

CSC 468-02: Final Project

University Chatbot

Nicholas Hines

Jared Miller

Brian Montecinos-Velazquez

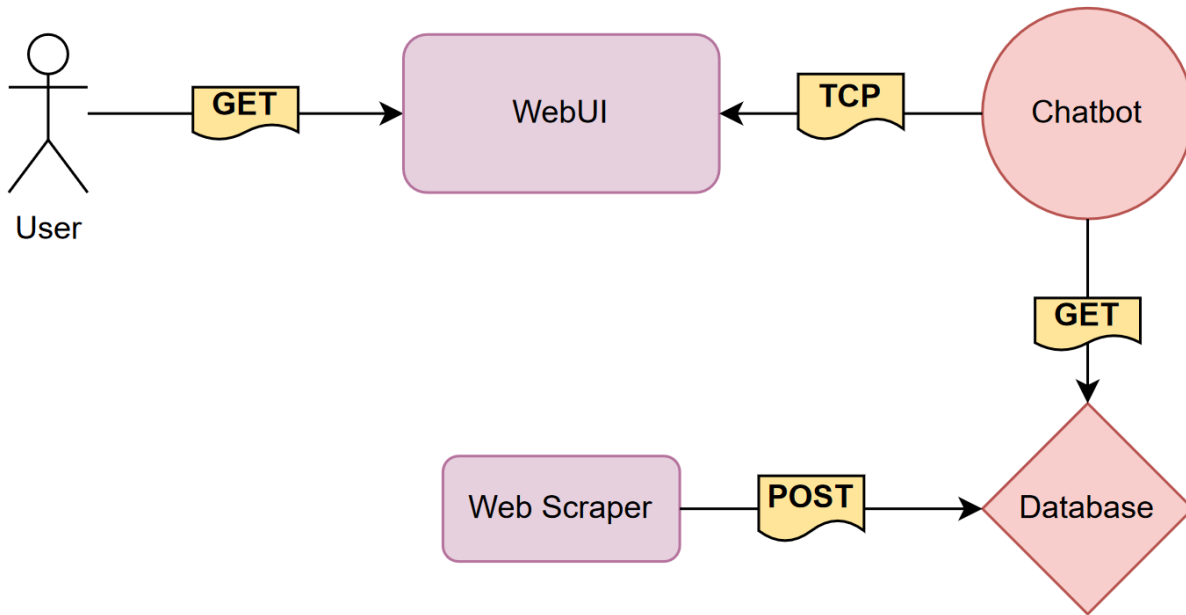
Dominic Pisano

Tyler Profitt

Summary

The University Chatbot is a cloud-based project executed over the CloudLab infrastructure. The project displays the utility of cloud-based development with concepts such as containerization and continuous integration/continuous delivery. The front-end will be a simple chat interface allowing users to ask about campus-related events. The backend will involve a deep-learning chatbot application that can process user input and query a database for information. The database will be populated by web scraping information from wcupa.edu related sites.

Chapter 1: Initial Vision



Chapter 2: Initial Description

The University Chatbot will provide users an interface to make specific queries about campus events through a natural language chat app. We intend to rely on existing open-source frameworks for natural language processing in the chatbot. The core of the project is containerization and Continuous Integration/Continuous Delivery. Development will rely on systems, such as Docker, Docker-Compose, Kubernetes, and Jenkins, to enable automation throughout the development process.

Our project will be composed of four primary components: a Web User Interface, Chatbot, Web Scraper, and Database. As outlined in the diagram in Chapter 1, a typical process flow will involve a user making a query in natural language through the WebUI. This query will be forwarded to the Chatbot for processing. Once the Chatbot has processed the intent of the query, it can form an official query to the Database and get the necessary documents to reply with. The Web Scraper will run separately to populate the Database with documents accessible from the university website. Our plan includes use of the languages, JavaScript, Python, and MySQL. The project will be orchestrated across a Kubernetes cluster and tested with Jenkins. The following is a description of each component.

WebUI

The WebUI will be a streamlined web-based application mimicking standard instant messaging applications available on the market. Through this chat interface, users will be able to make specific, university-related queries and receive accurate responses from the chatbot. Further, the UI will be built in JavaScript, with the framework React.JS and HTML. It will be designed to fit with the standard question and response format of natural language messaging and kept in the theme of West Chester University. The application will be orchestrated as a Kubernetes node where it can be exposed externally to users and communicate with the Chatbot. It will interact with the Chatbot by sending user inputs and displaying responses from the chatbot.

Chatbot

The Chatbot will be a sophisticated, deep learning program that can both read and respond to user text input in natural language. The Chatbot will be trained to understand the intent behind a user's request and formulate an appropriate reply in either natural language, image, or url links. Since we lack much experience with deep learning, we will expedite the development of the architecture needed using accessible natural language bot frameworks, such as Rasa. The Chatbot will be isolated as a Kubernetes pod and interact with the WebUI and Database through TCP requests.

WebScraper

The WebScraper will be a simple script to gather event data from the university website and populate the Database. It will be built with Python making use of dependencies for web scraping, dataframe creation, and mysql interactions: such dependencies include Beautiful Soup, Pandas, and SQLAlchemy. The program would target sites hosting information on campus events, documents, and newsletters, such as ramconnect.wcupa.edu. This data would be converted into a data frame and exported to the Database. The script will be executed on a separate Kubernetes pod linked to the Database.

Database

The Database will store university-related data to be accessed by the Chatbot. This system will be managed through a MySQL server hosted on a Kubernetes pod to ensure availability. It will directly interact with both the Chatbot and the Web Scraper to send and receive documents.

Chapter 3: Intermediate Milestones

Currently, we only have an initial version of the WebScraper which can successfully populate the Database with university events. We have successfully orchestrated this step with both Docker-Compose and Kubernetes. Currently, our project is able to deploy a MySQL server container and a WebScraper container as individual pods with services enabled to allow for TCP communication between them.

Development of the project has been slow due to lack of familiarity with the critical concepts of containerization, cluster networking, and CI/CD. Further, the MySQL server has been especially difficult to set up because of the need to integrate volumes for persistent storage. While the setup through Docker was not difficult, the Kubernetes setup was much more tedious. Most of the development has been accomplished through following tutorials, given the lack of familiarity with commands and setting up yaml files. One misstep was that most tutorials for setting up the server assumed knowledge of or the existence of a default Kubernetes StorageClass on the network.

Despite the slow progress, we have gained enough familiarity with establishing the necessary infrastructure that we are somewhat optimistic that we can complete the project (even if not to an optimal level). A significant hurdle will be the development of the Chatbot itself. The development of a sophisticated deep-learning chatbot may be more difficult than anticipated and we may need to scale it back to a simpler design. While the Rasa framework selected seems to be the most accessible option, we remain worried about our ability to realize our initial vision.

Chapter 4: Final State of Project

Our project has been successfully deployed in Kubernetes with CI/CD pipeline integration through Jenkins. We successfully deployed three out of four components as outlined in our initial vision. At the current state, the WebScraper, Database, and Chatbot are fully functional. The WebScraper is containerized and deployed as a Kubernetes job to populate the Database in a separate pod. The Chatbot is deployed as a multi-container pod that is configured to receive a socket.io connection from the WebUI and can query the Database. For the WebUI, we are able to launch an Nginx container hosting a static html page. This Nginx container has been deployed in a pod accessible to the outside world through a NodePort protocol. However, we lack the integration of a socket.io application within the WebUI to make a connection to the Chatbot. These four components have been separated into distinct branches in our repository with a Jenkinsfile, allowing for automated deployment through Jenkins.

Once Jenkins was installed on our experiment, we had to configure the settings to be able to get the branches to deploy. First, we set up SSH credentials for Cloudlab in Jenkins. We added new credentials with a generated RSA key with one of our Cloudlab usernames. Next, we made the built-in node on Jenkins a single executor and added the deploy label. The last part before adding the branches was configuring a Kubernetes cloud on Jenkins. This involved having to configure our container template to have one of our DockerHub usernames and token set up as well as the pod template. Once Jenkins was configured and after making sure that the Jenkinsfile in each branch was set up to connect to the correct IP and .yaml files, we made a pipeline for each branch in our repository and were able to build them without any issue. With this, our application had complete CI/CD pipeline integration.

The Chatbot is composed of two distinct containers: the rasa and action containers. These containers are launched within the same pod in Kubernetes. The rasa container holds the main NLU processing server for receiving input from the user. The action container holds a server for carrying out custom actions for the chatbot. When the rasa server predicts that a user wants to carry out a certain action, it sends out a request to the action server to carry out the defined action. Currently, the chatbot can only carry out a single action we developed. When it predicts that a user wants a list of events stored in the database, the chatbot is successfully able to query our database and provide the user with a list of event names from the database. The next logical action would be to allow the user to query for more

information on an event in that list. We began the process of creating this action but ran into many hurdles.

For this action to work, the rasa server would need to be trained to recognize event names as entities to be extracted from user input. This process of entity extraction proved to be much more difficult than anticipated. First, we would need to create a lookup table of values generated from the database to train the NLU model on. Ideally, we would dynamically create this lookup table when the application is deployed but for the sake of simplicity we chose to keep the training data constant during development. Despite the lookup table, the entity extraction failed leading us to attempt to reconfigure the NLU pipeline used for training. After many attempts at reconfiguration, we were unable to generate a successful entity extraction pipeline before the project deadline. The lack of experience with machine learning significantly stunted our ability to make the chatbot more sophisticated. Despite this hindrance, we consider the current chatbot deployment to be a complete success as a proof of concept for our initial vision.

In contrast, we regard the WebUI as a failure. While we are able to launch an Nginx container – allowing for external access to our application – we have no sophisticated way of accessing the Chatbot outside of its container. We relied on exec commands to launch a command line script within the container for testing the Chatbot. Our failure in creating a WebUI can largely be attributed to miscommunication. During the early stages of development, before the chatbot was functional, we were unsure of how we would develop an interface. We lost some effort trying to develop a React.js interface that we ended up scrapping because of failure to containerize. Once the chatbot was functional, we discovered that there were several widget applications advertised to be easy to integrate as script in our index.html page. With this discovery, we shifted priority to developing the chatbot over the WebUI. However, we realized (too close to the project deadline) that all the widgets available were deprecated and raised compatibility issues with the Chatbot's socket.io channel. In hindsight, the development of a socket.io application for interfacing with the Chatbot should have been prioritized sooner in the project timeline.

Reference

GitHub Repository: <https://github.com/bmv0161/csc468-final>

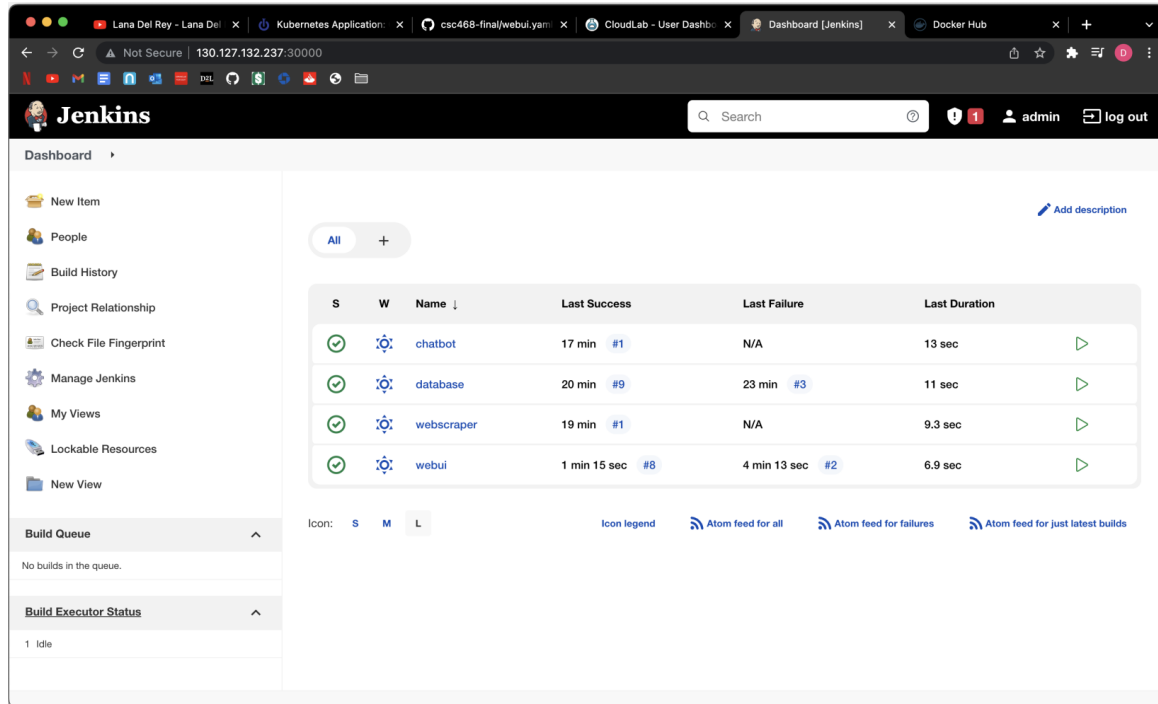


Figure: Jenkins Deployment

NICHOLAS HINES

Projects

Designed Website

- Used HTML for structure, CSS for design features
- Created website similar to the official Philadelphia Phillies website with multiple tabs and multiple pages
- Included seamless navigation between all pages
- JavaScript features including input validation, if statements, and a real world clock

Converted Code

- Converted code of a lexical analyzer from C to Java
- Involved testing of a variety of input files

Created Unique Software Development Process

- Created Fake company and outlined its entire software development process including development, release planning, iteration planning, testing, coding standards, and release
- Incorporated the Capability Maturity Model

Relevant Coursework

- Java Programming
- Data Structures and Algorithms
- Database Management Systems
- Programming Language Concepts/Paradigms
- Software Engineering
- Web Design
- Computer Security/Ethics
- Public Speaking

EXPERIENCE

LIFEGUARD

Hershey Entertainment and Resorts
Hershey, Pennsylvania

May 2017 - August 2020

- Learned and maintained proficiency in first responder skills such as First Aid and CPR to offer individuals in distress optimal support
- Demonstrated leadership skills as the most experienced lifeguard and instructed other co-workers when a manager wasn't present

✉ nhines030@gmail.com
☎ (717) 303-7683
📍 918 Roush Road
Hummelstown, PA 17036

EDUCATION

Bachelor of Science (B.S.)
Computer Science Candidate

WEST CHESTER UNIVERSITY
West Chester, PA
GPA: 3.72
Anticipated Graduation: May 2022

Awards & Honors

- Dean's List
Fall 2018, 2020 / Spring 2020, 2021

SKILLS

Languages (Proficient)

- Java
- SQL
- HTML

Languages (Familiar)

- JavaScript
- Python

Relevant Technologies

- Microsoft Visual Studio
- Git / GitHub
- Linux
- Microsoft Access
- NetBeans

Jared Miller



jdm750@yahoo.com



(215) 872-9731



Philadelphia, PA, 19154

LinkedIn:

www.linkedin.com/in/jmiller75

GitHub:

<https://github.com/jaredm750>

EDUCATION

WEST CHESTER UNIVERSITY OF PENNSYLVANIA

GPA: 3.608 / 4.0

West Chester, PA

*Bachelor of Science (B.S.) Computer Science
(Expected graduation May 2022)*

Relevant Coursework

- Java I, II, III
- Computer Security I, II
- Modern Malware Analysis
- Software Security
- Data Communications and Networking
- Software Engineering
- Data Structures and Algorithms
- Cloud Computing

Awards & Honors

- Dean's List 3x

Extracurricular Activities

- Computer Science Club
- Cyber Security Cub
- Competitive Programming Club
- Sigma Phi Epsilon Fraternity
- American Red Cross Club (ARCC)

ADDITIONAL SKILLS

Programming Languages / Software

- Java, Python, JavaScript, CSS/HTML
- Microsoft Office, VMware, Eclipse, Wireshark

CERTIFICATIONS

NSA/ABET Computer Security Certificate from West Chester University (Expected May 2022)

EXPERIENCE

SOFTWARE ENGINEER INTERN

HCL Technologies / January 2022 - Present

- Work in a team-oriented environment for all aspects of product development.
- Work closely with mentors to review problem report solution and implement the solution.
- Participate in software development using Agile/Scrum.
- Utilize frameworks such as Spring Boot, Angular, Node.js.
- Proactively identify and communicate risks and issues to mentors and create plans to mitigate risks.

IT TECHNICIAN INTERN

County of Chester, West Chester, PA / May 2021 - Present

- Devise solutions to issues related to Microsoft Office before deployment to hundreds of government employees.
- Patch software and install new versions to eliminate security problems and protect data.
- Diagnose, troubleshoot, and resolve hardware, software, or other network and system problems, and replace defective components when necessary.
- Follow up with clients to verify optimal customer satisfaction following support engagement and problem resolution.

INTRAMURAL SUPERVISOR

Student Recreation Center, West Chester, PA / May 2019 - May 2021

- Explained and taught regulations and rules to players.
- Worked as a team to ensure a safe environment for players.
- Trained new referees on safety regulations and game rules.
- Collected daily reports for improved efficiency during games.

PROJECTS & NOTABLE COURSEWORK

MODERN MALWARE ANALYSIS

MALWARE ANALYSIS TOOL

- Built a dynamic heuristic analysis tool for the detection of unknown malware.
- Downloaded a ransomware sample in a Windows XP environment and a monitor program in VirtualBox.
- Constructed a runnable program using Python to read through the file and determine if the activities are malicious.

SOFTWARE SECURITY

- Participated in 6 labs throughout the semester using Unix/Linux, including the command-line shell, and writing Python code to exploit certain vulnerabilities.
- Exploited attacks such as buffer overflows, SQL injection, and session hijacking by using advanced testing and program analysis techniques.

BRIAN MONTECINOS-VELAZQUEZ

Norristown, PA, 19401
(484) 674-5092
bmv0161@gmail.com
github.com/bmv0161



EDUCATION

Bachelor of Science in Computer Science

West Chester University of Pennsylvania

AUGUST 2019 – MAY 2022 (EXPECTED)

NSA/ABET Computer Security Certificate • Minor in Applied Statistics

GPA: 3.91/4.0 • **Dean's List:** from Spring 2020 to Fall 2021

Outstanding Senior Award 2021 - 2022

Relevant Coursework:

Computer Science I – III • Software Engineering • Computer Systems
Malware Analysis • Data Structures and Algorithms • Cloud Computing

Master of Science in Computer Science

West Chester University of Pennsylvania

AUGUST 2021 – MAY 2023 (EXPECTED)

GPA: 4.0/4.0

Relevant Coursework:

Software Security • Networking • Database Systems



EXPERIENCE/ACTIVITIES

General Merchandise Expert | Target Corporation

JUNE 2019 – CURRENT

- Required efficient time management to complete duties
- Developed strong communication skills for customer interaction
- Discovered significance of creating collaborative environment

Cyber Security Club | West Chester University

SEPTEMBER 2021 – CURRENT

- Participated in NCAE Cyber Games 2022:
 - Linux System Administration
 - Virtual Machine configuration
 - Capture the Flag challenges
- Participated in NSA Codebreaker Challenge 2021
 - Wireshark packet analysis and program debugging

Computer Science Club | West Chester University

FALL 2021

- Worked on game development with a group of peers
- Worked with Unity builds and GitHub



PROJECTS

University Chatbot (Spring 2022): led a small team to develop a cloud-based chatbot application

- Learned about cloud-based development: containerization, orchestration, CI/CD pipeline
- Development through GitHub, Docker, Kubernetes, Jenkins
- Worked with Python, MySQL, Rasa, JavaScript

Graduation Planner (Fall 2021): Led a small team to develop a sorting program to find graduation plan from map of course catalog

- Developed a graph data structure and applied topology sort
- Developed webscraping program to populate graph with data
- Worked with Java, Python, and organized a GitHub Project



SKILLS

Programming Languages:

JAVA • PYTHON • C

Technical Knowledge:

Object Oriented Programming • UML
Functional Programming • Assembly
Software Security • GNU Debugger
Windows Terminal • Linux CLI • Vim
Wireshark • IntelliJ • Eclipse
GitHub • DockerHub • Kubernetes

Statistical Analysis:

MySQL • R • SAS • SPSS • Excel
Model Building • Data Frame Handling
Database Management

Digital Communication:

Office Suite • Adobe Creative Suite

Spoken Languages:

English (Fluent) • Spanish (Fluent)

Dominic Pisano

(610) 719 7331 - Thorndale, PA - pisano.dominic@gmail.com

Experience

Wegmans

June 2021 – Present

Instacart Shopper

Downingtown, PA

- Utilize Instacart app to communicate with and fulfill customer orders.
- Properly handle product, use necessary equipment, and bag items to ensure they reach their destination safely.
- Accurately and efficiently scan items, make appropriate substitutions as needed, and process payments.
- Accurately and efficiently fulfill customer orders and stage them in fulfillment area.

CVS Pharmacy

June 2019 – July 2021

Shift Supervisor

Thorndale, PA

- Serve as manager and perform managerial duties in the manager's absence.
- Addresses all customer complaints in a professional and timely manner.
- Operate cash registers, stock shelves, and keep the store clean and organized.
- Supervise associates on shift and ensure the daily task list is completed.
- Reconcile cash register at the end of the shift and lock the store.

Rite Aid

May 2017 – June 2019

Shift Supervisor

West Chester, PA

- Removing outdated merchandise products from the shelves for damages/outdated.
- Set planograms for aisles, endcaps, and side wings according to the set layout.
- Pull merchandise from stock room to sales floor.
- Assist customers with the lottery machine, printing photos, and taking passport photos.

Cracker Barrel

June 2016 – May 2017

Server

Downingtown, PA

- Write guests' food orders on order slips and enter orders into computers for kitchen staff.
- Serve food or beverages to patrons.
- Perform food preparation duties such as preparing salads, appetizers, and cold dishes, portioning desserts, and brewing coffee.
- Clean tables or counters after patrons have finished dining. Stock service areas with supplies such as coffee, food, tableware, and linens.

Education

West Chester University | Bachelor of Science in Computer Science

Expected December 2022

Relevant Coursework: Calculus II, Discrete Mathematics, Data Base Management Systems, Data Communications and Networking, Data Structures and Algorithms, Introduction to Cloud Computing, Software Engineering, Software Testing

Delaware County Community College | Associate of Science in Computer Science

2021

Skills

- Proficient in Java, HTML, CSS, JavaScript
- Customer service skills
- Familiar with C, C++, Python
- Leadership skills

TYLER PROFITT

4015 Bethel Rd. Upper Chichester, PA, 19061 · 610-739-9951

Profitized5@gmail.com · <https://www.linkedin.com/in/tyler-profitt-361818223>

Dedicated student leveraging studies in Computer Science to seek real-world experience. Strongly determined to learn and understand more throughout the opportunities that I may be given.

EXPERIENCE

JANUARY 2022 – PRESENT

NETWORK SUPPORT TECHNICIAN, CHICHESTER SCHOOL DISTRICT

- Maintain network connectivity of workstations, servers, and network devices.
- Provide faculty and staff with security software and network configuration support.
- Diagnose and execute resolution for network and server issues.
- Perform laptop diagnostics and physical disassembly while providing replacements of internal components of a laptop.

EDUCATION

AUGUST 2020 – FALL 2022 (EXPECTED)

BACHAELOR'S: COMPUTER SCIENCE, WEST CHESTER UNIVERSITY

Dean's List Honoree:

- Fall Semester 2020
- Spring Semester 2020
- Cumulative GPA: 3.736

JANUARY 2018 – MAY 2020

ASSOCIATE'S: LIBERAL ARTS, DELAWARE COUNTY COMMUNITY COLLEGE

Dean's List Honoree:

- Fall Semester 2018
- Spring Semester 2019
- Cumulative GPA: 3.11

AUGUST 2013 – JUNE 2017

HIGH SCHOOL DIPLOMA, CHICHESTER HIGH SCHOOL

SKILLS

Languages (Familiar)

- Java
- Python
- HTML and CSS
- SQL

Programming Software

- Eclipse
- Jupyter Notebook
- Spyder
- Pycharm