

CS CAPSTONE PROGRESS REPORT

DECEMBER 4, 2017

CDK DATA STREAM AI

PREPARED FOR

CDK GLOBAL

CHRIS SMITH

PREPARED BY

GROUP 65

SIGFIND

INHYUK LEE

JACOB GEDDINGS

JUAN MUGICA

Abstract

This report is the cumulative summary of our groups progress throughout the fall term. We will be discussing the various assignments of the term and how we handled them as well as touching upon client interaction and group decisions. This report will also briefly talk about the upcoming winter break and what our plans are during the downtime between terms. At the end there will also be a week-by-week breakdown of events and a table indicating things that went well, poorly, and how to improve quality of life moving forward. Moving forward our focus will be shifting heavily towards coding accomplishments while this past term and report are favoring the documentation aspect.

CONTENTS

1	Project Purposes and Goals	3
2	Progress of this term	3
3	Problem and solution	3
4	Week activities summary	4
4.1	Week 1	4
4.1.1	Activities	4
4.2	Week 2	4
4.2.1	Activities	4
4.2.2	Problems	5
4.2.3	Solutions	5
4.3	Week 3	5
4.3.1	Activities	5
4.3.2	Problems	5
4.3.3	Solutions	5
4.4	Week 4	5
4.4.1	Activities	5
4.4.2	Problems	5
4.4.3	Solutions	6
4.5	Week 5	6
4.5.1	Activities	6
4.5.2	Problems	6
4.5.3	Solutions	6
4.6	Week 6	6
4.6.1	Activities	6
4.6.2	Problems	6
4.6.3	Solutions	7
4.7	Week 7	7
4.7.1	Activities	7
4.7.2	Problems	7
4.7.3	Solutions	7
4.8	Week 8	7
4.8.1	Activities	7
4.8.2	Problems	7
4.8.3	Solutions	7
4.9	Week 9	8
4.9.1	Activities	8

		2
	4.9.2 Problems	8
	4.9.3 Solutions	8
4.10	Week 10	8
	4.10.1 Activities	8
	4.10.2 Problems	8
	4.10.3 Solutions	8
5	Retrospective table	8

1 PROJECT PURPOSES AND GOALS

Our project, CDK Data Stream AI, is concerned with the construction of a program that can arbitrarily detect if a document has a signature line and if it has been signed. To accomplish this primary task the program must make use of an open source AI platform, function as a black box, and be modular enough that parts can be added or removed from the program. Accompanying this primary goal is a series of stretch goals that, should our primary concern be easy to implement, become our focus for the remainder of the coding section. These stretch goals include license validation, data security, image categorizing, and vehicle image processing. Should these be accomplished, additional interests have been stated in regards to automatic construction of consumer portfolios based on submitted forms.

A major concern for our client is that we do not get caught up on peripheral issues or stretch goals before the core task is completed. Weve been explicitly requested not to worry about container software, cloud operation, or user interface in particular. Another issue of which we must remain mindful is data sensitivity; any documents or forms were permitted to use for testing purposes cannot be made public. They are sanitized for us to legally view, but the forms are not open to the public. Lastly, CDK Global is aware that this project may not be feasible and in lieu of a completed product, they will accept a report indicating why we were unable to complete the project.

2 PROGRESS OF THIS TERM

All required documentation for fall term has been completed and submitted. All documents requiring client verification have been cleared except for the design document which was pushed into winter term. Regular meetings have been scheduled with our TA every week for 20 minutes as well as 30 minute meetings with our client. Communication has been setup for intergroup contact through the use of group text/chat services, and contact with our client has been established through the university-provided emails. Plans have been outlined for the duration of winter break with a strong focus on code writing.

The program itself has also made some progress throughout the term. We have selected OpenCV as our launching platform for the project and will be using its C++ variant as our leading choice. Our group has successfully completed setup of OpenCV on our personal machines and student FLIP server. This includes adding the ability to view images from PUTTY. Lastly, our most recent development for OpenCV is constructing an image window for any submitted image. Plans have been made to ensure progress continues throughout winter break with the intent to have PDF to JPEG conversion included into our program as well as the ability to construct reduced windows in OpenCV to pinpoint where a signature line is located.

3 PROBLEM AND SOLUTION

We encountered numerous problems throughout the term from document formatting to program initialization. Our first document, the problem statement, was originally constructed prior to meeting our client, and we had to make several rather generous assumptions regarding what exactly this project was about. Fortunately, we were able to meet with our client prior to the final submission where we learned that our previous understanding of the project was entirely

incorrect. Prior to this knowledge, we had assumed the project was about the construction of an AI to monitor potentially malicious code being sent throughout their network. The solution to this misunderstanding was in the meeting itself; we learned that this is a project for detecting signatures, and fortunately there was sufficient time before the final draft to rewrite it to conform to what our client wanted.

Another document error surfaced with our technology review. There was a miscommunication which resulted in each person writing about a single piece of the project and three subsequent technologies with each tech being 500 words with a total of 1500 per piece. After doing peer review and contacting our TA, we learned that it was three pieces at 500 words per piece. Corrections were made by significantly gutting our individual pieces and then constructing more components to the project.

As for our programming portion, there was one instance of misunderstanding with the client and one issue with setting up our respective OpenCVs. Initially, there was an assumption regarding the construction of a user interface and that the program would be communicating directly with the user. This is not the case. Our client corrected us indicating that there should be no UI beyond what is required for maintenance. To correct this error, our requirements document had a section covering our plans for UI layout removed prior to final submission. Lastly, a group member had difficulty with OpenCV setup which was the result of incorrectly calling the program on the wrong folder level, resulting in a botched installation. This was eventually remedied via a how-to guide written by the other members of the team.

4 WEEK ACTIVITIES SUMMARY

4.1 Week 1

4.1.1 Activities

An introduction to the course occurred throughout week one. During this time we were asked to go through each available project and vote on five that we liked, two we didn't like, and two students we did not want to work with. At this time our group had not been formed but all three members had an interest in artificial intelligence and as such voted on the CDK Data Stream project. We were also informed that we'd need to write weekly blogs on OneNote and be using Latex as our go-to for document writing.

4.2 Week 2

4.2.1 Activities

Teams were assigned and our we became Group 65 and proceeded to meet each other and make contact with our client. During this week, we planned a physical meeting for an hour in which we, as a group, discussed our interpretations of what the project is, what research we should be doing, and how best to contact our client.

4.2.2 *Problems*

Establishing a reliable method of communication between group members was difficult at the start, having never met before it was difficult to find each other in the classroom and reliably stay in touch through emails.

4.2.3 *Solutions*

We concluded that group texting was the most reliable method of reaching everyone and when contacting our client, we'd assign a single person to email him.

4.3 **Week 3**

4.3.1 *Activities*

Our first major assignment started this week, the problem statement, which would require us to write about what the project entails and how we intended to meet that goal. During this time, we also made contact with our client, Chris Smith, and scheduled a physical meeting at the end of the week. Clerical work was also requested of us, our TA was assigned to us and requested that our OneNote and GitHub be shared across instructors.

4.3.2 *Problems*

Our initial attempt at the problem statement was met with several issues, namely the lack of a strong idea as to what the project was, as well as an incorrect focus on format over content.

4.3.3 *Solutions*

Prior to final submission of the problem statement we had the chance to meet with our client and get clarity on the project. In addition, we had a peer review session in which we learned a great deal about the importance of the content of the document despite the lack of information to go on. Greater emphasis was made towards doing pre-emptive research on image detection to help in completing the paper.

4.4 **Week 4**

4.4.1 *Activities*

Final draft of problem statement was due this week. We also had group member, Mugica, meet with our client to receive a USB drive containing example documents that our project will work on. Research began in earnest for a suitable AI to carry out our image recognition needs.

4.4.2 *Problems*

Latex implementation proved difficult with our final pass on the problem statement. Issues with compiling with missing packages, formatting inconsistencies, and poor section construction resulted in a less than ideal paper.

4.4.3 Solutions

Format was simplified for the sake of completing the document, we also managed to determine the issue with user packages causing compiler errors. The lead issue for compiling was that third-party latex viewers would auto install missing packages whereas the school server would not. Once we isolated what was needed we manually included them and it functioned normally.

4.5 Week 5

4.5.1 Activities

Requirements document was now underway, at this time we focused heavily on discerning what was the primary goal of the project and what was the stretch goals. At this time we concluded OpenCV was likely our best bet for project completion.

4.5.2 Problems

Many goals set during our first meeting where too ambitious to be completable within our timeframe.

4.5.3 Solutions

Continued meetings with Chris Smith allowed us to narrow down our project goals by isolating precisely what they wanted, what was an added bonus, and what was of little concern for us.

4.6 Week 6

4.6.1 Activities

Researched OpenCV languages, implementations and began to look at the libraries offered by the framework. Finalized requirements document, and LaTeX implementation.

4.6.2 Problems

OpenCV can be implemented in a wide range of operating systems and languages. The task was to figure out a suitable OS as well as a language that we were all comfortable in understanding so as to ease the overall coding process. When attempting to implement OpenCV ran into various dependency issues especially with the `-sudo` command in Linux, since this command requires special permissions.

4.6.3 Solutions

The Linux implementation of OpenCV was chosen since it allowed us to run our program on OSUs flip server. C++ was chosen as a primary, with Python as a backup as per overall group member agreement. OpenCV utilizes the -sudo command to set up a visual framework for displaying images. Instead of this, we utilized X11, in particular Xming, and set up a putty instance that could this visual server to allow OpenCV to display images.

4.7 Week 7

4.7.1 Activities

Created technology review rough draft. Researched different types of neural networks, especially those suited for image processing. Weighed our options as to runtime environments as well as containerizing software. Implemented OpenCV in Linux as well as created sample program. Created documentation for my group to follow when implementing OpenCV.

4.7.2 Problems

The original layout of the technology review was rather unclear. Our whole group found out that we had done it incorrectly upon meeting with our TA. We had dedicated 1500 words to three technologies, when in reality there should have been 9 technologies, 500 words for every three.

4.7.3 Solutions

Took the useful information from our rough drafts, and with a clear idea of what the technology review was supposed to look like began finalizing the document.

4.8 Week 8

4.8.1 Activities

Technology review rough draft was completed, corrections where underway to meet the correct format.

4.8.2 Problems

Our client had to travel and we were having difficulty maintaining contact with him, informed our instructors of the concern as well as our TA.

4.8.3 Solutions

After attempting to call our client he was able to respond and sent the necessary verification emails to our instructors.

4.9 Week 9

4.9.1 Activities

Contact with client resumed as normal, corrections were completed on technology review and plans were made for what we'd be working on for the holiday break.

4.9.2 Problems

Client expressed concerns over our choice to use C++ as the language for the project, voicing more interest towards Java support.

4.9.3 Solutions

Explained our reasoning to which our client agreed with, concession was made to plan for a backup language should C++ become too unwieldy. Current plan is to move towards Python in that event.

4.10 Week 10

4.10.1 Activities

All focus has been put towards completion of the design document and progress report. Winter break was mapped out for individual work as well as group work. Communication will continue with client at a reduced rate throughout the break.

4.10.2 Problems

Design document formatting was very awkward for us, despite looking at an example document we still did not know what would work best for our situation.

4.10.3 Solutions

We merged what resources we found for the document and proceeded with that, the final design does carry a logical structure that followed the IEEE document provided.

5 RETROSPECTIVE TABLE

positives column	deltas column
Our group submitted all group assignments within time	Active communication between group members and client is needed
Project is an interest to all party members	Progress on work needs to be more evenly distributed over a week
Client is approachable and is excited about AI as well	More preparation for client absence will be needed
Progress on coding is already underway	More active use of OneNote and GitHub to share individual progress

TABLE 1
Retrospective table