluations of our learning in the software engineering process as well as the end product. We plan to have a playable demo to showcase our work.

## **An Android Application Suite to Supplement Elementary Math Education**

Author

**Larry Passantino**

Computing Sciences

The College at Brockport

Brockport, NY

Advisors

**Vishal Anand, T. M. Rao**

The College at Brockport

Brockport, NY

STEM education is vital to our future. STEM occupations are some of the fastest growing jobs in our country. Mathematics is a strong focus of the current elementary school curriculum. Using technology to enhance mathematics education is therefore a step in the right direction. With this as our focus, the purpose of this project was to develop a suite of Android applications to improve the mathematical skills of elementary school children. As our audience are children, we determined that the apps must run on a mobile device and have attractive and easy-to-use interfaces, provide drill exercises, hints and positive reinforcement when they are correct. We had to ensure that anything happening on the screen should be interesting enough to keep them engaged, requiring a balance of simplicity and challenge. We decided to use the Android platform because of its popularity. The work towards developing the apps began with the author learning the Android development tool. We have developed several apps, each dealing with a different mathematical concept and consolidated them into a single suite. In a typical interaction, the student chooses a subject from the main menu and the system presents a problem to solve. The student selects a possible answer and the system responds with an evaluation of the answer. This suite can be used by students to sharpen their mathematical skills and by the teachers as drill in the classroom. The suite currently contains apps that teach addition, place value, fractions (identifying, adding, subtracting), decimal / fraction equivalence, number ordering. We have demonstrated these applications to a group of children at Terry Taylor Elementary (Spencerport, NY), which included the author's daughter. Preliminary experiences show that the children found the apps interesting and kept them engaged. Further studies are needed to study the usefulness and efficacy of these apps.

Poster 5048

## **Pruning Neural Networks in the Neural Evolution with Augmented Topologies (NEAT) Algorithm**

Author

**Andrew Miller, Peter Finta**

Computer Science  
Widener University  
Chester, PA

Advisor

**Adam Fischbach**

Widener University  
Chester, PA

Simulating the human brain on a computer is a task sought out by computer scientists for over 60 years. One of the most common methods for accomplishing this task is to create a neural network, a collection of artificial neurons which can be trained to solve a task on a computer. Neural networks can be found all over, with applications ranging from image recognition to autonomous vehicles to gaming. One specific version of an algorithm which implements neural networks is known as NEAT; neural evolution with augmenting topologies. The NEAT algorithm was invented in 2004 as a solution to the issues of fixed topology algorithms which precede it. With fixed topologies, neural networks were more prone to getting stuck at local optima where the generated solutions must diverge from the global optimal solution before the global optimal solution can be generated. In NEAT, rather than deciding the shape of the network at launch, networks begin with the minimal valid size and growth occurs through mutations in a genetic optimization algorithm, which helps the system adapt to the complexity of the problem and reduce the time spent at local optima. One problem with implementing NEAT however, is how the actual policies for reproduction and mutation are vague. For example, not all possible network topologies are valuable for solving the problem, such as networks which lack the necessary connections to function or contain duplicate structures. Limiting networks to ones which are valuable can be a costly computation, and the NEAT algorithm does not specify if such an operation would be beneficial. Our goal is to determine if limiting the possible topologies of generated networks would decrease the number of generations required to generate the optimal solution to most problems enough to justify the increased calculation time per network. Accomplishing this meant writing two versions of NEAT and testing them with common benchmarks such as the XOR and pole balancing benchmarks. In addition to the benchmarks, we will also explore some additional applications of the machine learning algorithm such as handwriting recognition and game character control.

## **MiddGuard: A Framework for Collaborative and Extensible Visual Analytics**

Author

**Dana Silver**

Computer Science  
Middlebury College  
Middlebury, VT

Advisor

**Christopher Andrews**

Middlebury College  
Middlebury, VT

MiddGuard is a web-based visual analytics framework designed for customizable, collaborative investigations. I initially developed MiddGuard, working independently on the implementation, after collaborating with Professor Andrews on the VAST 2014 Challenge. Since then, Professor Andrews and another student researcher used MiddGuard to create an award winning submission for the VAST 2015 Challenge. Taking into account the difficulties the pair had during VAST 2015, especially that of processing data outside the framework, Professor Andrews and I brainstormed the high-level components that make up the next version MiddGuard. I am presenting my independent work on the low-level design and implementation of this version. MiddGuard provides a framework to ingest, process, and visualize data in a way that complements the analytic process. The core of MiddGuard is a graph composed of nodes, each of which performs an analytic operation or creates a visualization of the data. Investigators alter and extend the graph throughout the investigation based on previous analytic operations and views of the data. These operations replace the work external to MiddGuard that the team did to process data during VAST 2015. Developers contribute the graph's nodes, which perform analysis that is agnostic of the data source and destination. MiddGuard orchestrates the nodes' inputs and outputs and provides easy to use APIs to access the data within the node. It facilitates synchronous and asynchronous collaboration by keeping graphs, nodes, data, and visualizations in sync between connected investigators and stored in a central database. The framework includes a chat-like system for annotations. MiddGuard's abstractions for dealing with data and extensibility make it a powerful framework for developers and investigators in visual analytics.

## Engaging Students with Adaptive Learning

Author

**Philip Winchester**  
Computer Science  
Connecticut College  
New London, CT

Advisor

**James Lee**  
Connecticut College  
New London, CT

Adaptive e-Learning is a field that is growing more prevalent yearly. It's an educational method that is focused on computers customizing the class material to individual students. With extra research and development time a teacher can produce a tool that can adapt to each student's individual learning styles. By focusing on how the student learns and retains information the system will offer intelligent hints, leading the user to a better understanding of the subject. Keeping a user from becoming overwhelmed by a subject will ensure their continued participation and learning throughout the process. This style of system is generally designed towards higher education and specialized skills, which often utilizes outside technology to allow for more user information. We can develop a system similar to this that targets young students who are interested in Computer Science. The system will be a lightweight adaptive application and will focus on how the user interacts with a puzzle, not on how they interact with the screen. The application will be developed for use on iOS as a mobile app that will guide students through several Computer Science topics. The system will try to emulate the student's zone of proximal development and merge it with the three models of adaptive learning. By following the model of "Hour of Code" and adding in adaptive e-learning features we will create a system to help young students engage with computer science such that they feel supported. This project aims to show the viability of lightweight adaptive systems in educational apps to aid students engage more thoroughly with the material.

[1] Code.org. "Every Child Deserves Opportunity." Code.org. Zenedesk, 2015. Web. 04 Mar. 2016. <<https://hourofcode.com/us>>.

Poster 5051

**Bitcoin: Advancing Acceptance and Usability**

Author

**Tom Conlin**

Computer Science  
Connecticut College  
New London, CT

Advisor

**Christine Chung**

Connecticut College  
New London , CT

## Improving Navigation and Searchability in App Inventor

### Author

**Jacqueline Young, Cecellia Tsui**

Computer Science Department

Wellesley College

Wellesley, MA

### Advisor

**Franklyn Turbak**

Wellesley College

Wellesley, MA

App Inventor is a blocks-based programming environment that empowers users to create Android mobile apps. Programs are constructed by assembling visual blocks resembling puzzle pieces, which helps to eliminate syntactic errors and emphasize high-level programming abstractions. However, blocks take up a significant amount of room, often cluttering the workspace and making it challenging to navigate a large program. The main way to peruse blocks code is to drag a relatively small rectangular viewport over a large 2D workspace. Browsers provide a few additional navigation tools that are not well-suited to blocks programming. In our project, we implemented searching and zooming tools tailored to the blocks environment.

The browser's default Ctrl-F searching mechanism is not very helpful in a 2D blocks workspace because (1) the highlight color does not stand out amongst the colorful blocks, (2) it does not center matching blocks in the workspace, and (3) it does not handle matches inside collapsed blocks. Our new search mechanism solves all these problems. Non-matching blocks are greyed out, highlighting the brightly colored matching blocks. Arrow buttons and keys allow for navigation through searched blocks, which are centered in the workspace. As users navigate through blocks matching the search term, collapsed assemblies with matches expand and re-collapse as necessary.

The browser's default magnification mechanism scales the whole browser window and not just the blocks workspace, so it is not very useful for navigating through a large blocks program. Based on their experiences with online map navigation, users expect to be able to select a point in the workspace and zoom in and out around this point. Our new zooming feature allows users to select a particular block and scale the workspace around this block.

## **Predictive Link Following for Accessible Web Browsing**

Author

**Jiri Roznovjak**

Mathematics and Computer Science

Clark University

Worcester, MA

Advisor

**John Magee**

Clark University

Worcester, MA

We present a human-computer interaction research project intended to help people with severe disabilities use the web. Point-and-click interfaces present users with the task of positioning a pointer over a user-interface element, and then performing a selection operation. Users with low dexterity, involuntary movements, or tremors; or users of alternative interfaces such as accessible joysticks, trackballs, or computer-vision based interfaces can have difficulty with this interaction model while navigating web pages. Small links may be difficult to click, and cluttered pages can result in following an incorrect link. We design a mathematical model that based on contextual data attempts to predict what the next user action is going to be. This approach can be used for web browsing when it intercepts mouse events and instead follows links based on analysis of the user's mouse movement behavior. This way the link that the user intended to click is followed, even if the user did not click on that link. Our system can be integrated into web page scripting on the client side to provide ease of use without needing to install separate software. In a preliminary experiment with a computer-vision based mouse-replacement interface Camera Mouse, users were able to click particularly small targets more accurately with our system compared to using the Camera Mouse alone.

## **A Model and Cross Country Analysis of Multitasking with Technology in Academic Settings**

Author

**Yangyan Dong**

Computing Science Department  
The College at Brockport  
State University of New York

Brockport, NY

Advisor

**Mehruz Kamal**

Computing Science  
Department  
The College at Brockport  
Brockport, NY

In today's digital world, 'multitasking' has become a norm with college students across the world. Students are seen to be checking their email, logging onto social network sites, playing online games, sending text messages, and "listening" to in-class lectures. How effective are they at simultaneously managing all these activities? Is this hindering their ability to do their job well? These issues were the underlying motivation for our study. Our research was based on a previous study that developed an overarching model of multitasking that identified factors in various contexts. Our current study however, was focused primarily in the academic setting at the College level. The research questions we set out to investigate through our study were: *What are the factors that cause, and, are impacted by multitasking in academic settings? What role does technology-based factors have to play in the same context?* To address these questions, we carried out our study in two phases. The first phase involved refining and re-conceptualizing the earlier model of multitasking to be applicable only for the academic setting. This involved extracting relevant factors from extant literature and integrating them into a new conceptual model. The second phase involved conducting experiments involving the factors from the new model. Experiments were conducted at two colleges – one in China and one in the United States. Our contribution from this study firstly, is in the development of a thorough conceptual model of multitasking in the academic setting, and secondly, in the comparative analysis between the results from the two countries. The results of our analysis will have implications for better instruction design and greater student engagement in the learning process.

## Applying Novelty search to the construction of Ensemble systems

Author

**Hieu Le**

Department of Computer Science

Dickinson College

Carlisle, PA

Advisor

**Grant Braught**

Dickinson College

Carlisle, PA

Ensemble methods are widely applied in classification problems. In ensemble methods, votes from multiple classifiers are combined to overcome the possible deficiency of any single classifier. One important question is how to construct an ensemble system with individual classifiers that vote differently to avoid incorrect results by the ensemble.

Novelty Search is a recent approach in evolutionary algorithms in which networks evolve based on the differences in the behavior of individuals, rather than strictly on the objective fitness. This research applies the novelty search approach to evolve neural network classifiers to construct a pool of candidate classifiers for the ensemble. The ensemble then is built from the pool by adaptive boosting (ADABOOST). The results are compared to other published classification approaches such as evolutionary ensemble learning (EEL) and neural ensemble with arc boosting. Experiments are performed on three well-studied datasets, which are the liver-disorder, breast-cancer and Haberman's survival classification datasets retrieved from UC-Irvine machine learning repository.

Preliminary results have yielded an error rate of 7.57% on the breast cancer dataset, which does not on average improve on state-of-the-art methods (the error rate of EEL is 3.2% and arc boosting is 1.95%), but good performance on some experimental trials indicates the approach is promising.

## **Style and Sentiment Analysis in Movie Reviews**

### Author

**Stephen Jones**  
Computer Science  
Connecticut College  
New London, CT

### Advisor

**Ozgur Izmirli**  
Connecticut College  
New London, CT

Over the past several decades, the studies of linguistics and computer science have combined to quantify and disambiguate natural languages. In recent years, businesses have sought to leverage the massive amounts of text data available on the web using natural language processing to discern key insights. Specifically, the entertainment industry is one where the reviews of the product have a great effect on the revenue it generates upon its release. This project explores the ways in which movie reviews, and the language used therein, is correlated with the movies' commercial success as well as their overall reception. Initially, stylometric was performed to explore the authorship style of reviews. Later, stylometric and sentiment analyses were combined to find correlations between how a film is critiqued and how it is received by the general public. The results of the analyses will be supplemented by expansive data collected from the IMDB movie database to classify groups of movies based on a combination of box office success, critical acclaim, and genre. The poster will report on the use of stylometric features and sentiment analysis individually and then combined with machine learning techniques to show the degree to which these features can reveal any underlying relationships.

## Profiling Synchronization Patterns in Multithreaded Programs

Author

**Kasey Shen**

Computer Science

Wellesley College

Wellesley, MA

Advisor

**Benjamin Wood**

Wellesley College

Wellesley, MA

We built tools that profile the synchronization behavior of multithreaded programs. Error detection tools can use these patterns to make optimization decisions.

Multithreading is a model of concurrent software where multiple threads of execution share the same data. Parallel execution of threads allows for faster processing. However, there is always the possibility of multiple threads simultaneously accessing the same data, causing collisions. Mutual exclusion locks and other synchronization mechanisms can help prevent some of the collisions. Yet, they rely on the programmer knowing where to use them, which is notoriously difficult. Various error detection tools find these problems but sacrifice accuracy or speed. To optimize these tools, we can make informed changes based on profiled program behavior and choose the best tracking approach for each object.

We focused on profiling synchronization patterns when data are first shared. Specifically, we explored two primary questions regarding program lock characteristics: 1) how many locks are held by the thread making that first shared access? 2) is it possible to predict which locks protect which data? To profile the first behavior, we tracked the set of locks held when a thread first shared each object. To measure the predictability of locks protecting certain data, we built a tool that tracks which of the locks in this set continue to protect the object. The tool monitors the initial set size, changes in sizes, and the index of the protecting lock in the set.

Our preliminary analysis on a suite of multithreaded Java programs shows that the protecting lock index is consistent within a program. This result suggests that error detection tools could choose specialized analysis optimizations for each object by exploiting these patterns.

**International Union for Conservation of Nature Ecosystem Visualizer**

Author

**Zhuoming Tan, Son Le**

Computer Science Department

Clark University

Worcester, MA

Advisor

**John Magee, Florencia**

**Sangermano**

Clark University

Worcester, MA

We present an interdisciplinary research project as a visualizer that will help conservation scientists model and simulate ecosystems. It started as a request from the geography department of Clark University and is now being developed with close collaboration between computer science students and advisers in computer science and geography departments. To build a tool for scientists and field surveyers to facilitate the visualizing process of their sketches of ecosystems, we utilize Java framework JGraphX. It is capable of drawing and exporting biological system diagrams, whose specifications come from International Union for Conservation of Nature, to have a standardized way of modeling the system so that the publications and presentations share a common presentation approach. At current stage the software is capable of drawing diagrams via drag and drop from predefined gadgets list. Future planned functions include simple analyzing ability, which qualitatively indicates the effect the change of the factors will have on the biosystem. A prototype version has been shared with IUCN scientists and the first round of feedback is being incorporated into ongoing development.

## Whiteboard Scanning using Super-Resolution

Author

**Wode Ni**

Department of Computer Science  
Dickinson College  
Carlisle, PA

Advisor

**John MacCormick**

Dickinson College  
Carlisle, PA

We present a project on the application of Super-Resolution(SR) in a distance whiteboard scenario. To motivate the project, imagine a person taking a video or an image sequence of a whiteboard from a distance. Due to the limitation of camera capacity and the distance, the words in the whiteboard images are sometimes illegible. Though there exist applications to enhance the image quality, the resolution limit caused by the distance is difficult to overcome. Super-resolution, a class of techniques in the field of Computer Vision, enables us to enhance the quality of the images by utilizing the information of multiple low quality images and fuse them to obtain a higher quality image. In this project, we studied spatial domain SR algorithms described in [1] and [2], experimented on an implementation of it in OpenCV, and tried to apply it on our distance whiteboard data set. Due to the complexity of the OpenCV implementation, we first did black-box analysis on the performance of the algorithm. Quantitative analysis on the quality of the resulting images were then conducted to understand the effect of different parameter values for the algorithm. Finally, we attempted to improve the quality of the output by adjusting the algorithm to fit the whiteboard scenario by experimenting on different types of optical flow algorithms. We will display results and describe details of the experiments in the presentation.

[1] Farsiu, Sina, et al. "Fast and robust multiframe super resolution." Image processing, IEEE Transactions on 13.10 (2004): 1327-1344.

[2] Mitzel, Dennis, et al. "Video super resolution using duality based tv-l 1 optical flow." Pattern Recognition. Springer Berlin Heidelberg, 2009. 432-441.

## **Punctuated Anytime Learning to Evolve Autonomous Wheeled Agent Using Genetic Algorithms**

### Authors

**Moustafa Ndiaye**

Computer Science Department  
Connecticut College  
New London, CT

### Advisors

**Gary Parker**

Computer Science Department  
Connecticut College  
New London, CT

This research project explores Punctuated Anytime Learning (PAL), which is an approach to anytime learning for evolutionary robotics. Most implemented learning systems for evolutionary robotics use simulations for training. This implies physical environments are modeled and passed as parameters in simulations for learning. It is very difficult to have these models represent the environment and its uncertainties accurately. Thus when the learned control program is transferred to the robot, performance is not as effective. Moreover, if the learning takes place on robot, it takes significant time as a result the number of individuals needed to be tested in each population of each generation, which requires expensive time and causes wear to the robot. PAL handles this issue by having the learning take place on a separate off-line learning module, but periodic tests on the robot alter the learning system to improve learning. This requires less tests on the robot, does not require an extremely accurate model, and gives the system the ability to adapt to changes in the robot's capabilities and/or the environment. Prof. Parker has developed PAL and tested it in simulation, but it still needs to be tested using robot. Hence we have built a two-wheeled robot running an Arduino and are learning the control program using genetic algorithms. The robot's task is to perform a perfect square in an 8x8 feet colony space, while a 720p camera above tracks the robot's position and orientation, and reports this information to the learning system. Training takes place on a wireless PC, and communication between the control and learning systems is achieved with Xbee-RF modules.

## Optimizing Technical Indicator Parameters

Author

**Mengezi Ngwenya**  
Computer Science  
Connecticut College  
New London, CT

Advisor

**Ozgur Izmirlı**  
Connecticut College  
New London, CT

We hereby present our investigation of technical analysis parameter optimization, this applies high performance computing to the stock market. Technical analysis is concerned about forecasting financial markets or obtaining trading advice from visual inspection of past prices, without regard to any underlying economic analysis. The goal is to establish technical indicator parameters that maximize profits at specific market conditions. These are important because clearcut success in the market for technical analysts remains elusive. This study observes some of the basic generalizations about the market in the technical analysis calculations that limit the indicators' success. Unlike other similar works, technical indicators here are simulated with varying parameters and stock data and not investigated using an artificial intelligence trader on a single data series. The simulation runs are fine grained and are generally independent of each other. As such, GPU computing using CUDA was used to speed up the simulations. K-means clustering is then used to group the parameter combinations that are close to each other and yield similar returns. A compact cluster at a certain parameter range indicates the reliability of an indicator at producing the respective returns in that parameter range. As such clusters with the most positive returns give us a parameter range for taking bullish positions using that indicator, and the most negative returns provide a parameter range for taking bearish positions. Our results so far challenge the industry standard indicator parameters and suggest that when the market has been experiencing an up trend, it is more profitable to be risk averse. As in other works, the technical indicators are unable to definitively outperform the market. Our work has these indicators performing as well as the Dow Jones Industrial Average.