Software Requirements Specification

For

KLWINE

Version 0.5

* Prepared by:

-Aly Mohammed Reda

June , 2018

Faculty of Engineering, Helwan University

SRS Table of Contents

1.[**Preface 7**](#_Toc517134687)

[**Document Purpose 7**](#_Toc517134688)

[**Target Users 7**](#_Toc517134689)

[**2. Introduction 8**](#_Toc517134690)

[**2.1 Purpose 8**](#_Toc517134691)

[**2.2 Problem definition 8**](#_Toc517134692)

[**2.3 Scope 8**](#_Toc517134693)

[**2.4 Overview 9**](#_Toc517134694)

[**3. Glossary 10**](#_Toc517134695)

[**3.1 Acronyms, definitions and abbreviations. 10**](#_Toc517134696)

[**4. System Users 10**](#_Toc517134697)

[**4.1 End Users 10**](#_Toc517134698)

[**4.2 Users objectives 11**](#_Toc517134699)

[**5. User Requirements 11**](#_Toc517134700)

[**5.1 Functional Requirements 11**](#_Toc517134701)

[**5.1.1 A Developer can train a Semantic Classifier module: 11**](#_Toc517134702)

[**5.1.4 Algorithms to implement classifier 12**](#_Toc517134708)

[**5.2 Non-functional Requirements 12**](#_Toc517134709)

[**5.3 Constraints 13**](#_Toc517134710)

[**6. Interface requirements 14**](#_Toc517134711)

[**6.1User interfaces 14**](#_Toc517134712)

[**6.2-Hardware Interfaces 15**](#_Toc517134713)

[**6.3-Software Interfaces 15**](#_Toc517134714)

[**6.4-Communications Interfaces 17**](#_Toc517134715)

[**7. System requirements and specifications 18**](#_Toc517134716)

[**8. System Modeling 20**](#_Toc517134717)

[**9. Time Plan 24**](#_Toc517134718)

[**10. References 25**](#_Toc517134719)

**S**RS Table of figures:

[Figure 1: User interface (Before Classifying) 15](#_Toc517137648)

[Figure 2: User Interface (After classifying) 15](#_Toc517137649)

[Figure 3:system archetecture 16](file:///C:\Users\Yomna\Downloads\Merged_document.docx#_Toc517137650)

[Figure 5: use case Diagram 19](#_Toc517137651)

[Figure 6: context Diagram 21](#_Toc517137654)

[Figure 7: Activity Diagram 23](#_Toc517137656)

[Figure 8: Class Diagram 24](#_Toc517137657)

[Figure 9: WBS 25](#_Toc517137658)

**Figure10: gantt chart**

# 

# 1- Preface

## 1.1-Document Purpose

The purpose of this SRS is to clear the concept of Klwines Scraper how it works and how it can be used. It will illustrate the purpose and complete declaration for the development of system. It will also explain system constraints, interface and interactions with other external system.

## Target Users

-This document is composed by system engineers based on the requirements gathered from system owner Project Manager.

-This document is intended to be approved by Dr / Amr El-Sayed and his assistants.

## Revision History

|  |  |  |
| --- | --- | --- |
| Version | Description | Date |
| 0.1 | Initial | 1-6-2018 |
| 0.2 | Update structure | 8-6-2018 |
| 0.3 | Update structure | 10-6-2018 |
| 0.4 | Update structure | 11-6-2018 |
| 0.5 | Update structure | 13-6-2018 |

# 1. Introduction

## 1.1 Purpose

The purpose of this SRS is to clear the concept of Klwines Scraper how it works and how it can be used. It will illustrate the purpose and complete declaration for the development of system. It will also explain system constraints, interface and interactions with other external system.

## 1.2Problem definition

## In commerce website need to scrape data of klwines website into excel file with table contains product information and the user of the system should choose the type of product want to scrape from Klwines website data of the products (scraped table ) (Date , SKU , Vintage , Item , Name , Item , URL , List Price , Quantity On Hand , Allocation) after scraping user should can send email with file Excel or Excel and Json and should send email with attaching Excel or Excel and Json and the body of the email should contain five status of the comparison between the two last scraping process and user can scheduling scrape and email spirit type .

## By other word to scrap data from the website (Klwine) , store it in Excel or Excel and Jason formate And if the user want to send email with the data system compare the oldest data if the oldest file exist and if any change exist between the last two files update it and send the status by E-mail with last attached Excel file or last attached Excel and json file and if the user want can scrape the spirit type data and send email by Scheduling every period of time .

## 1.3 Scope

* This project aims to detect any change occur to the website (Klwine),and send the new scraped file periodically to the client in E-mail form attached by the Excel file & Json file contain the new scraped data .
* The main functionality isScrape website table manually and possible to store it on Excel and Json file and send email with attaching file and status of file comparison and automatically

( Scheduling ) Scrape website table of spirit type and send email every period of time

Figure 1: System Functions hierarchy

## 1.4 Overview

* The document follows the **IEEE** standards.
* **section 1:** describes the purpose, scope, definitions, acronyms and abbreviations of various terms used in the document
* **section 2:** describes system architecture, algorithms, Baseline used in this project.
* **section 3:** presents some helping information and diagrams that will facilitate the overall understanding.

## 1.5 Similar systems

###### **The following table will show some of feature of two similar systems which make the main job of our system Klwines Scraper .**

|  |  |  |  |
| --- | --- | --- | --- |
| **Features**  **Systems** | **Klwines Scraper** | **Linkedin scrapper** | **Twitter scraper** |
| **Extract Tables** | Yes | Yes | No |
| **Categories** | Yes | No | No |
| **Keyword Search** | No | Yes | Yes |
| **Save History** | Yes | Yes | No |
| **Internet Failure Detector** | Yes | Yes | No |
| **Gmail Login** | Yes | No | No |
| **Website Failure Detector** | Yes | No | No |
| **Export Data** | Yes | Yes | No |
| **Send E-mail with Attached** | Yes | No | Yes |
| **Scheduling** | Yes | No | No |

# 

# 1.6Glossary

## 1.6.1 Acronyms, definitions and abbreviations.

|  |  |
| --- | --- |
| *syncopation* | *Definitions* |
| IDE | *Integrated Drive Electronics* |
| Web Scrapping | a technology solution to extract data from web sites, in a quick, efficient and automated manner, offering data in a more structured and easier to use format, either for B2B or for B2C processes. |
| WAN | wide area network. |
| LAN | local area network |
| SATA | Serial AT Attachment. |
| SVGA | Super VGA cards |
| RAM | *Random-access memory* |
| IEEE | Institute of Electrical and Electronics Engineers |
| B2B | business to business |
| B2C | process to customer. |
| RAID | originally redundant array of inexpensive disks, now commonly redundant array of independent disks |

# 1.7 Alternatives

# 1.7.1 Scrape operation

# Using Scrapy Framework

# Using selenium

# Using selenium and Beautiful soap

# Using request and Beautiful soap (most appropriate)

**Request and Beautiful Library**

Using http request send to server and receive the html file manipulate it with Beautiful soap

**Advantage**

Faster than the other alternatives and because the website doesn't have any complex JavaScript so no need for selenium and Scrapy

**Disadvantage**

With the website have JavaScript prevent the server to send html file without JavaScript and prevent http request so request is most suitable for Klwines website not necessarily for other websites

# 1.7.1 Scrape operation

# Python script

# Python GUI Tkinter (used)

# Script in Flask Frame work to run in cloud

**Python GUI Tkinter**

Make a GUI to run the application in a personal computer having python > 3 and tkinter 3.6

**Advantage**

More easy to user to use GUI that write input into command line directly

**Disadvantage**

The performance maybe not the best python Flask script more faster because it is server based but in this situation the client doesn't care about the performance

The client need easy used application

# 2. System Users:

## 2.1 stake holders:

## 2.2 Users objectives

The objective of this project is to scrap data from website (klwine) ,save it and compare it with the old data and if any change occure update it .

* **system engineer**

-Gain Experience in software engineering and development.

##### **System tester**

###### -Gain Experience in testing of software

# 3. User Requirements:

## 3.1 Functional Requirements:

## 1- Scrapping tables

## 2- Store the data in excel sheet

## 3- Convert excel sheet into Json

## 4- Compare latest two Excel

## 5- Send E-mail with comparison status and attached files

## 6- Scheduling

## 3.2 Non-functional Requirements

* **Usability:**
* The Developer needs to know how to use the system functionalities in maximum time of 5 min.
* the UI must be simple, clear, friendly and easy to use.
* **Reliability:**
* The system must respond to failed internet connection immediately and return a warning message.
* The system must respond to failed website connection immediately and return a warning message.
* The system must respond to failed login to Gmail immediately and return a warning message.
* The system must respond to wrong inputs immediately and return a warning message.
* The system must be thoroughly tested using automated testing tools.
* the software will be able to do its job with an accuracy without failure or crash in.
* the software will be able to do logs to display of details of system operation
* the software will be able to do its job with an accuracy without failure or crash in.
* **Availability:**

-the system available all the time if internet connection exist and website worked and right login Gmail information.

* **Portability:**

-The system should be easily used in many different software environments with minimal time of modification: (1 day/person).

*.*

## 3.3 Constraints

In order to use Klwines Scraper, we provide some constraints that must be taken into account .these constraints are:

* **Hardware constraints:**

a. Needs hardware device to use the system.

b. Needs hardware memory usage.

* **Software Constraints:**a**.** used some specified Libraries :
* Python 3.6
* request
* beautifulsoup
* pickle
* tkinter
* socket
* glob
* os
* codecs
* pandas
* openpyxl
* datetime
* smtplib
* schedule

b. The Internet connection is also a constraint for the system. Since the system fetches data from the database of website over the Internet, it is crucial that there is an Internet connection for the system and user should have Gmail account to use system as sender.

**4. System Architecture**

* The used architecture is Pipe and Filter architecture.
* The reasons for using Pipe and Filter:

● Simple composition

● Reuse

● Prototype

● Easy growth & evolution

● Concurrency & parallelism

* disadvantage of using Pipe and Filter:

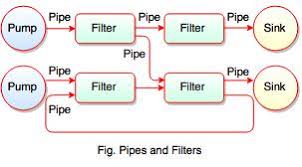
● Poor performance

● Not appropriate for interaction

● Low fault tolerance threshold

● Data transformation

● Increases complexity & computation



# 5. Interface requirements

## 1-User interfaces

## Manual Tab

## C:\Users\Aly\Desktop\web scraping\python win\GuI\gui\Git Klwines v 4.0\klwines V 5\klwines\version 9\klwines\New folder\New folder\klwines\klwines 9\klwines\Images\GUI\1 Manual.png

## Figure 1

## User can choose catergories and the defualt categorgy Distilled Spirits after choosing can scrape or connect or send and can select json and select keep and can add Gmail , Password and Send to information

## If scrape system scrape Excel file and if select json system will scrape excel file and convert it to json

## If connect system will check the Gmail login status to inform user the problem in login or user should enable third party to connect gmail

## if Send system check if the Excel file avialable in the system if not the system ask user to scrape first if yes the system check if there are old file and compare between the two Excel file and check if the user select json if yes excel file convert to json and the two files attached, email send with coparison status and attached files if json not selected excel file send after comparison and if there are no old file to comparse the system send email this is the first time more details will descriped figure 5

## if keep the system will remember Gmail user name and password

## if keep 2 the system will remember Send To

## if json select system will convert excel file into json if exist and if not ask user to scrape and after that convert it into json

## Schedule Tab

## C:\Users\Aly\Desktop\web scraping\python win\GuI\gui\Git Klwines v 4.0\klwines V 5\klwines\version 9\klwines\New folder\New folder\klwines\klwines 9\klwines\Images\GUI\2 Schedule.png

## Figure 2

## By defacult Schedule Tab Scrape and Email the Distilled Spirits into Excel File and attached it into an email with status and all of that every period of time take as inport from user

## Logs Tab

## C:\Users\Aly\Desktop\web scraping\python win\GuI\gui\Git Klwines v 4.0\klwines V 5\klwines\version 9\klwines\New folder\New folder\klwines\klwines 9\klwines\Images\GUI\3 Logs.png

## Figure 3

## This tab to show system logs if the user click button if the system connect to internet or login into gmail or connect website server and logs should contains more information about the history of system by user

## About Klwines Tab

## C:\Users\Aly\Desktop\web scraping\python win\GuI\gui\Git Klwines v 4.0\klwines V 5\klwines\version 9\klwines\New folder\New folder\klwines\klwines 9\klwines\Images\GUI\4 About.png

## Figure 4

## Email Send

## C:\Users\Aly\Desktop\web scraping\python win\GuI\gui\Git Klwines v 4.0\klwines V 5\klwines\version 9\klwines\New folder\New folder\klwines\klwines 9\klwines\Images\GUI\Emails.png

## Figure 5

## Email should contains subject with the name of category and contains the count of record for the category send and status and Email have five status of comparison between the latest Excel files and two status of Email send attached email send with Excel as a default and email send with Excel and Json according to user select and that mean 10 status for every category and since categories are 10 so all possible email send is 100

## 10 status for every category

### 1- Status: No Comparison First Time Scraping

### 2- Status: Same Data

### 3- Status: Data Updated

### 4- Status: Data Increment

### 5- Status: Data Decrement

### And the five status can be with Excel file only or Excel and Json

## 2-HardwareInterfaces:

Since the system must run over the internet, all the hardware shall require to connect internet will be hardware interface for the system.

As for e.g. Modem, WAN – LAN, Ethernet Cross-Cable.

For computers used:

• 100 MB of hard disk space & higher.

• Windows compatible mouse.

• 256 MB RAM & higher

• SVGA card with mono or color monitor.

• CD-ROM Drive for installation.

• Intel p.4 1.5 GH or above

• USB Flash drives

• External hard drives (USB – eSATA enclosures etc.)

• IDE, SATA, RAID controllers and attached devices

## 3-Software Interfaces:

## OS: Windows XP & higher.

## Language : python > 3

## IDE: Python shell or Pycharm

4-Communications Interfaces:

“The communication between the different parts of the system is important since they depend on each other”

• The system shall use the http request to ask server of klwines website to get html data and manipulate in the system and the communication is over the internet.

**4-software requirements and specifications**

## 5.1-system functional requirements

## 1- Scraping

## This function enable the user to extract the wanted table data from the website (Klwines), after check the internet connection and the website connection.

## 2- Store Data into Excel sheet

## this function used to save the scraped data in Excel sheet

## 3- Compare (Latest Two Excel File)

## this function used to compare the latest two excel file if the excel old excel file exist.

## 4- Convert (Json)

## This function used to convert the Excel file to Json file if the user want to send email with Json attached or if user just want to convert scraped file into Json.

## 5- Send E-mail:

## This function used to send an E-mail to the email user added attaching latest excel file or the latest excel file and the latest Json file and Json is just convert from Excel file.

## 5- Scheduling:

## This function to scheduling scrape and send Email with attached Distilled Spirits category Excel file every period time (minutes) added by user

## 5.2- use case diagrams

## C:\Users\Aly\Desktop\web scraping\python win\GuI\gui\Git Klwines v 4.0\klwines V 5\klwines\version 9\klwines\New folder\New folder\klwines\klwines 9\klwines\Images\Digrams\usecase.png

Figure1: use case Diagram

## 5.3-detailed description

# 5. System Models and Diagrams

**5.2-context Diagram**

Figure2: Context Diagram

**5.3-Class Diagram**

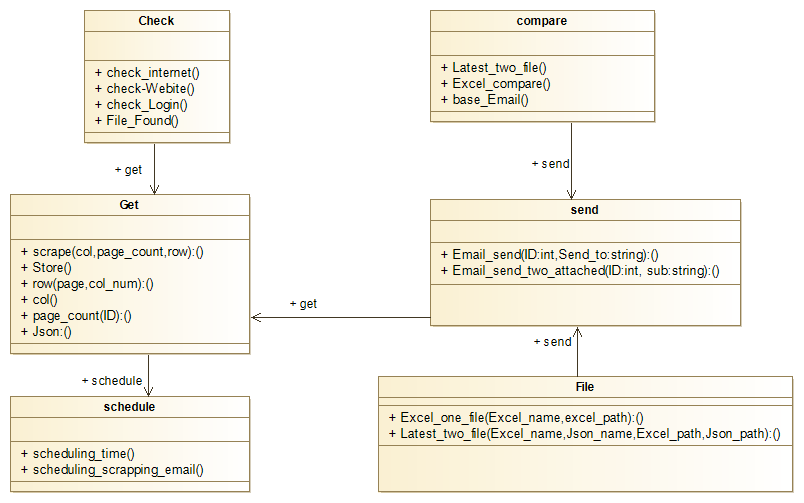
****

Figure3: class Diagram

**5.4-Sequence Diagram**

Figure4: Sequence Diagram

**5.5-Activity Diagram**

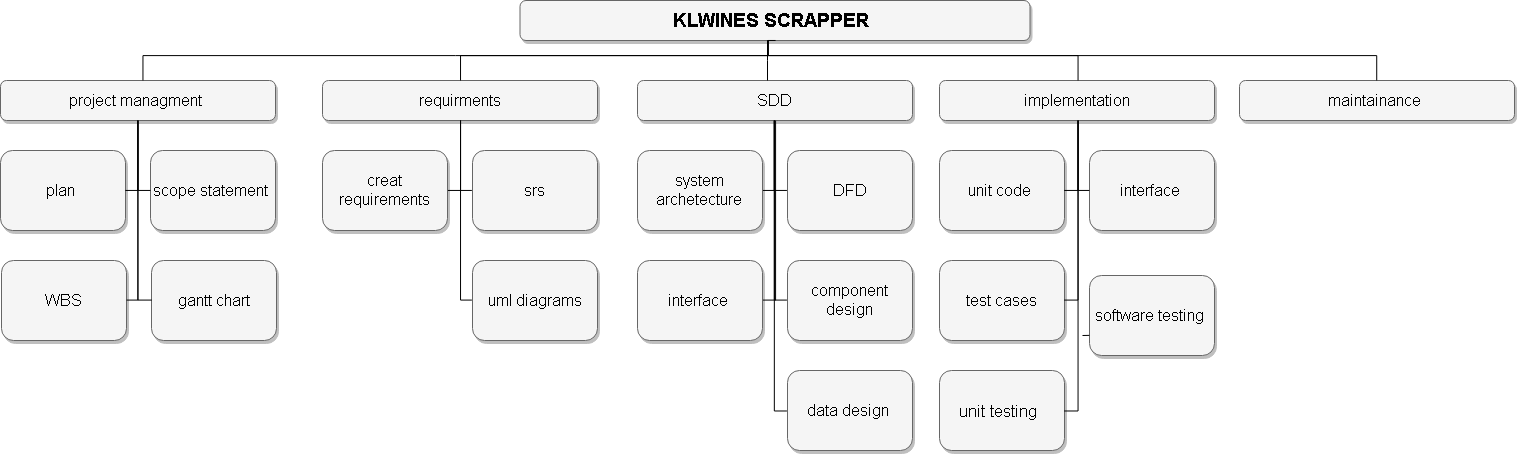
Figure5: Activity Diagram

**5.6-State Diagram**

Figure6: state diagram

**6. Work Plan**

**1-Work Breakdown structure**

Figure7: 

**2-Gantt Chart**

Figure8: Gantt chart

**3-Revision History**

|  |  |
| --- | --- |
| **Version** | **Description** |
| 0.1 | Scraping function |
| 0.2 | Scraping, Send E-mail attached with Excel file. |
| 0.3 | Scraping, Send E-mail attached with Excel file and Json file. |
| 0.4 | Scraping, Send E-mail attached with Excel file and Json file, Send the status of the sent file. |
| 0.5 | Scraping, Send E-mail attached with Excel file and Json file, Send the status of the sent file, Check if file exist in the E-mail if not send a message to the user "Please Scrape First". |
| 0.6 | Scrapping, Send E-mail attached with Excel file and Json file, Send the status of the send file, Check if file exist in the E-mail if not send a message to the user "please scrap first",  Enable the keep check box. |
| 0.7 | Scrapping, Send E-mail attached with Excel file and Json file, Send the status of the sent file, Check if file exist in the E-mail if not send a message to the user "please scrap first", Enable the keep check box. |
| 0.8 | Scraping , Send E-mail attached with Excel file and Json file ,Send the status of the sent file ,Check if file exist in the E-mail if not send a message to the user "please scrap first", Enable the keep check box, Check the website connection . |
| 0.9 | Scrapping , Send E-mail attached with Excel file and Json file ,Send the status of the send file ,Check if file exist in the E-mail if not send a message to the user "please scrap first" , Enable the keep check box, Check the website connection. |
| 1.2 | Logs tab contains all user activity in system |
| 1.4 | About tab contains system information |
| 1.6 | Scheduling tab make scrape of Distilled Spirits category and send email with attached Excel file of the data |

# 7-Future Work

In future versions we intend to:

* Add more features in scheduling tab to schedule more categories.
* Using threads in schedule and abort function
* Scheduling will be cloud based.
* GUI will have more attractive Design.
* System store data and in database and to make historical data.
* Make a login form for a system.
* Make the system faster in performance.

**8. References**

* **https://pandas.pydata.org/pandas-docs/stable/**
* **http://docs.python-requests.org/en/master/**
* **https://docs.python.org/2/library/os.html**
* **https://www.crummy.com/software/BeautifulSoup/bs4/doc**
* **https://docs.python.org/2/library/codecs.html**
* **https://docs.python.org/3/library/smtplib.html**
* **https://docs.python.org/2/library/pickle.html**
* **https://docs.python.org/2/library/datetime.html**
* **https://openpyxl.readthedocs.io/en/2.5/changes.html**
* **https://docs.python.org/3/library/glob.html**
* **https://docs.python.org/3/library/socket.html**
* **https://docs.python.org/2/library/tkinter.html**
* [**https://docs.python.org/2/faq/gui.html**](https://docs.python.org/2/faq/gui.html)
* **https://github.com/dbader/schedule**