Software design document for image-based web scraping software

**Prepared by:**

* **Ahmed Mohammed Abd El-Ghani**
* **Ahmed Mahmoud Agamy**
* **Hussein Muhammad El-Sayed**
* **Islam Ashraf Ismail**

**Guided By**

**Dr. Prof. Amr E. Mohamed**

Contents

[1. Introduction 2](#_Toc517291918)

[1.1. Purpose 2](#_Toc517291919)

[1.2. Scope 2](#_Toc517291920)

[1.3. References 2](#_Toc517291921)

[1.4. Overview 2](#_Toc517291922)

[2. Design consideration 2](#_Toc517291923)

[2.1. Constrains 2](#_Toc517291924)

[2.2. System environment 2](#_Toc517291925)

[2.3. Design methodology 3](#_Toc517291926)

[3. archtecture 3](#_Toc517291927)

[3.1. System design 3](#_Toc517291928)

[4. data design 4](#_Toc517291929)

[4.1. Context diagram 4](#_Toc517291930)

[4.2. Dataflow diagram 4](#_Toc517291931)

[4.2.1. Level 0 diagram 4](#_Toc517291932)

[4.2.2. Level 1 diagram 5](#_Toc517291933)

[5. component design 6](#_Toc517291934)

[5.1. Scrapping by text 6](#_Toc517291935)

[5.2. Scrapping by image 8](#_Toc517291936)

[6. software interface 10](#_Toc517291937)

[6.1. Object and actions 10](#_Toc517291938)

[6.2. User interface 10](#_Toc517291939)

[6.3. Screen image 10](#_Toc517291940)

[7. class diagram 11](#_Toc517291941)

# Introduction

## Purpose

The purpose of this document is to present a detailed description of the designs of image-based web scraping software, Firstly, this document is intended for the Team, to use the designs as guidelines to implement the project. Lastly, this document could be used for designers who try to upgrade or modify the present design of the system.

## Scope

This document gives a detailed description of the software architecture of the web-scrapping system. It specifies the structure and design of some of the modules discussed in the SRS. It also displays some of the use cases that had transformed into sequential and activity diagrams. The class diagrams show how the programming team would implement the specific module.

## References

* *IEEE Standard 1016-1998, IEEE Recommended Practice for Software Requirements Specifications, IEEE Computer Society, 1998.*

## Overview

This document is written according to the standards for Software Design Documentation explained in “IEEE Recommended Practice for Software Design Documentation”.

The next chapter, the Design consideration, this section gives an overview of some of the constrains and assumption that has been taken to consideration in the design.

The third to fifth chapter, contain discussions of the designs for the project with diagrams.

The sixth chapter, the Software interface design, contain the UI design samples from the system.

The seventh chapter, the Class diagram, this section contains the class diagram

# Design consideration

## Constrains

The system is designed to be built using python 3 with already made TensorFlow API with object recognition model and selenium web-driver for web-scrapping are used to generate the report/s.

## System environment

The System is designed to work on Microsoft windows and linex operating system

## Design methodology

The system is designed with flexibility for further development and/or modification. The system is divided into manageable processes that are grouped to sub-modules and modules that are built with abstraction.

# archtecture

## System design

A close up of a mans face

Description generated with high confidenceMain and subroutine call and return architecture style.

Figure 1. system design

# data design

## Context diagram

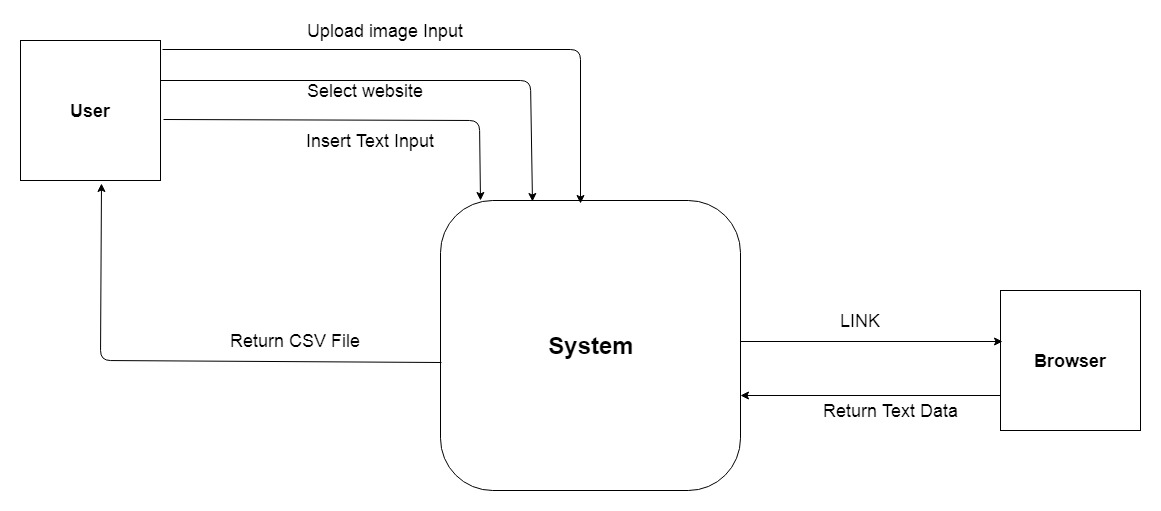


Figure 2. context diagram

## Dataflow diagram

### Level 0 diagram

A screenshot of a social media post

Description generated with very high confidence

Figure 3. level 0 diagram

### Level 1 diagram

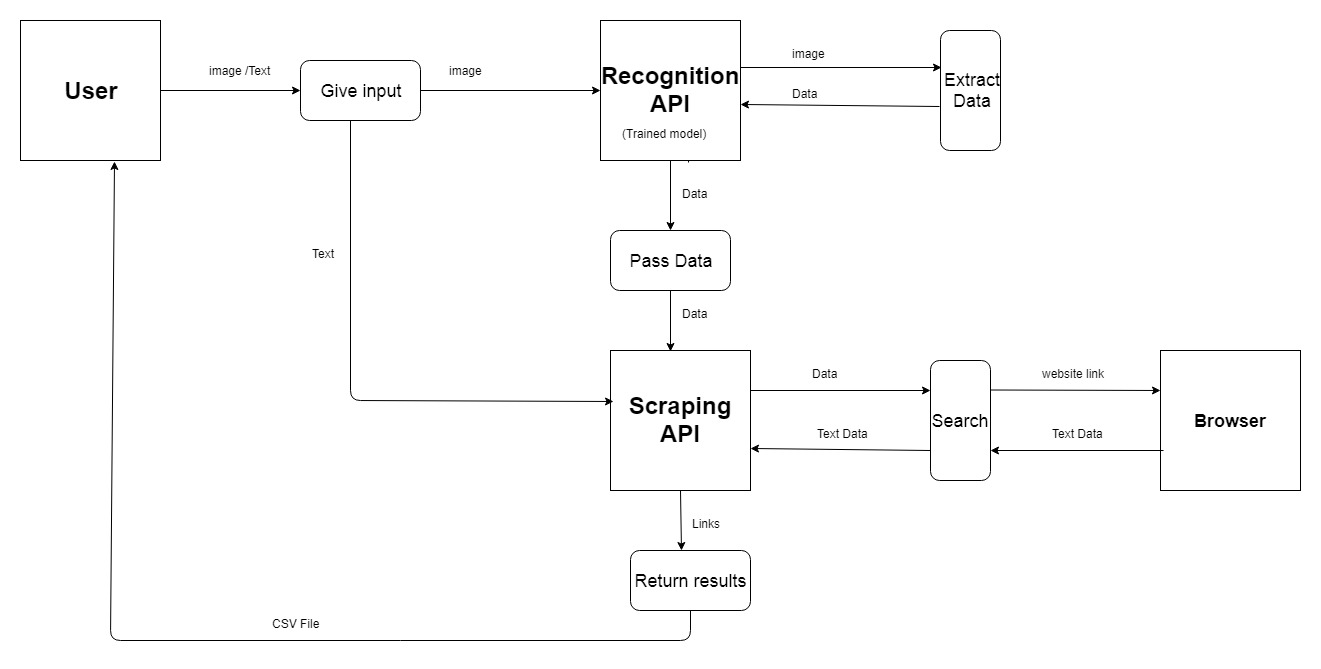


Figure 4. level 1 diagram

# component design

## Scrapping by text

A close up of text on a white surface

Description generated with high confidence

Figure 5. Activity scraping by text

A screenshot of a cell phone

Description generated with very high confidence

Figure 6. Sequence diagram 1

## Scrapping by image

A close up of a sign

Description generated with high confidence

Figure 7. Activity scraping by image

A close up of a map

Description generated with very high confidence

Figure 8. Sequence diagram 2

# software interface

## Object and actions

Selecting the scrapping type either by text entered in the “scrap by text input filed” or by image selected using the button open, when selecting scrap by image the selected image appears in the space below, “scrap by image input” (read only field) the path is displayed.

Then select the desired website through the check box (at least one option must be selected).

Start scrapping button, the software starts and generate the report files which is displayed in output list box.

## User interface

Used GUI components are buttons, text boxes, check boxes, labels, list.

## Screen image

A screenshot of a computer screen

Description generated with very high confidence

Figure 9. screen image 1

A screenshot of a computer

Description generated with very high confidence

Figure 10. screen image 2

# class diagram

A screenshot of a cell phone

Description generated with very high confidence

figure 11. class diagram

**souq** – scrap search result from souq website searching by a keyword

**ali** – scrap search result from Ali-Express website searching by a keyword

**new\_egg** – scrap search result from newEGG website searching by a keyword

**load\_image\_into\_numpy\_array** – transform an image to numpy array to process on it

**run\_interface\_for\_single\_image** – run detection api for the selected numpy-array image

**detect\_from\_image** – call suitable methods to apply the api and get updated image and string list of labels on image