Functional Specification

CA400 4th Year Project

OCR Text Recognition & Processing

Date Finished: 9/11/22

Student Name: David Weir

Student ID: 19433086

Student Name: Róisín O'Rourke

Student ID: 19360491

Table of Contents

1. Introduction

1.1 Overview

The system we are creating is an Optical Character Recognition (OCR) text to speech reading aid. It has the capacity to scan printed and handwritten text, detect the text language, translate the text into a set sample of languages and then convert the text to mp3/pdf. This will allow users to convert handwritten documents into pdfs and listen to them in audio format for studying purposes or to allow the visually imaired to understand handwritten notes.

In an educational setting, the system will be able to help students who wish to listen to their notes and study materials as they revise. The translation abilities of the system will also especially aid international students who can scan books/notes not provided in their home language and then listen to them in their own language. The system could also allow visually impaired lecturers to scan written exam papers and correct them by listening to them or it could assist visually impaired students in their studies.

1.2 Business Context

The system is aimed at users in third level education. Therefore businesses which may affiliate with the product are colleges and universities.

1.3 Glossary

Term	Definition
Optical Character Recognition (OCR)	The process of converting a digital image of text (usually in pdf, text or image format) into machine readable data (i.e. a text file).
Machine Readable Data	Machine-readable data, or computer-readable data, is data in a format that can be processed by a computer.

ISO 639	A standardised nomenclature used to classify languages. Each language is assigned a 2 letter (639-1) and a 3 letter (639-2 and 639-3) lowercase abbreviation. (e.g. English = 'en' and 'eng').
Machine Learning Model	A file that has been trained to recognize certain types of patterns. It is trained using a large dataset and taught to see these patterns so it can recognise them later in a previously unseen dataset.
Machine Translation	The use of software to translate text or speech from one language to another.
System Architecture	A system architecture is the conceptual model that defines the structure, behaviour, and more views of a system.
High Level Design (HDL)	Explains the architecture that would be used in the development of a system. Provides an overview of the system identifying the main components of the system.
Operational Scenario	A description of an imagined sequence of events including the interaction of the product or service with its environment and users.
Context Diagram	Displays the interactions between a system and external actors that the system will interact with.
Data Flow Diagram (DFD)	Represents the flow of data through a system. Provides information about the inputs and outputs of each entity.
GANTT Chart	A Gantt chart is a type of bar chart that illustrates a project schedule and also shows the dependency relationships between activities.

2. General Description

2.1 Product / System Functions

The system is a python based application that will be capable of detecting text from uploaded files using OCR and converting them into mp3 and/or pdf format. The program will be able to process images of printed and handwritten text and pdfs. It will then detect the language of the processed text. Using the detected language, the system will be able to translate the text into a number of different languages, if requested by the end user. Finally, it will convert the text to pdf format and/or create a text-speech mp3 file so it can be listened to by the user.

2.2 User Characteristics and Objectives

The expected user would be students or the visually impaired, however, the end product may be useful to many other user types including employees at work or the everyday person who wishes to convert their written lists or books to a pdf version accessible on their device. The user must be familiar with how to select files and upload them to the application from their device, however no other expertise will be required to successfully use the application.

2.3 Operational Scenarios

2.3.1 Upload Documents

Use Case Name	Upload Documents
Use Case Overview	The first step in converting documents into mp3 format or pdf is uploading them from the user's device. Once the user has uploaded the documents or images of text, the program will extract the text using OCR technology.
Actors	- User - GUI - OCR Model
Trigger	The user wishes to upload their documents in order to convert

	them.
Precondition	● The user has the application installed on their device. ● The user has images of text or documents that they wish to upload on their device.
Basic Path	● The user opens the application. ● The user selects whether they are uploading a document that contains print text handwritten text. ● The user selects the option to upload documents. ● The user selects what they wish to upload. ● The system extracts the text in the documents using OCR.
Alternative Path	Error notification appears to the user if upload is unsuccessful informing them the file upload has failed.
Postcondition	● The user has uploaded their selected documents. ● The text has been extracted from the documents.

TABLE TWO: Use Case 1: Upload Documents

2.3.2 Translate Language

Use Case Name	Translate Language
Use Case Overview	If the user wishes to translate the language of the text in the uploaded documents it is possible in this step. The system will first need to recognise the language in the documents and then will be able to translate text into one of a set of languages available.
Trigger	The text from the uploaded files has been extracted using OCR.
Actors	● User ● GUI ● NLP Model

Precondition	● The system can recognise the language used in the text in the documents. ● The language the user wishes to translate to is one of the languages the system is equipped to translate to.
Basic Path	● Language detection is run on the extracted text. ● The system recognises the language used in the documents. ● The user selects from the available languages the one which they want to translate their documents into. ● The system translates the text in the documents into the chosen language.
Alternative Path	● If the system cannot detect the language, it will ask the user to enter manually the language of the text in the documents. ● The user does not wish to translate the text, and this use case is skipped
Postcondition	● The text has been successfully translated. ● The system cannot detect the language used in the documents and cannot translate them. ● The system cannot translate the text as the language the user wants is not available on the system.

TABLE THREE: Use Case 2: Upload Documents

2.3.3 Edit Documents

Use Case Name	Edit Documents
Use Case Overview	Once the text has been extracted from the documents, the user will have the opportunity to edit the documents and correct any errors or make changes. These errors may result from inaccurate OCR from handwritten text or blurry images of text.

Trigger	The user has uploaded the text documents and the system has extracted the text from them.
Actors	● User ● GUI
Precondition	● The user has successfully uploaded the documents. ● The system has extracted the text from the documents.
Basic Path	● The system displays the extracted text in a preview form as a text file to the user. ● The user has the ability to make any changes to the file that they want. ● Once the user is satisfied with the document, they save the changes made.
Alternative Path	The user does not wish to make any alterations so this use case is skipped.
Postcondition	● The document has been updated and any changes the user wishes for have been made.

TABLE FOUR: Use Case 3: Edit Documents

2.3.4 Export Files

Use Case Name	Export Files
Use Case Overview	The final step in the process is exporting the files as a pdf and/or as an mp3 file to the user's local device. The user selects what format they want to export the files as and they are then converted and saved to the local device.
Trigger	The document has been uploaded to the application and any edits the user wants have been made.
Actors	● User ● GUI

Precondition	● The application has permission to export files to the user's local device.
Basic Path	● The user selects whether they wish to export the document as a pdf, or mp3, or as both. ● If audio format is selected, the system runs the file through a text-to-speech program to record an mp3 recording of the file. ● The file formats are downloaded to the user's local device. ● The user can view or listen to the document on their local device.
Alternative Path	N/A
Postcondition	● The user contains a copy of the document as a pdf and/or as an mp3 file so they can easily listen to it.

TABLE FIVE: Use Case 4: Export Files

2.4 Constraints

Time constraints

The application must be completed by the 28th of April 2022, as this is the due date of the assignment. Time will also need to be set aside to familiarise ourselves with the technologies being used in the project.

Accuracy constraints

We will be training a model to recognise handwritten text. This is an area of OCR that cannot, as of yet, be performed with complete accuracy. For example, it is extremely difficult to recognise cursive handwriting.

Language constraints

We will only be translating the text into a set sample of languages. Also, the model will only be trained to recognise handwriting written with English characters, so any handwritten notes uploaded to the application must be in English.

3. Functional Requirements

3.1 Accept different forms of input

Description: The proposed system must be capable of accepting pdfs and standard images as input files.

Criticality: **Essential -** This is a crucial functionality as without it the system will not be able to take in any form of user input.

Technical Issues: Difficulty in supporting file types.

Dependencies with other requirements: This function is not dependent on any other requirement.

3.2 Perform OCR on printed text

Description: The system performs an OCR script with a pretrained model to recognize print-style characters from the input file.

Criticality: **Essential -** This is a key functionality of the proposed system and must be implemented.

Technical Issues: The trained model must be robust enough to recognize characters from a range of fonts for all languages that will be supported by the system.

Dependencies with other requirements: This function is naturally dependent on Function 3.1 as it requires the input file to work.

3.3 Perform OCR on handwritten documents

Description: The system performs an OCR script with a pretrained model to recognize handwritten characters from the input file.

Criticality: **Essential -** This is a essential requirement to understand images of handwriting

Technical Issues: This is a new and unexplored region of OCR that is yet to be perfected. As such there are a number of possible technical issues including a difficulty in training a robust enough model.

We are also likely limited to only understanding modern English latin script characters and by extension limited to reading English handwriting characters only.

Dependencies with other requirements: This requirement is dependent entirely on Requirement 3.1 again, as we require the input file to work with.

3.4 Detect the language(s) of the inputted files

Description: The system must be capable of detecting the language of the input file and linking it to its corresponding ISO 639 language code.

Criticality: High - This is an important requirement of the system as other requirements are dependent on it.

Technical Issues: The detection process must be suitably fast and accurate.

Dependencies with other requirements: Depending on the input file type, this requirement will be dependent on Functions 3.2 and 3.3.

3.5 Translate the input file to a user provided language

Description: The user will have the option to translate their input file into a different language of their choice (a small number of languages will be supported as proof of concept)

Criticality: High - This requirement is an important secondary function of the system. It is an additional function on top of the OCR requirements and will not necessarily always be used by the user.

Technical Issues: Difficulties in training a suitable accurate and robust model for the chosen set of languages.

Dependencies with other requirements: This requirement depends on requirements 3.2, 3.3 and 3.4, to provide the machine readable form of the input file and the ISO 639 code of the original language.

3.6 Display a preview of the output

Description: This function will allow the user to preview the expected text output after OCR and optional translation, in a machine readable format, as a quality check measure.

Criticality: Medium - This is an important quality of life requirement but it is not crucial to the functionality of the system as a whole.

Technical Issues: Providing a suitable preview format and the possibility of altering the expected output before final processing could prove to be possible technical issues.

Dependencies with other requirements: This requirement is dependent on all requirements before it.

3.7 Process the output

Description: Process the output into its final mp3 and/or pdf format.

Criticality: Essential - This is an essential requirement to provide the user access to the final output.

Technical Issues: Implementing an accurate and quality text-to-speech reader functionality for the mp3 file.

Dependencies with other requirements: This is dependent on all preceding requirements.

4. System Architecture
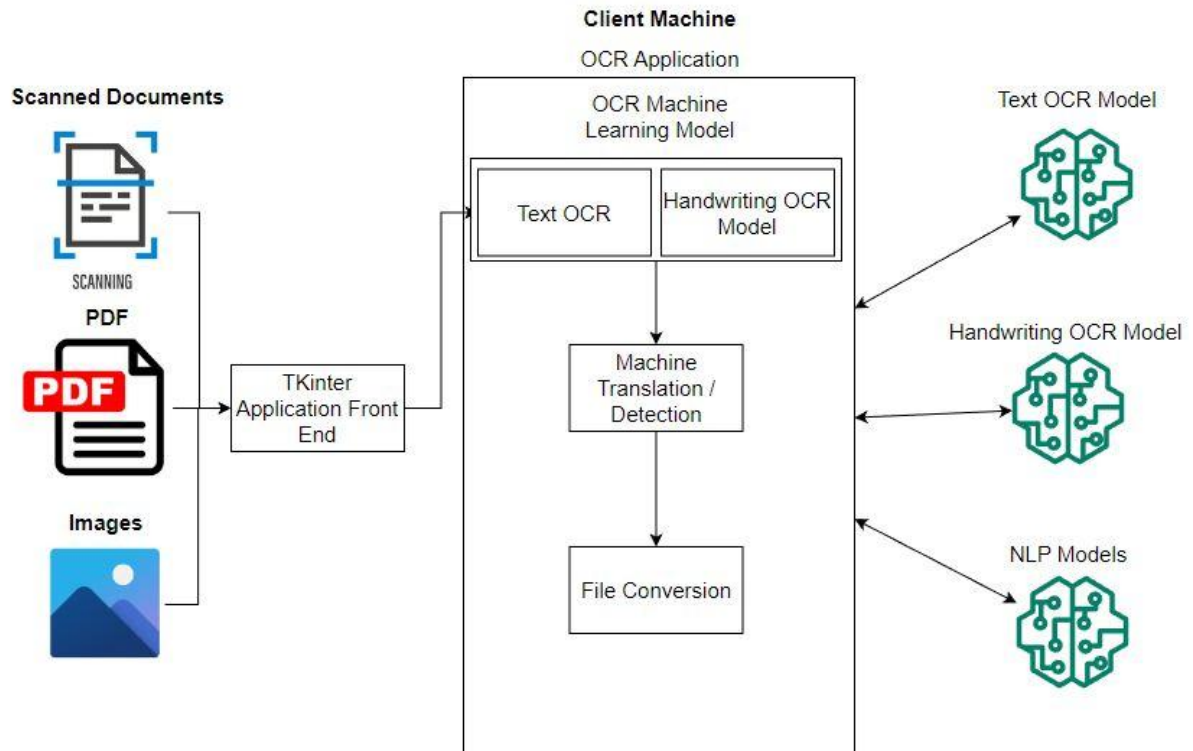
DIAGRAM ONE: System Architecture

5. High-Level Design (HLD)

5.1 Context Diagram
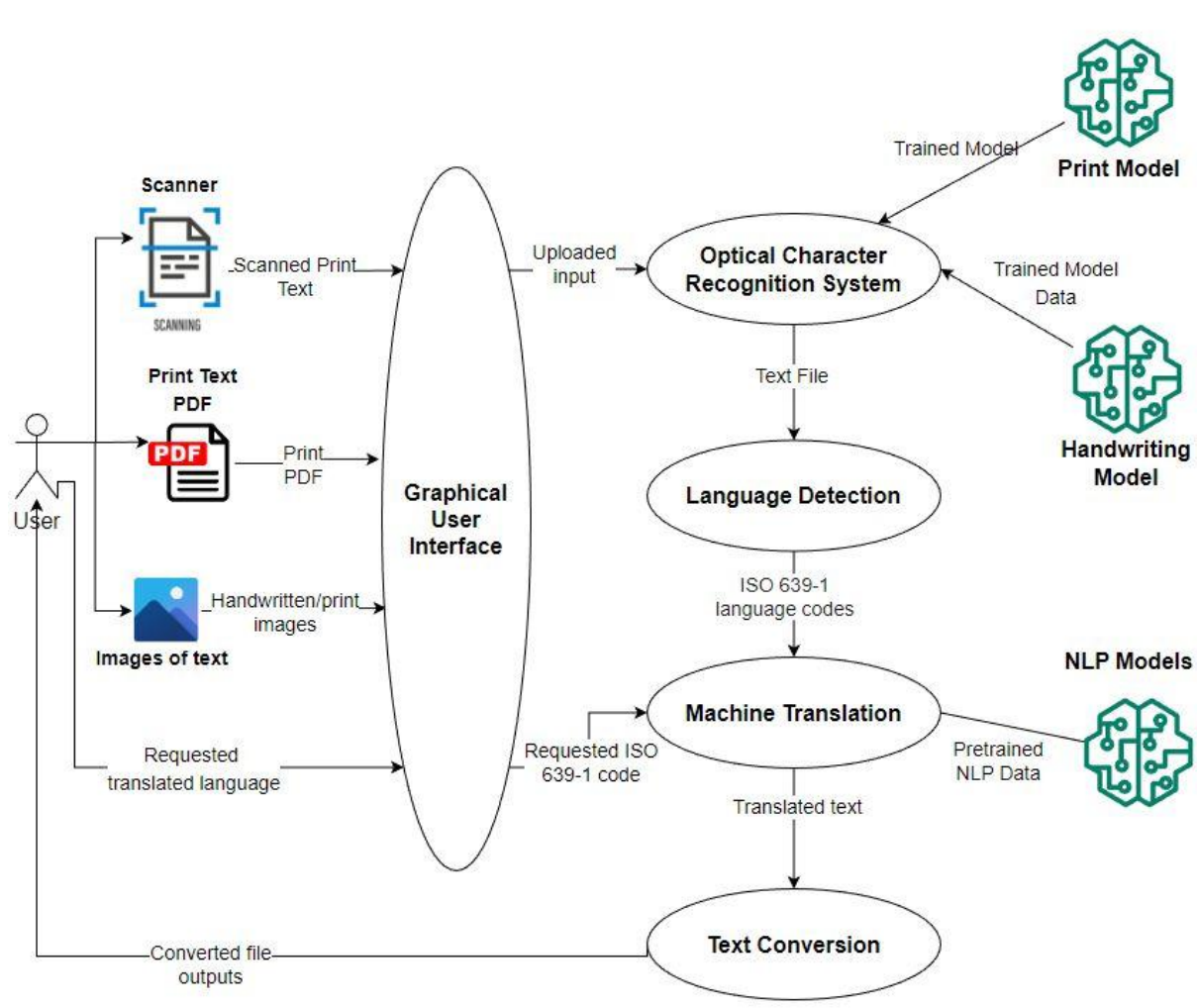
DIAGRAM TWO: Context Diagram

5.2 Data Flow Diagram (DFD)
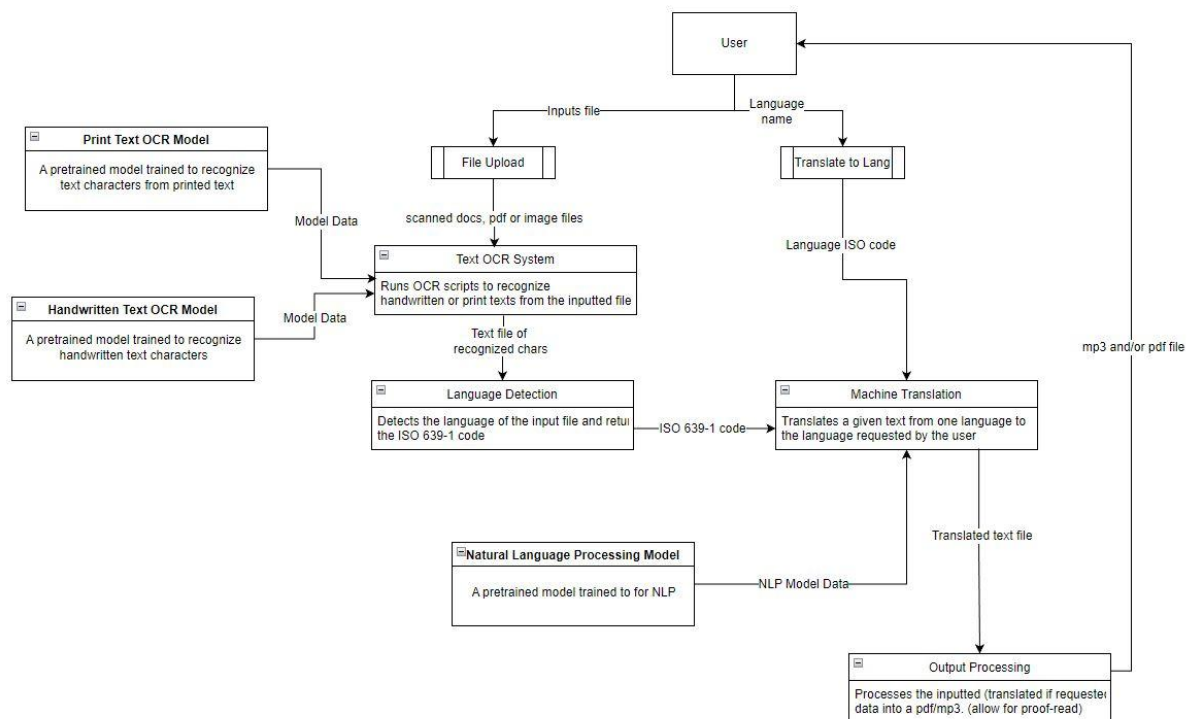
DIAGRAM THREE: Data Flow Diagram

5.3 High Level Design Description

- The user opens the Python application and accesses the GUI.

- The user chooses either the print text option or the handwriting option and uploads a file (of one of the supported file types).

- The system accepts the input and performs the indicated OCR script on the input file, creating a machine readable file (likely a .txt file).

- The system runs a script to determine the language of the input file and stores the correct ISO 639 language code.

- The user is asked if they want to translate the file into one of the supported languages.

  ○ If they say no this step is skipped.

- If they say yes, they are prompted to choose one of the provided languages. The system then translates the machine readable file into the requested language.

● The system displays a preview form of the machine readable file to the user to allow them to correct any errors.

● The system then runs the file through a text-to-speech program to record an mp3 recording of the file, and stores the mp3 and/or the pdf to the user's local machine.

6. Preliminary Schedule

6.1 GANTT Chart

CA400 4th Year Project	start	end	12%
OCR Text Recognition	22/09/22	06/05/23	12%
Research + Discovery	22/09/22	20/10/22	100%
Define Project Scope	22/09	30/09	100%
Supervisor Meeting	05/10	05/10	100%
Project Proposal form	10/10	13/10	100%
Proposal Panel Meeting (Approved)	20/10	20/10	100%
Functional Spec & Design Phase	31/10/22	22/12/22	12%
High Level Design & Diagrams	31/10	06/11	90%
Design Review	07/11	07/11	0%
Design Revisions (if needed)	08/11	09/11	0%
Finalize Functional Spec	10/11	10/11	0%
Semester 1 Assignments + Study	12/11	09/12	0%
Semester 1 Exams	10/12	22/12	0%
Development Phase	02/01/23	12/02/23	0%
Development Phase 1 (early de...	02/01/23	19/01/23	0%
Print Based OCR	02/01	04/01	0%
Handwritten text OCR	07/01	15/01	0%
Language Detection	04/01	06/01	0%
Phase 1 Testing	17/01	19/01	0%
Review	20/01	20/01	0%
Phase 1 Revision	23/01	30/01	0%
Development Phase 2	03/02/23	12/02/23	0%
Machine Translation	03/02	10/02	0%
Output Conversion	03/02	04/02	0%
Front End Development	06/02	12/02	0%
Review Phase 2	15/02	15/02	0%
Phase 2 Revision	20/02	24/02	0%
Testing + Revision Phase	25/02/23	26/03/23	0%
Inhouse Testing	25/02	28/02	0%
Revisions	02/03	05/03	0%
Review	06/03	06/03	0%
Testing	06/03	12/03	0%
Revisions	13/03	26/03	0%
Deployment Phase	28/04/23	28/04/23	0%
Deployment/Submission	28/04	28/04	0%
Semester 2 Exams	08/04/23	23/04/23	0%
Study Period	08/04	23/04	0%
Exam Period	24/04	06/05	0%