Process Definition Document (PDD)

*Process Name: YouTube Scraper*

Table of Contents

[**Introduction**](#_fmc2ik42b62t) **1**

[Purpose of the Document](#_8b0nhjcbe7cw) 1

[Objectives](#_soa72miybokv) 1

[Process Key Contact](#_e5eh7vtp3elw) 1

[Minimum Prerequisites for Automation](#_8uc76jjm25ud) 1

[**As-Is Process Description**](#_mdr6kpc5a5r5) **2**

[Process Overview](#_vrc3lxjwb5na) 2

[Applications used in the Process](#_in5ehl2op8tm) 3

[As-Is Process Map](#_3xrlx7nhtlp7) 3

[**To-Be Process Description**](#_gmvdjkbe065o) **3**

[Detailed Process Map](#_nmn90y3pi1ee) 3

[Robot Type](#_cx5ym07ptgjk) 4

[Business Exceptions Handling](#_7u1z1cuc6dh1) 4

[Known Exceptions](#_jr6jw3koor93) 4

[Unknown Exceptions](#_57bwdp6ycy5h) 4

[System Exceptions Handling](#_1os2bz2dwbrb) 4

[**Other Observations**](#_bsbyy5x0t0oj) **5**

[**Additional sources of process documentation**](#_ms60s1nz1shm) **5**

# Introduction

## Purpose of the Document

The Process Definition Document outlines the business process chosen for automation using UiPath Robotic Process Automation (RPA) technology.

The document describes the sequence of steps performed as part of the business process, the conditions and rules of the process prior to automation and how they are envisioned to work after automating it, partly or entirely. This specifications document serves as a base for developers, providing them with the details required for applying robotic process automation to the selected business process.

## Objectives

The business objectives and benefits expected by the Business Process Owner after automation of the selected business process are:

* The primary purpose of this automation is to allow users to easily obtain transaction-level details about a YouTube Search with applicable filters for further data analysis.
* Secondary purposes include: collecting aggregate-level details about YouTube Search results over time, scraping transaction-level details for a specific topic or timeframe, and identifying trends or patterns within search results based on applicable filters.

## Process Key Contact

The specifications document includes concise and complete requirements of the business process and it is built based on the inputs provided by the process Subject Matter Expert (SME)/ Process Owner.

The Process Owner is expected to review it and provide signoff for accuracy and completion of the steps, context, impact and a set of process exceptions. The details are to be included in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| Role | Name | Contact Details  (email & phone number) | Notes |
| Process Owner | Jesse Peterson | [petersoninquiries@gmail.com](mailto:petersoninquiries@gmail.com)  402-123-3492 |  |
| Business Analyst | Jesse Peterson | [petersoninquiries@gmail.com](mailto:petersoninquiries@gmail.com)  402-123-3492 |  |

## Minimum Prerequisites for Automation

|  |  |
| --- | --- |
| Met (Y/N) | Prerequisites |
| Y | A filled in and completed Process Definition Document |
| Y | Closure of any open process questions |
| Y | Environment set up |
| Y | Test Data to support development and testing |
| Y | User access and creation of user accounts (licences, permissions, restriction to create accounts for robots) |

# As-Is Process Description

## Process Overview

General information about the process selected for RPA prior to automation.

|  |  |  |
| --- | --- | --- |
| # | Item | Description |
| 1 | Process Full Name | YouTube Scraper Automation |
| 2 | Process Area | Web Scraping and Automation |
| 3 | Department | R&D |
| 4 | Process Short Description  (operation, activity, outcome) | This automation uses the ‘Legacy’ version of UiPath and the ‘UiPath.System.Activities ‘ library to prompt the user to enter data into a UI form using the ‘UiPath.Form.Activities’ library, performs a YouTube search and scrapes the results using the ‘UiPath.UIAutomation.Activities’ library, saves the results to a local file using the 'UiPath.Excel.Activities’ library, then sends a summary report through email using the ‘UiPath.Mail.Activities ‘ library. |
| 5 | Role(s) required for performing the process | A single UiPath Studio or StudioX license owned by any business user. |
| 6 | Process schedule and frequency | Run once per day, daily |
| 7 | # of items processed /reference period | Scrapes 1 YouTube video in 15 seconds |
| 8 | Process execution time | 15 seconds per video and 30 seconds to start / complete the automation |
| 9 | Peak period(s) | Peak periods would be during normal 9-5 business houes |
| 10 | Transaction Volume During Peak period | Automation can scrape up to 320 videos per hour. |
| 11 | Total # of FTEs supporting this activity | 1 FTE per bot |
| 12 | Expected increase of volume in the next reference period | Volumes could be increased to scraping 3000 videos per day / per bot |
| 13 | Level of exception rate | Less than 1 / 100 records should have an exception. |
| 14 | Input data | User submitted form data for searching and filtering YouTube |
| 15 | Output data | 1 log file with the users original input and 1 stats report file with the YouTube search results sent via an email with Summary statistics about the automation. |

\*Add more rows to the table to include relevant data for the automation process. No fields should be left empty. Use “n/a” for the items that don't apply to the selected business process.

## Applications used in the Process

The table includes a comprehensive list of all the applications that are used as part of the process to be automated to perform the given steps in the flow.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Application Name & Version | System Language | Thin/Thick Client | Environment/ Access Method | Comments |
| 1 | UiPath Studio | VB | Thick | Local Access |  |
| 2 | Google Chrome | N/A | Thick | Desktop Access |  |
| 3 | YouTube | N/A | Thick | Browser Access |  |
| 4 | Gmail | N/A | Thick | Browser Access |  |
| 5 | Text Editor | N/A | Thick | Local Access |  |

\*Add more rows to the table to include the complete list of applications.

## As-Is Process Map

**High Level As-Is Process Map:** This chapter depicts the As-Is business process at a High Level to enable developers to have a high-level understanding of the current process.

3. Automation Loads YouTube and Applies Users Inputs in Search

2. Automation writes Startup Form input to Log File

1. User fills out Startup Form



5. Automation writes YouTube Results to Log File

6. Log File is Upload to UiPath Orchestrator Queue

4. Automation Scrapes each record in the YouTube Search Results



8. Automation sends email via Gmail with Start Form Log File and YouTube Results Log File

7. Automation pulls Log File Transaction from UiPath Orchestrator Queue

**Detailed Process Map:** This chapter depicts the As-Is business process at a detailed view to enable process owners to document their process

|  |  |  |  |
| --- | --- | --- | --- |
| # | Step Action/Description | Screenshot | Remarks |
| 1 | User fills out Startup Form |  | ‘Count’ field will be used to determine how many records are queried. |
| 2 | Automation writes Startup Form input to Log File |  | Log File is saved into a local ‘ Logs ‘ folder at the root of the users Desktop. |
| 3 | Automation Loads YouTube and Applies Users Inputs in Search |  | Automation uses Google Chrome by default |
| 4 | Automation Scrapes each record in the YouTube Search Results |  |  |
| 5 | Automation writes YouTube Results to Log File |  | YouTube Results are save to a local ‘Status\_Reports’ folder |
| 6 | Log File is Upload to UiPath Orchestrator Queue |  | Results are uploaded as transactions to a queue named ‘YouTube\_Scraper\_Queue’ |
| 7 | Automation pulls Log File Transaction from UiPath Orchestrator Queue |  | Transactions are pulled individually from a UiPath Orchestrator queue. |
| 8 | Automation sends email via Gmail with Start Form Log File and YouTube Results Log File |  | User will be prompted to log into Gmail account |

# To-Be Process Description

## Detailed Process Map

**High Level To-Be Process Map:** This chapter depicts the To-Be automation process at a High Level to enable developers/COE to have a high-level understanding of the to be developed process.

The ideal ‘to-be’ state of this automation is to transform it into an unattended bot run on a schedule using the YouTube API and storing the final results all in Orchestrator in both Queues and Storage Buckets.

.

3. The unattended bot asynchronously calls YouTube

2. Orchestrator executes the search term on a schedule using an unattended bot.

1. Search Terms are created by users and uploaded to UiPath Orchestrator as an Asset



**OR**

6. A notification is sent to the process owner informing them that the automation has completed.



5. Results are moved to a storage bucket everyday for analysis by other applications or services.

3. YouTube API is used to submit request to endpoint and automation parses result-set

4. Parsed results are stored in UiPath Orchestrator Queue







**Detailed Process Map:** This chapter depicts the To-Be automation process at a detailed view to enable developers/COE to see the workflows involved in the RPA solution

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Workflow Name** | **Description** | **Pre-conditions** | **Post-actions** | **Arguments** | **Notes** |
| Assets | User creates file or stores individual keywords and filters in Orchestrator as an Asset | None | Asset is created | Keywords; List of Strings  Filters; List of Strings |  |
| Jobs | A UiPath Orchestrator job is kicked off on a schedule | Assets created in Orchestrator and job configured | Starts YouTube Scraper automation | Keywords: List of Strings  Filters; List of Strings |  |
| YouTube | Asynchronous call to YouTube to load a ‘headless’ version of the app | Job started and assets configured, keywords and filters passed as an argument. | Loads a headless version of YouTube, stores results of query in DataTable | Keywords: List of Strings  Filters; List of Strings |  |
| API | Web request to YouTube REST API to query search results | Job started and assets configured, keywords and filters passed as an argument. | Submits REST requests to YouTube API, , stores results of query in DataTable | Keywords: List of Strings  Filters; List of Strings |  |
| Queues | Resulting DataTable from call to YouTube is uploaded to UiPath Orchestrator Queue as transaction | YouTube Search results are stored in DataTable | Queue Created or Updated with Search Results | Search Results: Data Table  Queue Name: String | Queue can be created manually before running the automation or dynamically by the automation. |
| Storage Buckets | Search Results are moved from Queue to a Storage bucket after 24 hours | Search Results stored in Queue as transactions | Search Results moved to Storage Bucket and deleted from Queue | Queue Name; string  Storage Bucket name; string | Results deleted from Queue after moving to Storage Bucket |
| Gmail | A notification is sent to the user informing them that the automation has completed | Search Results are successfully uploaded to Queue and job scheduled to move them to storage bucket | Email sent to developer | Email address; string  Queue Name; string  Storage Bucket name; string  Email Subject; string  Email Body: List of Strings |  |

## Robot Type

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # | Attended | Unattended | Trigger | Comments |
| 1 | YouTube Scraper Bot | N/A | N/A | N/A |

## Business Exceptions Handling

The Business Process Owner and Business Analysts are expected to document below all the business exceptions identified in the automation process. These can be classified as:

### Known Exceptions

The table below reflects all the business process exceptions encountered during the process evaluation and documentation. These are known exceptions that occurred before. For each of these exceptions, define a corresponding expected action that the robot should complete if it encounters the exception.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BE # | Exception Name | Step | Parameters | Action to be Taken |
| 1 | InvalidDataException | Enter\_Form\_Process.xaml | out\_LogFilePath,  out\_SearchTerms,  out\_Duration, out\_UploadDate, out\_Type,  out\_Features,  out\_SortBy, out\_SearchCount,  out\_EmailAddress | Generate an Error Log message and prompt the user to change the input dataset. |
| 2 | MailException | Send\_Status\_Report | in\_EmailAddress | Prompt the user to retry the automation with the correct email address or restart the automation from the beginning. |
| 3 |  |  |  |  |

### Unknown Exceptions

For all other unanticipated or unknown business (process) exceptions, the robot should:

* For all other unknown business process exceptions, the robot should throw a generic Exception and prompt the user to restart the automation from the beginning.

## System Exceptions Handling

A comprehensive list of all errors, warnings or notifications should be consolidated here with the description and action to be taken, for each, by the robot.

Errors identified in the automation process can be classified as:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SE # | Exception Name | Step | Parameters | Action to be Taken |
| 1 | SystemException | Main.xaml | None | Generate an Error Log message and prompt the user to restart the automation from the beginning |
| 2 | WriterException | Enter\_Form\_Process and Send\_Status\_Report | out\_LogFilePath,  out\_SearchTerms,  out\_Duration, out\_UploadDate, out\_Type,  out\_Features,  out\_SortBy, out\_SearchCount,  out\_EmailAddress | Generate an Error Log message and prompt the user to restart the automation from the beginning |
| 3 | BrowserNotFoundException | Load\_YouTube\_Process | N/A | Generate an Error Log message and prompt the user to restart the automation from the beginning |
| 4 | SelectorOperationException | Scrape\_Results\_Process | out\_Title,  out\_Creator,  out\_Description,  out\_VerifiedStatus,  out\_WatchCount,  v\_TagStrings,  out\_TagList | Generate an Error Log message and prompt the user to restart the automation from the beginning or use the ‘ignore’ feature to skip the current record. |
| 5 | OrchestratorHttpException | Post\_Data\_To\_Orchestrator | IO\_DTResults,  v\_SuccessTransactionCount, |  |

For all the other unanticipated or unknown system exceptions, send an email to **<peteroninquiries@gmail.com>** and attach a screenshot of the error message.

# Other Observations

Include below any other relevant observations you consider needed to be documented here.

* All automations are using their default timeouts periods, so if a specific activity fails to execute or cannot find the appropriate selectors, it will timeout after the default interval.

# Additional sources of process documentation

* To learn more about how CSS selectors work, you can view the following page: [CSS Selectors Reference (w3schools.com)](https://www.w3schools.com/cssref/css_selectors.asp)