IMDb Movie Rating Automation and Data Extraction

This project automates the process of reading movie titles and ratings from an Excel file, searching each title on IMDb, allocating the specified rating, retrieving detailed information (description, directors, writers, stars), and saving all the data into a new Excel file. At the end of the process, the project automatically sends an email with the output Excel file attached.

Contents

1. Introduction

1.1 Purpose

The Process Definition Document outlines the business process chosen for automation. The document describes the sequence of actions performed as part of the business process, the conditions and rules of the process prior to automation (AS IS) as well as the new sequence of actions that the process will follow as a result of preparation for automation (TO BE).

The PDD is a communication document between:

- The RPA Business Analyst and the SME/Process Owner. The goal is to ensure that the RPA Business Analyst has the correct understanding of the process and has represented it accurately.
- The RPA Business Analyst and the Development team (represented by the Solution Architect and RPA Development Lead). The goal is to ensure that the process is documented appropriately and to a sufficient level of detail so that the Solution Architect can then create the solution based on the PDD content.

1.2 Objectives

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

- Reduce processing time.

- Ensure consistent and accurate entry of ratings and extraction of movie data by eliminating human errors in manual data handling.

1.3 Key Contacts

Add here any stakeholders that need to be informed or to approve changes to the process:

Role	Name	Contact Details (email, phone number)	Notes
Mentor	**Tudor Popescu**	**tudor.popescu@tquila-automation.com**	
Student	Andreea-Beatrice Croitoriu	croitoriu.va.andreea@student.utcluj.ro	

1.4 Minimum Pre-requisites for the Automation

a) Filled in Process Definition Document
b) Test Data to support development
c) User access and user accounts creations (licenses, permissions, restrictions to create accounts for robots)
d) Credentials (user ID and password) required to logon to machines and applications

2. AS IS Process Description

Open Excel to Create the Output Excel:

- Initialize and create an Excel file that will be used to store the extracted movie details and ratings.

Search on Google for the IMDb Website:

- Use Google to locate and navigate to the IMDb website where the movies will be searched and rated.

Check if the User is Signed In on IMDb:

- Verify whether the user is already signed in to IMDb. If not, perform the sign-in process using stored credentials.

Open Input Excel to Extract the Movie Name and Grade:

- Access the Input Excel file, read the list of movies along with their respective grades (ratings) for processing.

Search the Movie by Name and Give the Rating:

- Use the movie name from the Input Excel to search for the movie on IMDb.
- If the movie is found, verify the grade (rating) is between 1 and 10.
- If the movie already has a rating, remove the existing rating and assign the new rating. If not, assign the new rating directly.

Extract Movie Details:

- After assigning the rating, extract details about the movie, including its description, directors, writers, and stars.
- Store this information in the Output Excel file along with any business exceptions if the movie was not found or if an invalid rating was provided.

Handle Multiple Movies:

- Repeat the above steps for each movie listed in the Input Excel file until all movies have been processed.

Close IMDb Tab:

- Once all movies are processed, close the IMDb browser tab to ensure no open sessions remain.

Save the Output Excel:

- Save the Output Excel file containing all the processed movie details and ratings.

Send Email with Output Excel:

- Send an email to the designated recipients with the Output Excel file attached, summarizing the processed movie data.

End Process:

- Conclude the process, ensuring all applications are properly closed and resources are released.

2.1 Process Overview

Section contains general information about the process before automation.

Item	Description/Answer
Process Full Name	***IMDb Movie Rating Automation and Data Extraction***
Process Area	n/a
Department	n/a
Short Description (operation, activity, outcome)	***This project automates the process of reading movie titles and ratings from an Excel file, searching each title on IMDb, allocating the specified rating, retrieving detailed information (description, directors, writers, stars), and saving all the data into a new Excel file. At the end of the process, the project automatically sends an email with the output Excel file attached.***
Role(s) required in applications to perform the process	n/a
Process schedule and frequency	n/a
Number of times the process is ran by selected frequency	n/a
Process execution time	***2 min./movie***
Process Restrictions	*The process can be done only in case of a movie or/and grade that exists, with an IMBd account created and an email adress.*
Peak Period (s)	***n/a***
Peak Volume Approximate increase	n/a
Number of persons performing the process	***1***
Expected Volume increase during next periods	*20-30%*
Percentage Un-handled exceptions	***15%***
Input data description	*Input Excel contains movie titles and a grade for each movie*
Output Data description	*Output Excel contains movie titles, and the data extracted for each movie from IMDb website (description, directors, writers, stars)*

2.2 Applications Used

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 actions in the flow.

Application Name	Version	Application Language	Thin/Thick Client	Environment/ Access method	Comments
Excel	*Microsoft Excel 2016 or later (including Office 365)*	*English*	Thick Client	Local Installation or Virtual Desktop Infrastructure	
IMDb	n/a	English	Thin Client (Web-Based)	Direct Browser Access	Requires an account
Outlook	*Microsoft Excel 2016 or later (including Office 365)*	English	Thick Client	Local Installation or Virtual Desktop Infrastructure	Requires an account

![UiPath Reboot Work.™ logo]
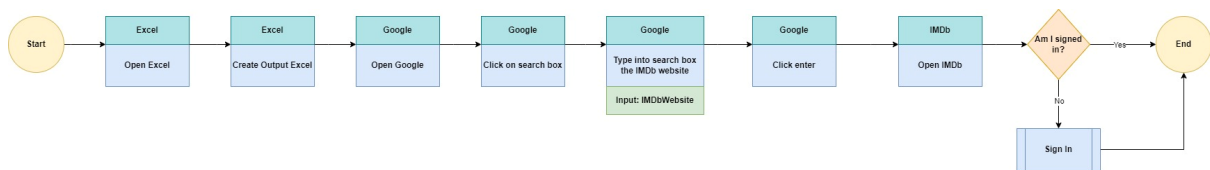
2.3 AS IS Process Map

This section contains various process maps contributing to a better understanding of how the process is performed pre-automation.
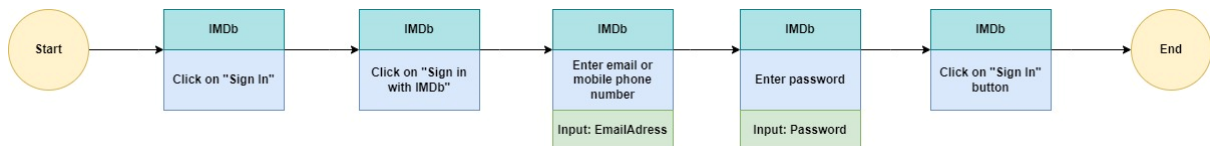
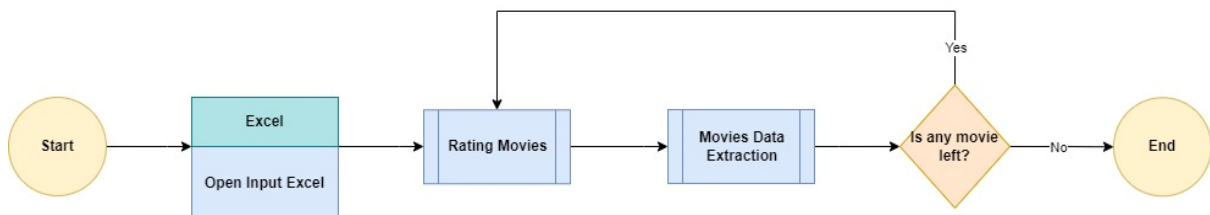### 2.3.1 High Level Process Map

2.3.2 Detailed Level Process Map
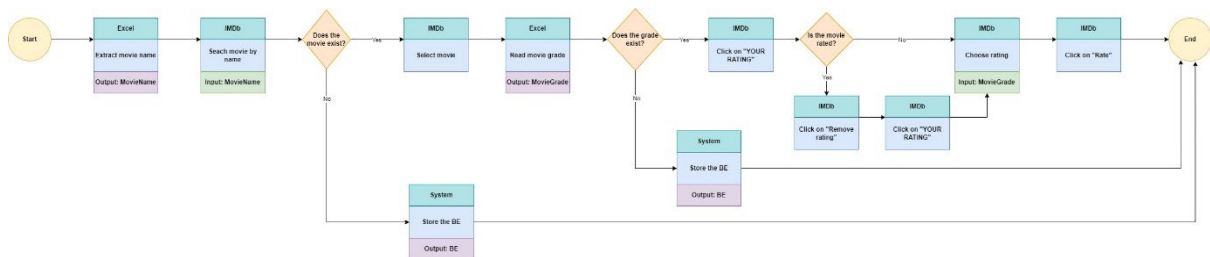
- Init Applications SubProcess:
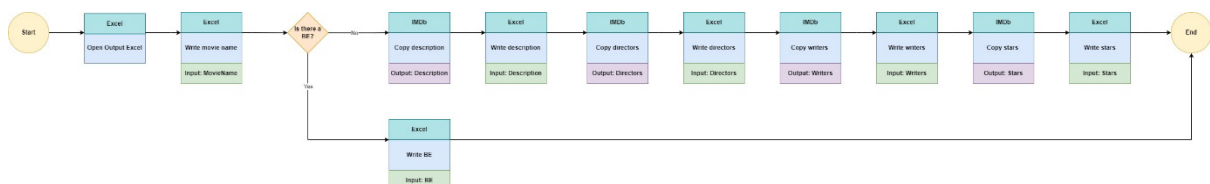

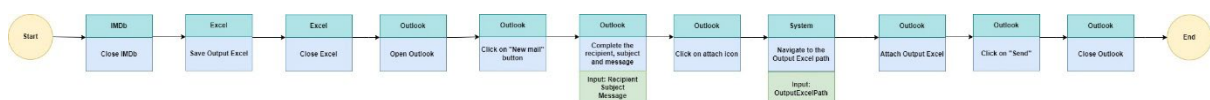 o Sign In SubProcess:

- Process Films SubProcess:

  o Rating Movies SubProcess:

  o Movies Data Extraction SubProcess:

- Close Applications SubProcess:

2.4 Process Statistics

High Level statistics

Processes	Windows	Actions	Mouse clicks	Keys pressed	Text entries	Hotkeys used	Time
{total_applications_count}	{total_windows_count}	{total_actions_count}	{total_clicks}	{total_keys_pressed}	{total_text_entries}	{total_hotkeys}	{process_execution_time}
3	4	46	19	40	12	2	2 min./movie

Detailed statistics

Window name	Mouse clicks	Text entries	Key pressed
{#windows}{name}	{total_clicks}	{total_text_entries}	{total_keys_pressed} {/windows}
Excel Main Window	7	6*nr_input	40
IMDb Browser Tab	6	3*nr_input	40
Outlook Email Window	6	3	40

2.5 Detailed AS IS Process Actions

#Action	Input	Description	Details (Screen/Video Recording Index	Exception Handling	Possible Actions
1	Movie Name and Grade (Excel File)	*Extracts movie name and grade from the input Excel file.*	n/a	Retry reading file; log error if fails.	Check file existence; validate data.
2	IMDb Website	Opens IMDb and checks if the user is signed in.	n/a	Retry sign-in process; trigger alert.	Validate IMDb page load; handle sign-in.
3	Movie Name	Searches for the movie by name on IMDb.	n/a	Handle "movie not found" with a BE.	Verify search results; handle exceptions.
4	Movie Grade	Rates the movie on IMDb according to the provided grade.	n/a	Validate rating input; handle BE for invalid ratings.	Ensure rating is applied; remove previous rating if necessary.
5	IMDb Movie Page	Extracts movie details: description, directors, writers, stars.	n/a	Handle missing details; log incomplete data.	Capture required data; manage missing information.
6	Output Excel Path (Asset)	Writes movie data to the output Excel file.	n/a	Retry writing file; log error if fails.	Check file path; ensure data integrity.
7	Email Content	Sends an email with the output Excel file attached.	n/a	Retry sending email; log and alert if fails.	Verify email content; manage attachments.

2.6 Input Data Description

The following table should contain details regarding the inputs that every action of the process takes.

#Action	Sample	Input Type	Location	Are inputs Natively Digital*?	Are the inputs Structured*?
1	Movie Name and Grade	*Excel File*	Local Disk	Yes	Yes
2	IMDb Website	Web Data	Online	Yes	No
3	Movie Description, Directors, Writers, Stars	Web Data	Online	Yes	No
4	Email Content	Text	Online	Yes	No

* Native Digital: This is data that was originally created digitally e.g. excel, database or application reports etc. The non-native digital inputs are usually scanned images.

* Structured Data: has a predictable format and exists in fixed fields (e.g. an excel cell or a field in a form) and is easily detectable via search algorithms.

3 TO BE Process Description

In this section the proposed improvements to the process, actions to the process will be outlined as well as the actions proposed for automation and the type of robot required. **This will be cross-checked by the Solution Architect.**
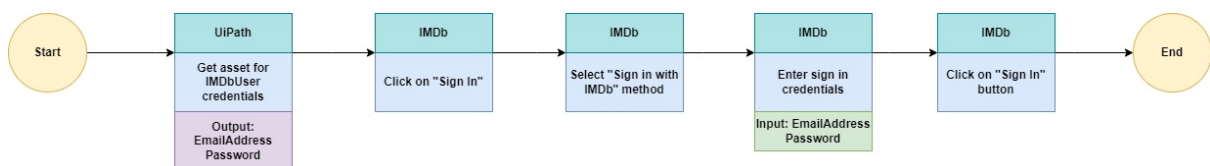

3.1 Detailed TO BE Process Map

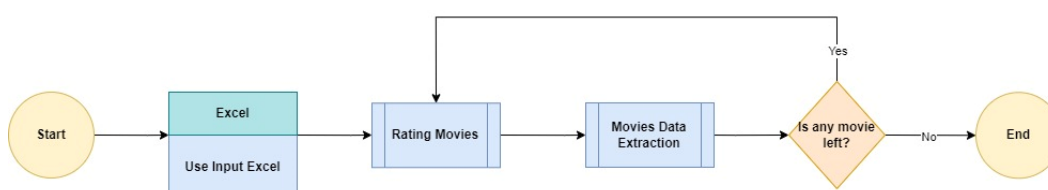A detailed process map of the process as it will look like post-automation will be outlined here.
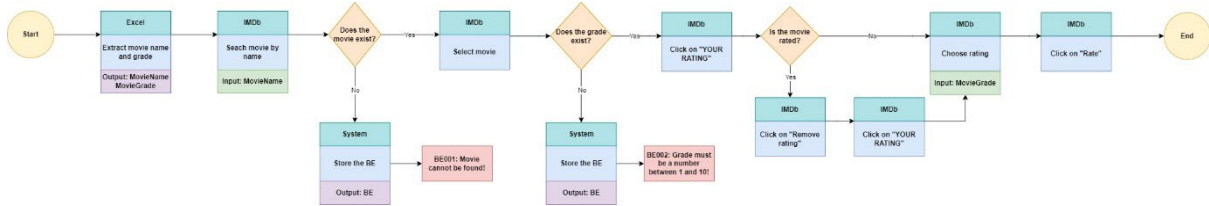
- Init Applications SubProcess:
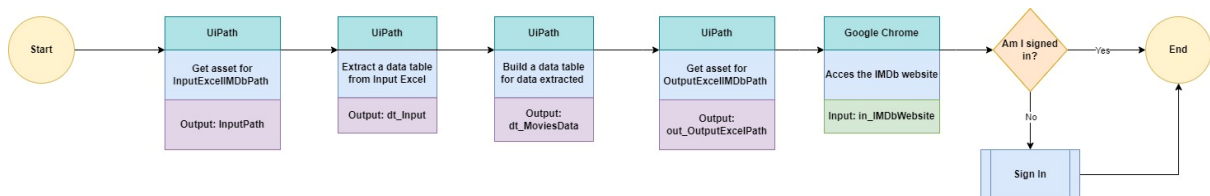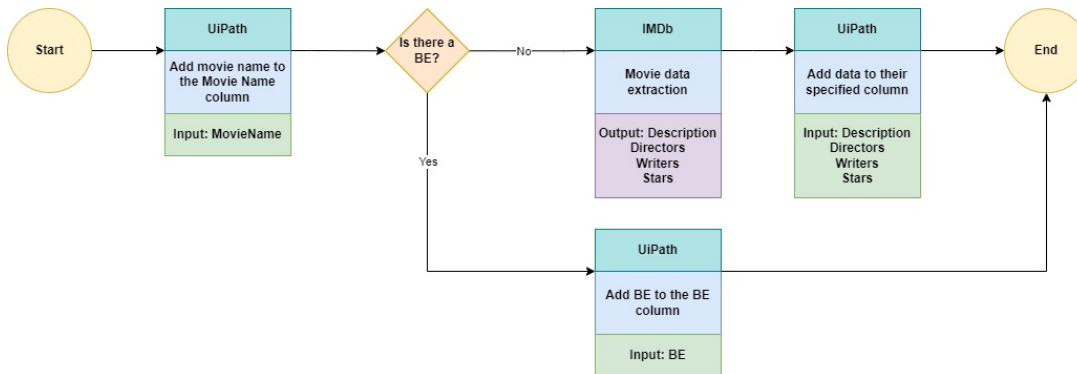

- Sign In SubProcess:
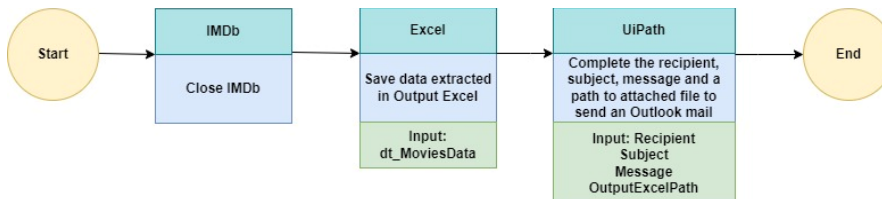

- Process Films SubProcess:

o Rating Movies SubProcess:

o Movies Data Extraction SubProcess:

- Close Applications SubProcess:

3.2 Exception Handling

The Business Process Owner and Business Analysts are expected to document below all the business exceptions identified in the automation process. Exceptions are of 2 types and both need to be addressed:

Known exceptions = previously encountered. A scenario is defined with clear actions and workarounds for each case.

Unknown = New situation that was not encountered before. It cannot be predicted and in case it happens it needs to be flagged and communicated to an authorized person for evaluation.

3.2.1 Known Business Exceptions
Details regarding how the robot should handle the exceptions.

Exception Name	Action	Parameters	Actions to be taken
Movie Not Found	*Search Failure*	*Movie Name*	*If a movie search fails (i.e., movie is not found), log the error and proceed to the next movie in the list.*
Invalid Rating Input	*Rating Error*	*Movie Grade*	*If the rating provided is not between 1 and 10, log the error, notify the user to correct the rating, and proceed to the next movie.*

3.2.2 Unknown Business Exceptions

An umbrella rule that includes a notification needs to be designed for all other exceptions that could happen and cannot be anticipated.

e.g.: for all other cases which do not follow the rules defined an e-mail should be sent to: exceptions@company.com with a screen shot and robot should proceed to next transaction.

4 Other

In this section the proposed improvements to the process, actions to the process will be outlined as well as the actions proposed for automation and the type of robot required. **This will be cross-checked by the Solution Architect.**

4.1 Additional sources of process documentation

If there is additional material created to support the process automation please mention it here, along with the supported documentation provided.

Additional Process Documentation		
Video Recording of the process (Optional)	Acme-System1-Process-WI5-Manual-Walkthrough	Insert any relevant comments
Business Rules Library (Optional)	Insert link to Business rules library	Insert any relevant comments
Other documentation (Optional)	Insert link to any other relevant process documentation (L4, L5 process description, fields mapping files etc.)	Insert any relevant comments
Standard Operating Procedure(s) (Optional)		Insert any relevant comments
High Level Process Map (Optional)		Insert any relevant comments
Detailed level process map (Optional)		Insert any relevant comments
Work Instructions (Optional)		Insert any relevant comments
Input Files (Optional)		Insert any relevant comments
Output Files (Optional)		Insert any relevant comments

*Add more rows to the table to reflect the complete documentation provided to support the RPA process.