Department of Computer Science & Engineering
The University of Texas at Arlington

Project Charter
CSE 4316: Senior Design I
Spring 2021

Optical Profilers
Document Classifier

Muhammad Daud
Xavier Wells
Bishwamitra Sapkota
Koshish Khadka

Revision History

Revision	Date	Author(s)	Description
0.1	02.14.2021	XW	document creation

CONTENTS

1 Product Concept **7**
- 1.1 Purpose and Use . . . 7
- 1.2 Intended Audience . . . 7

2 Product Description **8**
- 2.1 Features & Functions . . . 8
- 2.2 External Inputs & Outputs . . . 8
- 2.3 Product Interfaces . . . 8

3 Customer Requirements **9**
- 3.1 Build UI and API . . . 9
  - 3.1.1 Description . . . 9
  - 3.1.2 Source . . . 9
  - 3.1.3 Constraints . . . 9
  - 3.1.4 Standards . . . 9
  - 3.1.5 Priority . . . 9
- 3.2 Utilize Storage Solution . . . 9
  - 3.2.1 Description . . . 9
  - 3.2.2 Source . . . 9
  - 3.2.3 Constraints . . . 9
  - 3.2.4 Standards . . . 9
  - 3.2.5 Priority . . . 9
- 3.3 Building classifying text model . . . 9
  - 3.3.1 Description . . . 9
  - 3.3.2 Source . . . 10
  - 3.3.3 Constraints . . . 10
  - 3.3.4 Standards . . . 10
  - 3.3.5 Priority . . . 10

4 Packaging Requirements **11**
- 4.1 Source Code and Availability . . . 11
  - 4.1.1 Description . . . 11
  - 4.1.2 Source . . . 11
  - 4.1.3 Constraints . . . 11
  - 4.1.4 Standards . . . 11
  - 4.1.5 Priority . . . 11
- 4.2 Documentation . . . 11
  - 4.2.1 Description . . . 11
  - 4.2.2 Source . . . 11
  - 4.2.3 Constraints . . . 11
  - 4.2.4 Standards . . . 11
  - 4.2.5 Priority . . . 11

5 Performance Requirements **12**
5.1 Classification Accuracy . . . 12
5.1.1 Description . . . 12
5.1.2 Source . . . 12
5.1.3 Constraints . . . 12
5.1.4 Standards . . . 12
5.1.5 Priority . . . 12
5.2 Database Query Time . . . 12
5.2.1 Description . . . 12
5.2.2 Source . . . 12
5.2.3 Constraints . . . 12
5.2.4 Standards . . . 12
5.2.5 Priority . . . 12
5.3 Document Classification Time . . . 12
5.3.1 Description . . . 12
5.3.2 Source . . . 13
5.3.3 Constraints . . . 13
5.3.4 Standards . . . 13
5.3.5 Priority . . . 13

6 Safety Requirements **14**
6.1 Laboratory equipment lockout/tagout (LOTO) procedures . . . 14
6.1.1 Description . . . 14
6.1.2 Source . . . 14
6.1.3 Constraints . . . 14
6.1.4 Standards . . . 14
6.1.5 Priority . . . 14
6.2 National Electric Code (NEC) wiring compliance . . . 14
6.2.1 Description . . . 14
6.2.2 Source . . . 14
6.2.3 Constraints . . . 14
6.2.4 Standards . . . 14
6.2.5 Priority . . . 14
6.3 RIA robotic manipulator safety standards . . . 14
6.3.1 Description . . . 14
6.3.2 Source . . . 15
6.3.3 Constraints . . . 15
6.3.4 Standards . . . 15
6.3.5 Priority . . . 15

7 Maintenance & Support Requirements **16**
7.1 Requirement Name . . . 16
7.1.1 Description . . . 16
7.1.2 Source . . . 16
7.1.3 Constraints . . . 16
7.1.4 Standards . . . 16
7.1.5 Priority . . . 16

8 Other Requirements **17**
8.1 Innovation . . . 17
8.1.1 Description . . . 17
8.1.2 Source . . . 17
8.1.3 Constraints . . . 17
8.1.4 Standards . . . 17
8.1.5 Priority . . . 17

9 Future Items **18**
9.1 RECOGNITION OF SIGNATURE . . . 18
9.1.1 Description . . . 18
9.1.2 Source . . . 18
9.1.3 Constraints . . . 18
9.1.4 Standards . . . 18
9.1.5 Priority . . . 18
9.2 MATCHING THE SIGNATURE . . . 18
9.2.1 Description . . . 18
9.2.2 Source . . . 18
9.2.3 Constraints . . . 18
9.2.4 Standards . . . 18
9.2.5 Priority . . . 18

List of Figures

1 X conceptual drawing 7

1 Product Concept

This section describes conceptual viewpoint of the system, its purpose and use, and intended audience. The system, document classifier, will be able to recognize files and categorize them into documents. From the high-level perspective, the system includes Identification of files and their classification. Purpose and use and intended audiences of the system are described below.

1.1 Purpose and Use

The name of the system we are going to design, and implement is document classifier. The purpose of the system is to classify files into specific documents. The uploaded files by the user will be classified based on text and the system is also expected to detect the presence of a signature. From the user point of view, user is supposed to input files for classification, allow system to process and share or save categorized documents.

1.2 Intended Audience

The system can be used in variety of fields by different customers. This benefits sectors, where documents are regularly used. Many fields including science, technology, business, management, and arts require the use of documents. Managing and separating these documents can be tedious sometimes. The system can be used from individual level to organizational level and intended for general use.

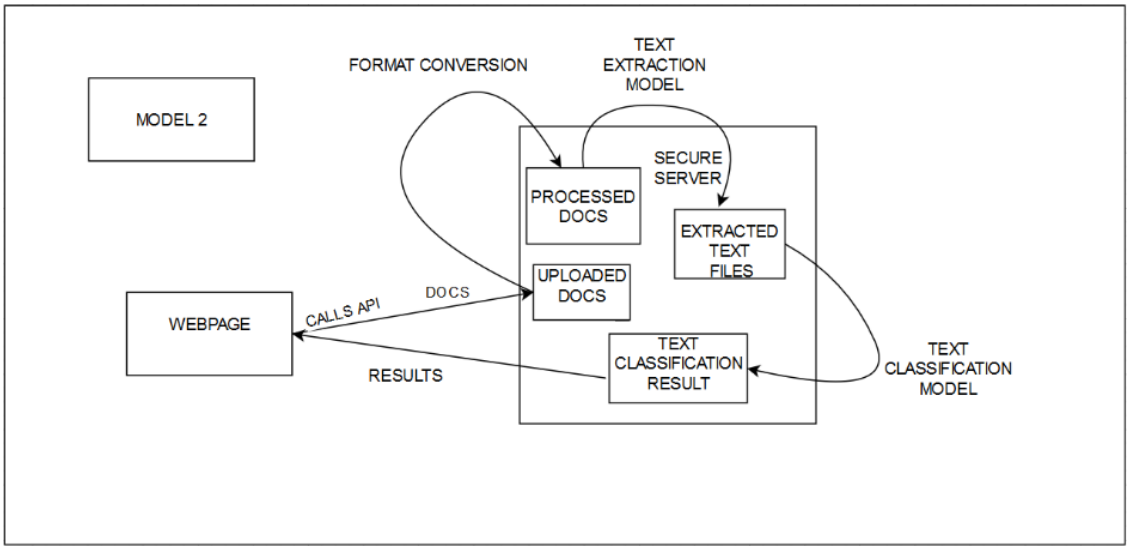

Figure 1: X conceptual drawing

2 Product Description

This section of the document will provide the reader with an overview of the Document classifier system. The operational goals of the project as well as the features of the product will be described in depth below. Most of the following will come directly from the "IdentiDoc" application specification. At the current time, the only detailed description available is from the IdentiDoc specification so the below will be a copied section, including figure 1.

2.1 Features & Functions

The Document classifier system is made up of several separate components. The front-end of the system will consist of a simple web application. This web application will facilitate all user interaction with the Document classifier system. The web application will be able send requests to upload documents and query for previous classification results. A RESTful API will facilitate all requests between the front-end and the back-end. Uploaded documents will be stored in a secured storage solution along with previous classification results. After uploading documents, they will be pre-processed to improve classification accuracy. The text will then be extracted from the documents and used to classify them. If the document is of a recognized type, then it will be scanned for a signature. These components are represented in Figure 1.

2.2 External Inputs & Outputs

The Document classifier system is only expecting pictures or pdf scans of documents as input from the user. The user will then receive the classification result of the system as output. The original file will be stored on the server. The user will also be able to query for classification results and the original documents.Data will exist in other forms in intermediate classification steps, but will not be stored.

2.3 Product Interfaces

The user will primarily just see the web interface for document classifier. Document classifier is not intended to have any administrator or maintainer, but if the system did, they would also see the storage solution and the classification model. The storage solution will be a database that will hold files and classification results. The classification result will include all pre-processing and text mining components. The uploaded documents will be stored in the database along with the classification results.

3 Customer Requirements

The original system requirements were given to team 1 for this project and they were given by the company State Farm. The basis of this project is creating an application to recognize and process text from a document to classify the document. In addition to recognizing & classifying a document, the program should also be able to identify if a signature is present within a certain section of the document.

3.1 Build UI and API

3.1.1 Description

We have to build a user interface to interact with for this program. We have to make it so the user interface can interact with the API. The API needs to have the ability for documents to be uploaded in the UI.

3.1.2 Source

The source of the requirement is StateFarm.

3.1.3 Constraints

There is not a detail for constraints to this requirement as long as it is appropriate for the project and professional looking.

3.1.4 Standards

N/A: no standards for this requirement.

3.1.5 Priority

Moderate priority: it should function properly, it is the backbone and the front end of the program.

3.2 Utilize Storage Solution

3.2.1 Description

The API needs to be able to write to a storage solution. It also needs to be able to connect the UI base to the storage system.

3.2.2 Source

The source of the requirement is StateFarm

3.2.3 Constraints

There is not a detail for constraints to this requirement as long as it is appropriate for the project and professional looking.

3.2.4 Standards

N/A: no standards for this requirement.

3.2.5 Priority

Low priority: it is nice to have but not necessarily a need but good to have due to storage need for the documents once classified.

3.3 Building classifying text model

3.3.1 Description

Building the structure to the classifying process, needing access to the storage location and process text documents. The model has to recognize a single document with some basic static and some variable

text as the same document. The scan should also be able to tell if the document is recognized or not, and it needs to have the ability to recognize if a signature exits.

3.3.2 SOURCE

The source of the requirement is StateFarm.

3.3.3 CONSTRAINTS

There is not a detail for constraints to this requirement as long as it is appropriate for the project and professional looking.

3.3.4 STANDARDS

N/A: no standards for this requirement

3.3.5 PRIORITY

Critical priority: This is a must have or the product will fail because this is the whole reason that the project even exists.

4 Packaging Requirements

All deliverables will be received by Dr. Shawn Gieser, and tentatively State Farm, by August of 2021. The system thus far will include continued work completed by the Identidoc Senior Design II team: the UI, API, and classification model. Our team will update the API, UI, and classification model as well as tentatively make a mobile application. This will expand on the proof of concept, IdentiDoc, made by the current Senior Design II team. Both of the requirements below are modified versions of those in the original Identidoc specification.

4.1 Source Code and Availability

4.1.1 Description

The customer will be given access to the code repositories at the conclusion of the project. The most recent version of the system will be live until Septemeber 2021

4.1.2 Source

Identidoc Team, Optical Profilers

4.1.3 Constraints

N/A

4.1.4 Standards

N/A

4.1.5 Priority

High

4.2 Documentation

4.2.1 Description

All documentation including the project charter, System Requirements Specification, Architectural Design Specification, and Detailed Design Specification will be submitted to Dr. Shawn Gieser, and tentatively State Farm, as PDFs.

4.2.2 Source

Optical Profilers

4.2.3 Constraints

N/A

4.2.4 Standards

N/A

4.2.5 Priority

High

5 Performance Requirements

The following is directly from the IdentiDoc team, "The performance of any system is dependent of several metrics like the user bandwidth, server processing speed, robustness of the algorithm. Provided that these metrics are optimal for system performance,our system is expected is perform best in terms of space and time complexity. These requirements are arbitrarily set by the identiDoc Team and will be used as a guideline to build most efficient system based on provided constraints. These metrics are subject to change based on the constraints. If the system is capable of delivering better performance within the same constraints, we will make sure the system does. The performance of our system shall be quantified in terms of the performance of several computational units of our system". Our goal as the document classification team is too further improve upon those metrics.

5.1 Classification Accuracy

5.1.1 Description

The system should have at least a 90% classification accuracy when classifying known document types.

5.1.2 Source

identiDoc Development Team

5.1.3 Constraints

The quality of training documents will skew the accuracy of classification.

5.1.4 Standards

N/A

5.1.5 Priority

High

5.2 Database Query Time

5.2.1 Description

A response should return to the user within two seconds while querying the database for previous results.

5.2.2 Source

identiDoc Development Team

5.2.3 Constraints

The processing capacity of server and the database structure considered shall have an effect on this performance metric.

5.2.4 Standards

N/A

5.2.5 Priority

Low

5.3 Document Classification Time

5.3.1 Description

After a document is uploaded, the system should classify it within two seconds.

5.3.2 Source

identiDoc Development Team

5.3.3 Constraints

The processing capacity of server and the model used for classification shall be taken into consideration.

5.3.4 Standards

N/A

5.3.5 Priority

Low

6 Safety Requirements

The team, Optical Profilers will work on a system that will solely work-on/develop a software system. As required by the CSE Senior Design course and in case the team has to work on any laboratory equipment, electrical, or robotic manipulator, the team will follow the following procedures.

6.1 Laboratory equipment lockout/tagout (LOTO) procedures

6.1.1 Description

Any fabrication equipment provided used in the development of the project shall be used in accordance with OSHA standard LOTO procedures. Locks and tags are installed on all equipment items that present use hazards, and ONLY the course instructor or designated teaching assistants may remove a lock. All locks will be immediately replaced once the equipment is no longer in use.

6.1.2 Source

CSE Senior Design laboratory policy

6.1.3 Constraints

Equipment usage, due to lock removal policies, will be limited to availability of the course instructor and designed teaching assistants.

6.1.4 Standards

Occupational Safety and Health Standards 1910.147 - The control of hazardous energy (lockout/tagout).

6.1.5 Priority

Critical

6.2 National Electric Code (NEC) wiring compliance

6.2.1 Description

Any electrical wiring must be completed in compliance with all requirements specified in the National Electric Code. This includes wire runs, insulation, grounding, enclosures, over-current protection, and all other specifications.

6.2.2 Source

CSE Senior Design laboratory policy

6.2.3 Constraints

High voltage power sources, as defined in NFPA 70, will be avoided as much as possible in order to minimize potential hazards.

6.2.4 Standards

NFPA 70

6.2.5 Priority

Critical

6.3 RIA robotic manipulator safety standards

6.3.1 Description

Robotic manipulators, if used, will either housed in a compliant lockout cell with all required safety interlocks, or certified as a "collaborative" unit from the manufacturer.

6.3.2 Source

CSE Senior Design laboratory policy

6.3.3 Constraints

Collaborative robotic manipulators will be preferred over non-collaborative units in order to minimize potential hazards. Sourcing and use of any required safety interlock mechanisms will be the responsibility of the engineering team.

6.3.4 Standards

ANSI/RIA R15.06-2012 American National Standard for Industrial Robots and Robot Systems, RIA TR15.606-2016 Collaborative Robots

6.3.5 Priority

Critical

7 Maintenance & Support Requirements

The team, Optical Profilers is working on prebuilt software. Based on the information provided by the team who prebuilt the application, this software already requires good maintenance on the signature recognition and a better software UI. As of now, these prerequisite maintenance will be performed by all the members of the Optical Profilers. After the delivery of the product to the sponsors, users will notice the problems they face with the software and notify the sponsors and sponsors will notify the team Optical Profilers, who will further maintain the product. Optical Profilers will mainly use software tools for maintenance and the team has a version control platform that contains all the necessary information/manuals/guides that are needed to run/maintain the product and in case a new member is added to Optical Profilers, they can understand the DevOps using the provided version control platform.

7.1 Requirement Name

7.1.1 Description

Until the end of the Senior Design 2 Course, the team Optical Profilers will definitely provide long-term support for the Document Classifier system. Based on an abstract idea, if the user experiences any bugs or crashes in the software, the user can send crash reports directly to the developers or this team via the Document Classifier application, who will work to fix/maintain the system.

7.1.2 Source

Optical Profilers Dev Team

7.1.3 Constraints

Further maintenance and support cannot be performed in case the team meets the following constraints:

- Sponsors do not accept or use the system provided by the team.
- Current member derailing or leaving the project.
- Improper communication medium between users and developers.
- Sponsors not fully co-operating/sponsoring the team.

7.1.4 Standards

ISO 9001

7.1.5 Priority

High

8 Other Requirements

After the completion of the main project, which is being handled by team 1, there will be 1 more requirement handled by team two which is innovation which will deal with improving the final product in any way possible due to time constraint or another reason..

8.1 Innovation

8.1.1 Description

There needs to be improvement to the UI which will make the application look much more professional and fun to use. The API needs to be more accurate and able to recognize documents more efficiently and scan more types of documents. There needs to be some sort of innovation to the portability of the software meaning a mobile application.

8.1.2 Source

The source of the requirement is StateFarm or will be UTA.

8.1.3 Constraints

Time constraint summer 2021.

8.1.4 Standards

N/A: no standards set for this requirement.

8.1.5 Priority

Moderate priority: This is important to have because it is the goal of team two to make improvements to this project.

9 FUTURE ITEMS

This section provides the list of future requirements, which will improve and upgrade the implementation of the system. Future requirements will be implemented according to the customers demand. Implementation of these requirements depends on time availability of developers.

9.1 RECOGNITION OF SIGNATURE

9.1.1 DESCRIPTION

The system will be able to read the file, detect and verify the presence of a signature.

9.1.2 SOURCE

Optical Profilers Dev Team

9.1.3 CONSTRAINTS

N/A

9.1.4 STANDARDS

N/A

9.1.5 PRIORITY

Future. Depends on time availability

9.2 MATCHING THE SIGNATURE

9.2.1 DESCRIPTION

This requirement will compare signature with others users.

9.2.2 SOURCE

Optical Profilers Dev Team

9.2.3 CONSTRAINTS

N/A

9.2.4 STANDARDS

N/A

9.2.5 PRIORITY

Future. Depends on time availability

REFERENCES