2023 - 2024

COMP S450 Applied Computing Project

DREAMS: A ONE-STOP DOCUMENT RECOGNITION, EXTRACTION AND MANIPULATION SOLUTION FOR LOGISTICS INDUSTRY

Interim Report

Group members:

1155180	Eraka Yip	(BCOMPHIT1)
1265005	Timothy Chan	(BSCHCN)
1284810	Andy Lau	(BCOMPHIT1)

Project Supervisor

Dr. Voyce Li

Course Coordinator

Dr. Jeff Au Yeung

Abstract

The logistics industry is an important sector in Hong Kong's economy. To facilitate sustainable growth and maintain competitiveness, embracing new innovations is necessary.

As a logistics solution provider, we consistently support the growth of our port community. Through our efforts, we have successfully implemented an export and import platform that can be utilized by various stakeholders in the industry. However, due to a lack of correlation between booking data and shipper information, the workflow cannot be fully digitized and still requires manual input from the shipper.

In this project, we are developing a document processing system that allows end users to extract booking document information in a digitalized and automated manner. This system is supported by state-of-the-art, fine-tuned AI custom models for object detection (YOLOv8), document classification, and token classification (LayoutLMv3). The results can seamlessly integrate into the existing logistics platform for terminal operations, enhancing data security and reducing manual input for end users.

Table of Content

1. Introduction

1.1. Overview

In Hong Kong, maritime and port industry has taken the key role in supporting the tiny city as one of the most important trading hubs in the world. Based on statistics in 2022, the whole industry contributed 4.1% (equivalent of HKD 111.8 billion) of the GDP and 2.1% (about 78,400 jobs) of the total employment [1].

Within the industry, export and import services are the most important sector. Thus, the government has been committed to promoting smart ports to optimize service capacity and efficiency. A further approach is to establish an automated container terminal, where terminal operations are fully digitalized and automated, i.e. without human intervention [2].

As a logistics solution provider, we always support the growth of our port community. With our effort, we have successfully implemented an export and import platform that can be utilized by different stakeholders of the industry. Using the platform, most of the terminal operations are now digitalized and paperless. For example, in the past, manual inspection of documents was required for entry and exit at the terminal gate. This is no longer required. Instead, the platform consolidates necessary data to the terminal for review. Now, end users can simply use the platform's mobile application for a variety of terminal operations, saving a lot of time and reducing human error.

However, as the platform runs, there are still some unsolved problems that hinder the smooth run of the platform. To understand the problems clearly, we first introduce the stakeholders of the platform.

Stakeholders	Roles & Responsibilities
Terminals 	**Logistics:** – Control gate in and gate out of a port. – Provide container storage for pickup and return. – Vessel arrival / departure. **Platform:** – Provide data for the platform that shows information to end users (Shippers and trucking company).
Shipping Lines 	**Logistics:** – Provide vessels for ocean freight shipping. – Work with terminals on available timeslots for vessel arrival / departure and containers loading / discharge. **Platform:** – Provides data for terminal operation.
Shippers 	**Logistics:** – Request bookings from shipping lines for ocean freight shipping of their goods. **Platform:** – End users of the platform. – Since booking data in the platform does not include shipper information, shippers have to input booking number in the booking document manually to grant permission to retrieve it for further logistics process (forward to trucking company).

Trucking companies 	**Logistics:** – Pick up or return containers. **Platform:** – End users of the platform (Shipper may delegate the role) – Drivers can use a mobile app provided by the platform for gate-in and gate-out terminals. – Receive booking from shipper via the platform and assign drivers to pick up or return containers.
Platform provider 	**Platform:** – Establish an online export and import platform for end users (i.e. shippers and trucking companies) to complete export and import operations in a digitalized and paperless manner. – Continuous system integration and maintenance tasks of the existing platform.

Table 1: Roles and Responsibilities of stakeholders

The following diagram shows the necessary operations among stakeholders in the booking retrieval procedure. The green part is the problematic area to be discussed.

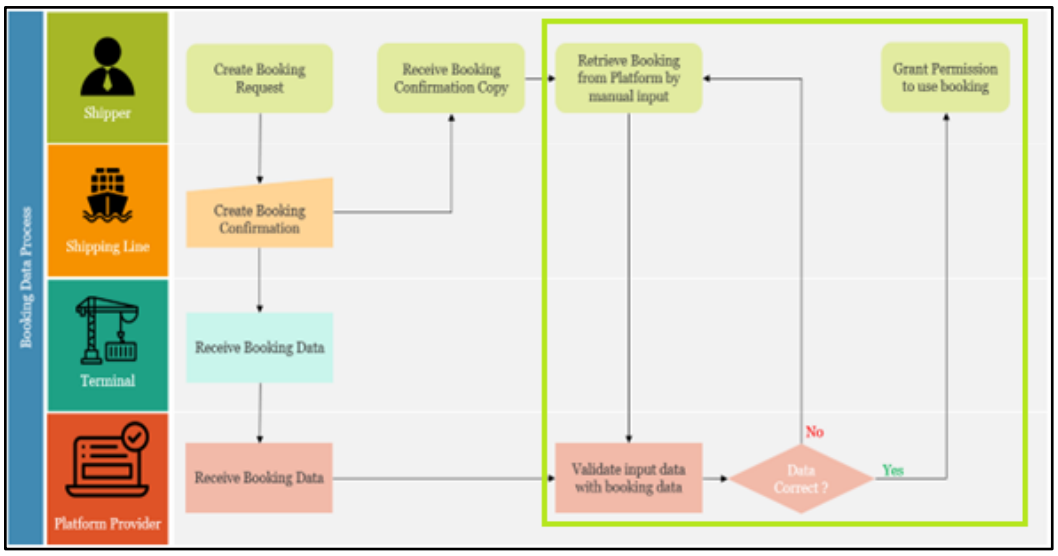


Figure 1: Flowchart of the Booking retrieval procedure

In general, end users (i.e. shippers) carry out the following steps for terminal operations using the export and import platform. The first step is the most important step because any mistake here will lead to failure of the shipment operation.

1. The shipper retrieves the booking on the platform.
2. If the retrieval is successful, the shipper forwards the booking to the trucking company in the platform.
3. The trucking company assigns a truck tractor and driver to pick up the container from the terminal.
4. After the container has been picked up and the cargo has been packed into the container, either the shipper or trucking company submits the declaration for the return of the container to the terminal.
5. After the container is returned to the terminal, the shipper will get the trans-shipment receipt and then wait for the container to load on the vessel.

So, what are the problems for the current booking retrieval procedure?

Under the business logic, when a booking has been successfully retrieved by a user, other users cannot retrieve the same booking data. This is to prevent multiple retrievals of the same booking for terminal operations.

However, the data relationship is not transitive. Shipper requests booking from shipping lines and shipping lines cooperate with the terminal on the booking do not imply that the platform can own the shipper data for each booking. As shipping lines do not provide shipper information to the platform, there is a loss of relation between the booking information and the shipper who owns the booking in the platform.

Currently, the platform assumes that only the booking owner can get the booking document and the booking number. The shipper is required to manually input the booking number in the booking document to establish the relation, linking as the owner of the booking retrieval and granting for remaining logistics processes.

There are two problems for the current practice.

1. **Low data security**
   Booking number is sequential for each shipping line. Therefore, unauthorized users can get access to other booking information by randomly guessing the valid booking numbers, at least in theory. As booking information contains sensitive information of the shipper, it reduces data security of the whole platform. As a responsible company we should always protect the data integrity of our users.

2. **A non-digital procedure**
   Unlike other steps of terminal operations which have been digitalized, the booking retrieval step is manual. It requires manual input of the end user, reducing efficiency and user experience. Also, it increases the risk of manual input errors, which may retrieve booking from other shippers without intention.

Therefore, it is necessary to design a new solution to enhance data security and reduce manual, repetitive input tasks of the end users.

We aim to solve the problems using state-of-the-art AI models for classification and extraction of booking documents under the new stage of AI wave.

1.2. Project Aim

The aim of this project is to develop an AI-based document processing system that allows shippers to use a web application to retrieve booking documents in an automated and digitalized manner for terminal operations, reducing manual input and enhancing data security.

Feb 05, 2024: [The aim of this project has remained unchanged. Rephrase only.]

1.3. Project Objectives

1. **Collect data and build dataset**
   Four document formats are considered in our projects. We prepare 10 booking documents per shipping line, convert to JPEG format and propagate using image augmentation to mimic the variety and quality of actual booking documents.

2. **Setup AI training environment**
   Prepare a basic machine learning environment, additional pre-trained models, and the libraries that the models required.

3. **Train, test and evaluate models**
   There are two types of models in our system for document classification and data extraction. We train with different sizes of dataset and other hyperparameters.

4. **Select the best models**
   Models with the optimal performance are selected for system implementation.

5. **Package and evaluate the solution**
   We have to package the two AI models and run the models on a server. Also, it is necessary to create a web interface for shippers to submit documents and a database for persistent storage.

   After the prototype is created, we test and evaluate the solution based on functional requirements and non-functional requirements of the system.

Feb 05, 2024: [The core project objectives have remained unchanged, except for the scope of datasets and models focused on.]

1.4. Value Propositions

The success of this project can bring us a platform with enhanced data security, protecting user data from leakage. In addition, it reduces manual input and the associated human error, resulting in an increase in platform efficiency.

The project outcome may contribute to the smart ports in Hong Kong.

2. Background or Literature Review

2.1. Review of Existing or Related Solutions for the Problem

As the platform provider, we proposed three solutions to tackle the problems using system approach and data approach.

Since the root cause is a loss of relation between the booking and the corresponding booking owner, we aim to reconstruct the relation with the support from the shipping lines. Another approach is to reconstruct the relation through provision of more information by the end users.

1. **Unique ID as the booking key**
   Shipping line can send a unique ID to both shipper and terminal as the booking key. The unique ID will be stored in our platform database. This establishes a relation between the shipper and the booking which can be used to retrieve booking in the platform using the unique ID.

   This solution ensures data security and keeps the workflow simple. However, it requires each shipping line to change their own system. Since shipping lines are generally global companies, it is very difficult for them to change their system for just a use case in a local company. Also, there are so many shipping lines in Hong Kong, making this solution unrealistic in terms of system implementation.

2. **An extra field for company ID in the booking request**
   This is a simplified version of solution 1. Instead of system change, the shipping lines provide a field for shippers to input their company ID in the booking request. The data is then sent to the terminal to be used for validation.

   This solution requires the shipping line to add an input field, which is less development effort than the previous solution. However, it is difficult to control if the shipper inputs the company ID incorrectly.

3. **Shipper input more information**
   Since it is difficult to change the system of shipping line, another approach is to request shippers to input more information to retrieve the booking.

   This method does not require any enhancement of the system of shipping lines and keeps data security since booking request with multiple information required is very difficult to guess out correctly. However, it makes the workflow complicated, which is more difficult to retrieve a booking and against our digitalization and automation direction.

Given the existing limitations, we looked for technologies that can automatically analyze the booking document submitted by the shipper and extract the booking data that can be incorporated in our current digitalized operations. These technologies can simplify the workload of end users, enhance data security and maintain our digitalization policy workable.

We searched for two supporting technologies available on the market and they gave us some insights on the project direction.

1. **SaaS (Software-as-a-Service)**
   Nowadays, many companies have migrated their on-premises data centers to cloud to enjoy the benefits of cloud services such as high flexibility and scalability [3]. These can help companies reduce their maintenance cost and can respond to sudden demand change quickly.

   Similarly, the big-three cloud providers – AWS, Microsoft Azure and Google Cloud Platform are also offering their intelligent document processing (IDP) solution [4]. When end users upload documents to the cloud, the solution will analyze the document layout and extract the key-value pairs based on pre-built AI models. When necessary, custom model can be trained to achieve better performance and fit the specified use case [5].

2. **AI-powered document processing software**
   There are several AI-powered document processing software available in the market. This kind of software makes use of their custom AI models and optical character recognition (OCR) to analyze the document and extract the key-value pairs. Studies showed some outperform current available AI models [6].

There are several considerations when using these technologies as an enterprise solution.

a) **Data governance policy**
   We have to update our client's sensitive data to the public cloud for document processing, which are not allowed in the IT policy of terminals.

b) **Black-box solution**
   We have no idea of the details of the AI models, including the algorithm and model architectures. We also lose an opportunity to learn new technology.

c) **Lack of control**
   As a platform provider, we should have a better control on the models used. However, in a SaaS solution, the cloud provider is responsible for providing the updates and patches.

From the review of existing solutions and related technologies, we have the following conclusions.

1. System change of shipping lines and voluntary submission of shipping information by the shippers are not feasible because of the global scale of shipping lines and increasing workload to end users.

2. AI is the key to the success of the project. The solution can reduce the workload of the end users, validate the booking document while extracting key data that can be incorporated into the current platform.

3. As an enterprise solution, custom AI models should be developed and deployed to avoid the problems arising from using public SaaS solution / software shown above.

Which kind of AI models should we use in the project? To answer this question, we should introduce basic terminology of AI and its recent development.

Artificial Intelligence (AI) refers to the simulation of human intelligence processes by machines. Within AI, machine learning is the mainstream. Computers learn from data to perform tasks without explicit programming. Traditional machine learning algorithms include decision trees, support vector machines, and linear regression.

In recent years, deep learning, which is a subset of machine learning, has gained dominant attention. It employs artificial neural networks such as convolutional neural networks (CNNs) for image detection and recurrent neural networks (RNNs) for text translation. These models achieve remarkable results but require large, labeled datasets, powerful GPUs, and lengthy training times.

In 2017, Google introduced the transformer architecture, revolutionizing natural language processing (NLP). By using multi-head attention mechanisms, transformers address issues like memory loss and lengthy RNN training [7].

It opened a new chapter in deep learning. Since then, many big techs have developed their open-source, transformer-based model. Instead of building models from scratch, we fine-tune these pre-trained models on smaller datasets for specific tasks. This approach is known as transfer learning. This can save a lot of training time and computation resources [8].

These are some candidate models for the project.

Task	Models
1. Object detection	YOLOv8 (based on CNNs)
2. Document classification	LayoutLMv3 (Transformer-based) DONUT (Transformer-based)
3. Token classification	LayoutLMv3 (Transformer-based) DONUT (Transformer-based)

Table 2: Tasks and candidate models.

For object detection, YOLO (You-Only-Look-Once) model is a candidate model [9]. We discovered that there were three FYPs in last year utilizing the YOLO model for object detection. Previous success stories give us more confidence in applying this model in our project.

For document classification and token classification, two candidate transformer-based models are LayoutLMv3 [10] and DONUT [11]. They achieved outstanding results in benchmark tests but the underlying principle is different from each other.

2.2. Highlight of the proposed solution

The proposed solution is to develop a document processing system for extraction of the booking document uploaded by the shippers. The extracted data can be incorporated into the current platform for terminal operations.

We made a sketch on our proposed solution to facilitate discussion on system design.

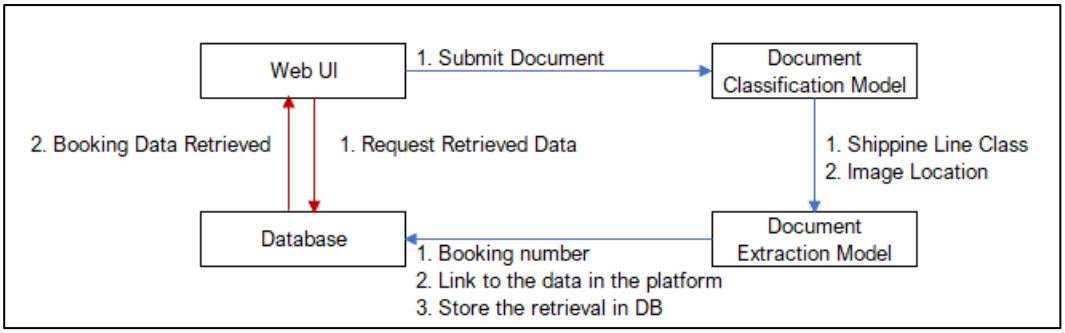


Figure 2: A sketch on the proposed solution (This is not a formal diagram)

From the figure, the following core components are necessary to build our solution.

1. **Web UI**

   The shippers have to upload a booking document in PDF format in a web interface and view retrieved booking afterwards. It is necessary to build a web interface for submission and data retrieval.

2. **AI Models**

   Our document processing system consists of two AI model components - A document classification model for shipping line classification and a document extraction component for booking number extraction.

   The document classification component aims to classify booking documents into shipping lines. An easy approach is to identify the company logo of the booking document, which is unique for each shipping line. There are various object detection models available on the market including the YOLO (You-Only-Look-Once) models.

   The document extraction component aims to extract the booking number for each booking document. It requires an understanding of the text information and the document layout. LayoutLMv3 and DONUT are candidate models for this component.

   We fine-tune the models using labelled dataset propagated from the original booking document samples.

3. **Web Server / Model Server**
   To host the web application and run the AI models, we have to set up a web server for the web application and a model server or service endpoints for calling the AI models and returning results to the web application.

4. **Back-end database**

   It is necessary to create a database for storage booking data retrieved by the users on the booking retrieval system.

The methodology and system design are further elaborated in Part 3.

3. Methodology

3.1. Requirements, Supporting Technologies and Technical Gap

A. Requirements

To develop a system, it is necessary to define the system requirements. It is the foundation of all software engineering activities.

In this project, we have to develop a document processing system web application using AI-based technologies that allows shippers to retrieve booking document information in an automated and digitalized manner.

The primary actor is the shipper, who uploads booking documents for booking retrieval using the web interface provided by the system. The custom AI model is a supporting actor that acts as a service endpoint providing document classification and extraction service for the system.

After requirement elicitation, we conclude with the following use case diagram.

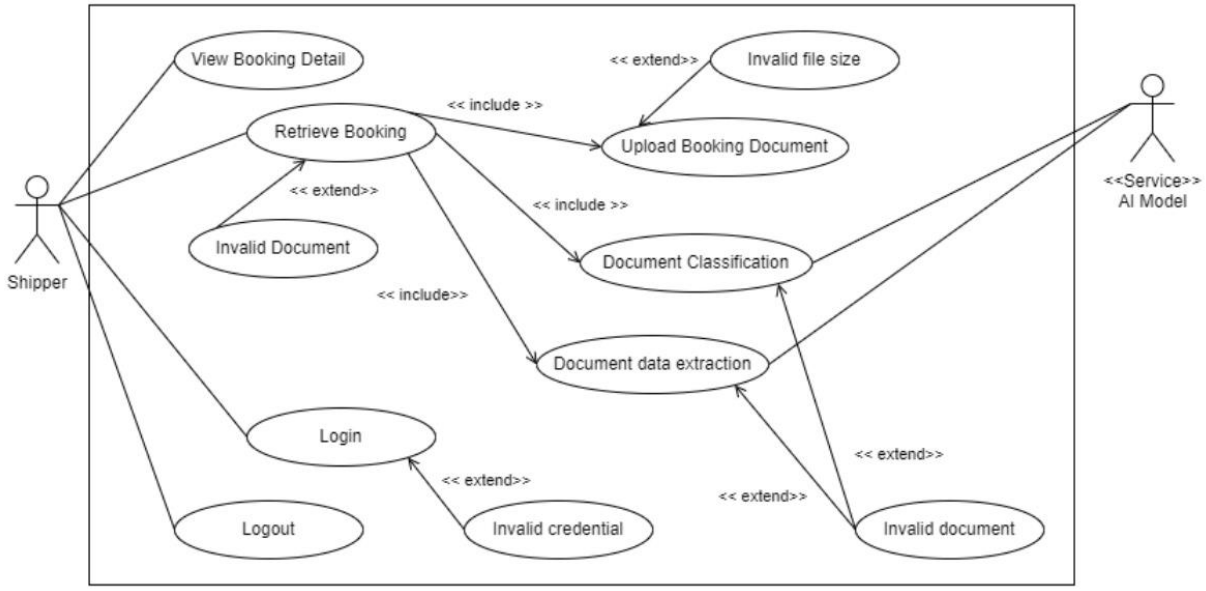


Figure 3: Use case diagram of the document processing system

The following tables summarize the use case description of core use cases.

Use Case name	Retrieve Booking
Primary actor	Shipper
Participating actor	AI Model <<Service>>
Flow of events	1. The shipper uploads the booking document (include Upload Booking Document) to the AI model service. 2. The AI model performs document classification by the booking document. (include Document Classification). 3. The AI model performs data extraction by document classification result. (include Document Data Extraction). 4. The web UI shows the result from the AI model service. 5. The retrieved booking is linked to the corresponding booking data in the extract and import platform.

	6. After retrieval and linking, the Shipper can view the booking details (use case: View Booking Detail)
Pre-conditions:	1. Shipper has logged in to the system with valid user id and password. 2. Shipper has a booking document ready to upload.
Post-conditions:	1. Shipper has successfully submitted the booking document and received the predicted result on the upload panel.

Use Case name	**Upload Booking Document**
Primary actor	Shipper
Participating actor	System
Flow of events	1. The Shipper clicks the “Choose file” button in the user booking list page to upload a booking document. 2. The system prompts the Shipper to select a booking document in the local drive to upload. 3. The Shipper selects the desired booking document and clicks the “Upload” button. 4. The system validates the file extensions and file sizes. 5. The system saves the document.
Alternative Flows:	At step 4, if the uploaded booking document has an invalid format, the system will reject the uploaded document and show an error message (Invalid document) to the shipper. Then back to step 1.
Pre-conditions:	1. Shipper has logged in to the system with a valid user id and password.

Use Case name	**Document Classification**
Primary actor	AI model
Participating actor	None
Flow of events	1. The AI model converts the first page of booking document to image. 2. The AI model predicts the class of shipping line from the image. 3. The AI model validates the confidence score. 4. The AI model outputs the image and the class of the shipping line.
Alternative Flows:	At step 3, if the confidence score of the booking document is not enough. The AI model will return a fail response to the system with fail reason “Invalid Document”.
Pre-conditions:	1. The AI model receives a request with a booking document.

Use Case name	**Document Data Extraction**
Primary actor	AI model
Participating actor	None
Flow of events	1. The AI model selects the data extraction model based on the class of the shipping line. 2. The AI model applies OCR to the image. 3. The AI model use data extraction model to get target booking number. 4. The AI model validates the confidence score. 5. The AI model outputs the class of the shipping line and the booking number.

Alternative Flows:	At step 4, if the confidence score of the booking document is not enough. The AI model will return a fail response to the system with fail reason “Invalid Document”.
Pre-conditions:	1. The AI model receives the image and class of the shipping line.

Use Case name	**View Booking Detail**
Primary actor	Shipper
Flow of events	1. Shipper selects the option to view booking data 2. The system displays the details of each booking, including the booking number, vessel information and date of retrieval. 3. The Shipper can filter and sort the retrieved booking by different criteria. 4. After retrieval, the Shipper can log-out the system or return back to the main page.
Pre-conditions:	1. Shipper has logged in to the system with valid user id and password. 2. Shipper submitted booking for retrieval previously.

Table 3: Use case descriptions

Functional requirements:

1. Shipper can login the document processing system with a valid company account. Only authenticated users can access the system.
2. Shipper can logout the document processing system after use.
3. Shipper can upload booking document(s) in valid PDF format on the system platform. The maximum size of a PDF file is limited to 2MB.
4. The system can classify the shipping line of the booking document and extract the booking number as the key for retrieving booking.
5. The system shows an error message when the uploaded booking document is invalid.
6. Shippers can view the details of retrieved booking.

Non-functional requirements:

1. The shipper can view the details of retrieved booking after submission of the booking document for no more than 20 seconds.
2. The system is easy to learn. The average training time should be less than 1 hour.
3. The system availability must be greater than 95%.
4. The accuracy of document classification must be greater than 85% for general booking documents.

B. Supporting Technologies

To establish the system, the essential supporting technologies are as follows:

Core infrastructure

1. PHP Web Server
   Shipper accesses the system platform on the Internet. Hence, we need to host the web application in a web server with associated dependencies. We can either deploy the web application in the cloud or on a personal server.

2. WSGI (Web Server Gateway Interface)
   It is necessary to have a WSGI between the web server and the python application to run the application successfully. We use the Python Flask Framework (WSGI supported) to create AI service endpoints for the web application to call.

3. MySQL database
   We also need a MySQL database for user accounts, the booking details from terminals and the booking ownership. Considering the workload and data storage, a simple MySQL database is sufficient.

Hardware components

1. GPU (Nvidia RTX 4070 Ti for demo system)
   In machine learning, matrix multiplication and addition are fundamental for tasks like weight updates, forward and backward passes in neural networks, and optimization algorithms.

   GPUs consist of thousands of processing units called CUDA cores that work concurrently. This parallel architecture allows them to handle large-scale matrix operations efficiently. Nvidia RTX 4070 Ti is a powerful GPU designed specifically for high-performance computing and deep learning tasks. We use Nvidia RTX 4070 Ti to accelerate training and execution speed of AI models.

Software components

1. PHP Web MVC Framework
   To demonstrate the prototype of the system, we build a simple web application using the PHP Laravel Web MVC (Model-View-Controller) Framework. With the use of a framework, we can develop the content management system with minimal functions quickly, reducing both development and maintenance efforts.

   Laravel Breeze is a minimal, simple implementation of Laravel's authentication features, including login, registration and password reset. We apply this framework plug-in to create a simple login page for authentication of the prototype.

2. Python and its available modules, frameworks and transformer-based models
   Python is well-known for machine learning projects. The language is close to natural language and there are so many useful modules and frameworks available on Python, including image augmentation, pdf conversion, dummy dataset generation and frameworks for AI model development such as PyTorch.

   The pre-trained models built on these frameworks can also be employed by just a few lines of codes. For example, YOLOv8 for object detection and LayoutLMv3 for document classification and token classification. The details of modules used are included in part 3.3 Implementation Issue.

3. Dataset labelling tools (Roboflow, Label Studio)
   To train custom models such as YOLOv8, we have to label the position and class of our dataset, which is tedious without appropriate tools.

   Fortunately, there are labelling tools such as Roboflow and Label Studio that reduce the effort by using drag and drop to select the appropriate region.

C. Technical Gap

There are hundreds of shipping lines worldwide. Booking document from each shipping line contains core information such as booking number, vessel name, loading and discharging venues, estimated date of arrival / departure and so on. However, field positions and general layout varies from one to another.

To classify booking documents into shipping lines, it is necessary to find out the unique characteristics of booking document for each shipping line. For example, position of a certain field, company logo, shipping line address and so on.

A proper way is to identify the company logo, which is unique for each shipping line booking document. Object detection is an area where AI models can achieve similar or even better performance than human beings. The execution speed of an AI object detection model is faster than models that focus on text and layout.

Therefore, the document classification component will be performed by YOLOv8, the latest version of YOLO (You-Only-Look-Once) object detection model. YOLOv8 is user-friendly and easy to configure, train and execute.

Using other characteristics such as field position involves more complicated setting and training. Although the execution speed is lower than YOLOv8, we will apply LayoutLMv3 document classification model as a cross checking to enhance correctness of the component if the booking document has passed the YOLOv8 examination.

The success of this project also depends on whether our system can extract the booking number for each type of booking document. There are various state-of-the-art technologies and pre-built transformer-based models available. We can fine-tune these models to fit our use case. One of the differences is whether the model applies OCR (Optical Character Recognition) for text analysis before prediction. For example, LayoutLMv3 versus DONUT.

LayoutLMv3 is a pre-trained multimodal transformer model that uses unified text and image masking. It can perform a variety of tasks including document classification, document parsing and token classification. It requires OCR for text analysis before the tasks. The analysis combines both text analysis and layout analysis. It achieves state-of-the-art performance in various tasks such as form understanding, receipt understanding, document visual question answering, and document layout analysis [10].

DONUT, on the other hand, is an OCR-free model that does not require OCR for text recognition of input image. The disadvantages of OCR are computationally expensive and any error in the OCR will propagate to the subsequent process [11].

Both models have their advantages and disadvantages. We selected LayoutLMv3 because there are more use cases provided on the Internet and it has a better performance for text analysis of structured data, e.g. form content.

3.2. System Design

A. System components

The component diagram classifies the system into several components based on the functionality. Each component is responsible for a particular task in the system and interacts with each other to complete the system aim.

In our system design, there are six components. The interactions are shown in the following figure.

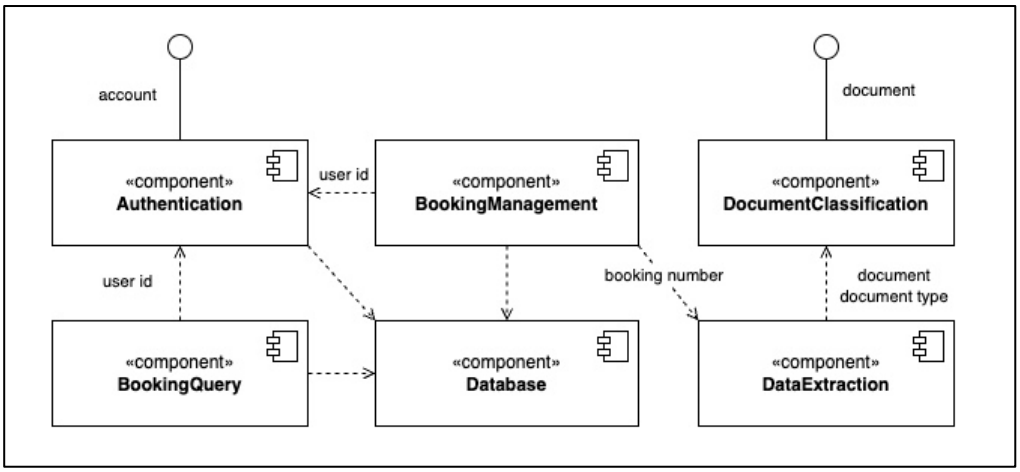


Figure 4: Component diagram of the document processing system

1. The **DocumentClassification** component provides an interface to receive digital booking document in PDF format.
2. The **DataExtraction** component depends on the output of the document type (i.e. shipping line class) in the document classification component.
3. The **BookingQuery** component is for querying booking information retrieved by the user id.
4. The **BookingManagement** component manages the ownership of the booking. The ownership of a booking should be the user who uploaded a document with the booking number.
5. The **Authentication** component provides an interface to receive account credentials for linkage of user id in BookingQuery component and BookingManagement components.
6. The **Database** component stores the information necessary for booking query, authentication and booking management.

B. Data Flow of the system

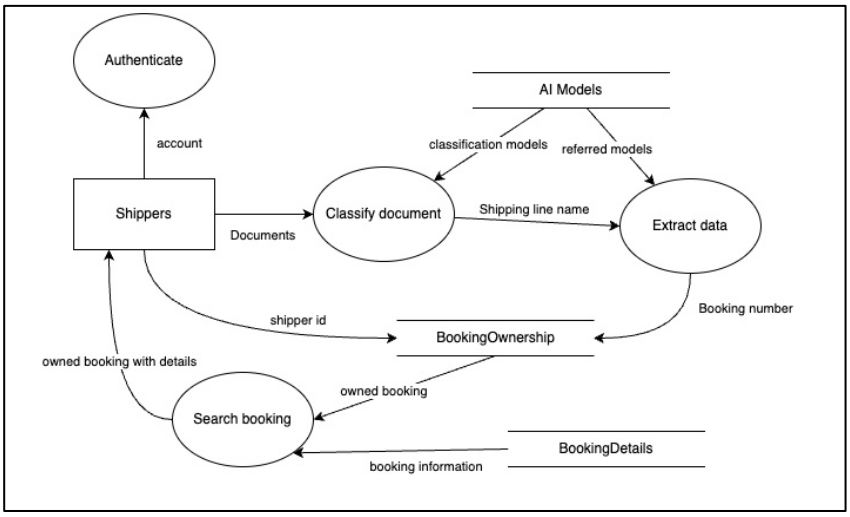


Figure 5: Data flow diagram of the document processing system

1. Shipper logins the web application using the company's user id and credentials.
2. Shipper get the access rights to upload the booking document for booking retrieval.
3. The booking document is classified into shipping lines by loading and running the AI classification model.
4. The booking document, with known shipping line class, is extracted by loading and running the AI document extraction model to extract the booking number.
5. The BookingOwnership is established between booking number and the shipper's user id.
6. The key is used to extract the booking information from booking details in the database of the export and import platform.
7. The retrieved booking of the shipper is stored in the database for review afterwards.

C. System Architecture

The system architecture of the document processing system is a typical three-tier architecture, consisting of the presentation tier, application tier and the data tier.

The presentation tier is responsible for the user interface and communication with the application. It is used to display the booking information and allow the users to upload the booking documents.

The application tier is responsible for processing, rule checking and notification by the application. There are two servers in the application tier. One is the web server for running the PHP Laravel Web application. Another one is to host python application for document classification and document extraction, serving as a service endpoint to the system.

The data tier is responsible for storage, retrieval and management of persistent objects or information. All data retrieved successfully will be saved in the database for the user to review after uploading. The benefits of a three-tier architecture is that each tier can be developed separately without impacting other tiers.

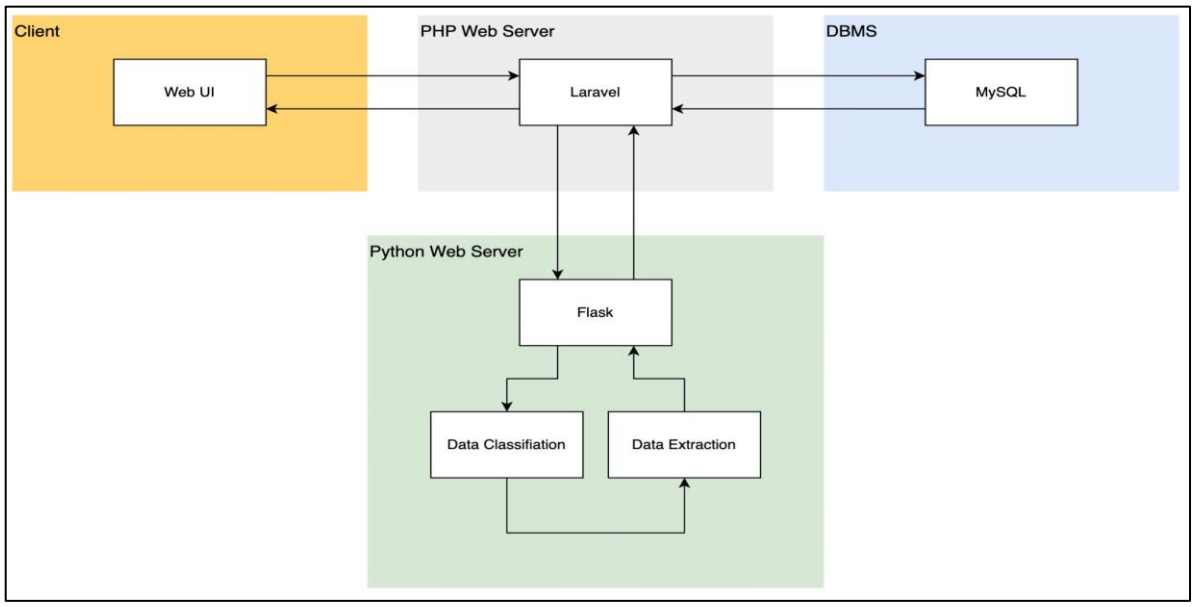


Figure 6: System architecture of the document processing system

3.3. Implementation Issue

To implement the prototype of the document processing system, we use the following programming languages and run-time dependencies:

Component	Name of the dependencies / technologies
Web Server	PHP Web Server
AI Model service endpoints	Python Flask framework (WSGI supported), backed up by Nvidia RTX 4070 Ti to execute the models.
DBMS	MySQL
Web Application	Laravel (Web MVC) Framework
Programming languages	PHP (for core web application) Python (for machine learning / AI tasks)
Libraries	Pdf2image (for conversion of pdf to jpeg) cv2 (for image augmentation) ultralytics (for YOLOv8 training and production run) Pytesseract (OCR Engine) PyTorch (Machine learning framework)
Pre-trained AI models	YOLOv8 (Object detection) LayoutLMv3 (Document classification and token classification)

Table 4: Various components for System Implementation.

3.4. Evaluation Methods and Design

For the prototype system, we will evaluate the functional requirements and non-functional requirements to see whether the system achieves the requirements.

For functional requirements, basic functionalities such as uploading booking documents in PDF format and viewing retrieved booking are evaluated. We also evaluate the cases when the uploaded document is invalid and when the AI models cannot return a predicted result.

Since shippers are the end users of the system, we can invite 20 shippers' representatives to carry out a user acceptance test (UAT) using the booking documents from the four shipping lines and their working procedures. After that, all shippers will fill in a questionnaire providing test results for each test case. If the average score exceeds 80%, then most of the functionalities are provided in the prototype system. Only minor changes will be implemented and the project is considered successful in terms of functionalities.

Functional Requirements Evaluation:

Evaluated Task	Contents
Function Testing	**Participants:** We will invite 20 shippers to use the prototype system. **Testing Details:** Shippers will test all necessary functions specified in the functional requirements. **Feedback:** Shippers will fill in a questionnaire to evaluate whether the system meets their expected functionalities. **Success Criterion:** If the average score exceeds 80%, the system implementation is considered successful with only minor changes needed.

Table 5: Functional Requirements Evaluation

Also, it is necessary for the system to satisfy the non-functional requirements to provide a good user experience and performance to the end users. As shown in clause 3.3, we will mainly focus on the average booking retrieval time, the ease of learning the system and the system uptime percentage. The evaluation methods and corresponding success criterion are shown in the below table.

Non-functional Requirements Evaluation:

Evaluated Task	Contents
Average Booking Retrieval Time	We will measure the time required to upload 30 booking documents and get the retrieve booking for each of the four shipping lines. **Success Criterion:** The average time for booking retrieval of one booking document should be less than or equal to 20 seconds
Usability	**Participants:** We invite 20 shippers to join the usability test. **Feedback:** Shippers will fill out a questionnaire to evaluate the time required to learn all functionalities of the system. **Success Criterion:** The average training time for shippers should not exceed 1 hour.
System uptime percentage	Check the system log to see if any abnormality occurs within three days of evaluation. **Success Criterion:** The uptime percentage must be greater than 95%.

Table 6: Non-functional Requirements Evaluation

Since it is a project that involves the use of custom AI models, it is necessary to evaluate the performance of the models, including the execution speed and the performance metrics such as precision, recall, accuracy, specificity, and F1-score.

We will prepare another 100 booking documents and 20 documents with empty class as the background. We will evaluate different performance metrics and our focus is on the accuracy. The accuracy of the model must be greater than 0.80 for each model.

Custom AI Models Performance Evaluation:

Evaluated Task	Contents
Performance Metrics	**Testing Set:** We'll prepare 100 extra booking documents for statistical analysis, with 25 booking documents from each class of shipping lines. In addition, 20 empty datasets are included. **Metrics:** Precision, recall, accuracy, specificity, and F1-score will be assessed for the custom AI models. **Success Criterion:** The accuracy of the model must be greater than 0.80.

Table 7: Custom AI Models Performance Evaluation:

This comprehensive evaluation approach ensures both functionality and performance are thoroughly assessed.

References

[1] Transport and Logistics Bureau, The Government of the Hong Kong Administrative Region, *Action Plan on Maritime and Port Development Strategy*. Hong Kong: Transport and Logistics Bureau, Dec. 2023.

[2] DollarAndSense Business, *Guide To Tuas Mega Port – World's Largest Fully Automated Port In Our Backyard,* 2023. [Online]. Available: https://dollarsandsense.sg/business/guide-to-tuasmega-port-worlds-largest-fully-automated-port-in-our-backyard [Assessed Feb. 2, 2024].

[3] Sydle One, *13 Advantages of Cloud Computing*, 2022. [Online]. Available: https://www.sydle.com/blog/advantages-of-cloud-computing-6154ae3d0e94dd1b0fa01fd6 [Assessed Jan. 16, 2024]

[4] Google Cloud, *Compare AWS and Azure services to Google Cloud*, 2023. [Online]. Available: https://cloud.google.com/docs/get-started/aws-azure-gcp-service-comparison [Assessed Jan. 20, 2024]

[5] Google Cloud, *Document AI*, 2023. [Online]. Available: https://cloud.google.com/document-ai?hl=en#key-features [Assessed Jan. 20, 2024]

[6] Ajay Bhatt, "Document Automation Using Artificial Intelligence", *IJRASET*, Volume 10 Issue IX, 2022.

[7] A. Vaswani, *et al*. "Attention is all you need." *Advances in neural information processing systems*, vol. 30, 2017

[8] Hugging Face, "How do Transformer work?", 2021. [Online]. Available: https://huggingface.co/learn/nlp-course/chapter1/4 [Assessed Jan. 20, 2024]

[9] Deci AI, *The History of YOLO Object Detection Models from YOLOv1 to YOLOv8,* 2023 [Online]. Available: https://deci.ai/blog/history-yolo-object-detection-models-from-yolov1-yolov8 [Assessed Feb 02, 2024]

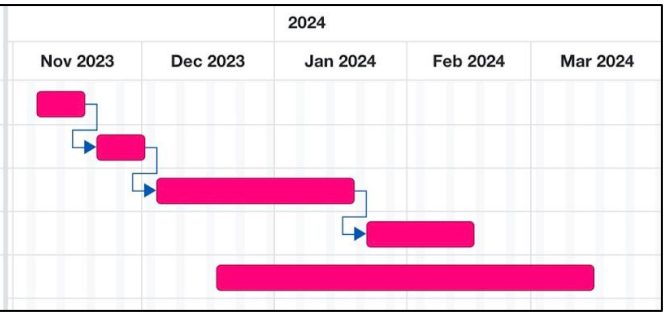
[10] Huang, Y, et al. "Layoutlmv3: Pre-training for document ai with unified text and image masking." *Proceedings of the 30th ACM International Conference on Multimedia*. 2022.

[11] Kim, G, et al., "OCR-free Document Understanding Transformer", *European Conference on Computer Vision*, Cham: Springer Nature Switzerland, 2022.

Appendix A. Overview of Project Progress

Original Gantt Chart

Name	Start Date	End Date
1. Collect data and build datasets	Nov 06, 2023	Nov 17, 2023
2. Setup AI training environment	Nov 20, 2023	Dec 01, 2023
3. Train, test and evaluate models	Dec 04, 2023	Jan 19, 2024
4. Select the best models	Jan 22, 2024	Feb 16, 2024
5. Package and evaluate the system	Dec 18, 2023	Mar 15, 2024



Tasks completed and tasks ongoing:

The following lists the tasks completed:

1. Collect data and build datasets (Completed on schedule, Nov 17, 2023)
    a. Collection of samples of booking documents from four shipping companies (COSCO, Evergreen, OOCL and TSLines). The number of companies changed from 5 to 4 for concise purpose.
    b. Propagate the dummy dataset to 20 booking documents per company using Python script.
    c. Conversion of PDF to JPEG for the first page of each document and carry out image augmentation to mimic variations and document quality.
    d. Grouping the images into training set, validation set and testing set for document classification component and document extraction component of the system.

2. Setting up the AI training environment (Completed on schedule, Dec 01, 2023)
    a. Setting up the environment to utilize GPU for training our AI model.
    b. Searching for some techniques that may speed up the training and execution speed. Finetuning the hyperparameters during the training and detection part.

3. Train, test and evaluate models (Completed 75% of the tasks)
    a. Train the YOLOv8 object detection model with different sets of custom dataset and batch number and evaluate the confidence of results from the testing dataset.
    b. Evaluate the YOLOv8 object detection model versus LayoutLMv3 in document classification tasks.
    c. Train the LayoutLMv3 token classification model and test with current dataset to see whether the booking number can be correctly extracted.

4. Select the best models (Completed 50% of the tasks)
    a. Select the best model of YOLOv8 model for document classification and screening task.
    b. Select the temporary best LayoutLMv3 models for document classification and data extraction based on the accuracy.

5. Package and evaluate the system (Completed 50% of the tasks)
    a. Setup PHP Web server, Python WSGI server and MySQL Server on the local machine.
    b. Completed authentication module. (Allow users to register/login/logout)

c. Embedded the selected AI model to perform both document classification and data extraction features in Python WSGI server.
d. Completed the normal flow of the system. (Allow users to upload documents and query their owned bookings)

The following lists the ongoing tasks:

For task 3 and task 4, the ongoing tasks are to evaluate the best model of trained LayoutLMv3 and then select the best model of trained LayoutLMv3. (~25% of task 3 and ~50% of task 4)

For task 5, the system still needs to add more modules to optimize user experience, handle exception cases when confidence score is not enough during retrieve booking. Also, validation more fields in booking documents that can increase data security for the system.

Appendix B. Revised Project Plan

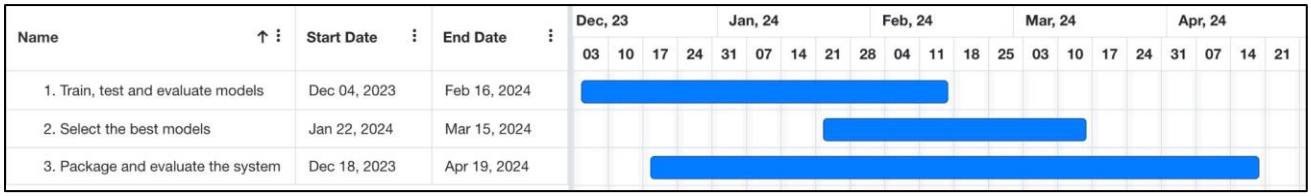
Revised Gantt Chart for remaining tasks

We completed two tasks before submitting this report, which are ‘Collect data and build datasets’ and ‘Setup AI training environment’.

There are three main remaining tasks. The task ‘Train, test and evaluate models’ is nearing completion. To prepare the demo of our project, we proceed with ‘Package and evaluate the system’ ahead of the planned schedule. Then, the system will be enhanced by handling invalid documents and adding other features.

Also, we have to spend more time on ‘Select the best models’ in order to ensure the performance of the AI models.

We have to spend more time on “Train, test and evaluate models” and “Select the best models” than expected. Therefore, the task on “packaging and evaluate the system” will be delayed by one week.



Roles

Roles	Member(s)	Remarks
Team coordinator	Eraka Yip	Manages the project in general and keeps records, reports, and other documents in order, and prepare the submission of reports
Team members (Roles: Designer, Programmer, Tester and System Analyst)	Eraka Yip, Timothy Chan, Andy Lau	All team members will participate in designing, coding, testing and deploying the system.

Responsibilities and Task Assignment

Tasks	Responsible Member(s)	Target Date
Train, test and evaluate models	Eraka Yip, Timothy Chan, Andy Lau	Feb 16, 2024
Select the best models	Eraka Yip, Timothy Chan, Andy Lau	Mar 15, 2024

Tasks	Responsible Member(s)	Target Date
Package and evaluate the system	Eraka Yip, Timothy Chan, Andy Lau	Apr 19, 2024

Remarks: Tasks such as report writing and presentation demonstration are not included in the Gantt chart since they are not our project objectives. However, they are work-intensive and crucial for the success of our project.