



ARAUJO AUTOMATION

123 Automation Road.
São Paulo, Brazil
(55) 123-456-789

Fresh News 2.0

Process Definition Document

Table of contents

Table of contents.....	2
Document History.....	3
Overview.....	4
As-is Process.....	5
To-be process.....	6

Document History

DATE	VERSION	NAME	NOTES
2024-07-15	1.0	PEDRO	Initial version.

Overview

The goal of this PDD is to document the process as it is executed today, and how we envision it to work after automating it.

It describes a process called **Fresh News 2.0**, which scrapes a news website and creates an excel file with the ingested contents.

As-is Process

The main steps:

1. Open the site by following the link

2. Enter a phrase in the search field

3. On the result page

If possible select a news category or section from the

Choose the latest (i.e., newest) news

4. Get the values: title, date, and description.

5. Store in an Excel file:

- title
- date
- description (if available)
- picture filename
- count of search phrases in the title and description
- True or False, depending on whether the title or description contains any amount of money

> Possible formats: \$11.1 | \$111,111.11 | 11 dollars | 11 USD

>

6. Download the news picture and specify the file name in the Excel file

7. Follow steps 4-6 for all news that falls within the required time period

To-be process

The process should run automatically from start to end, provided a initial parameter configuration is performed. It will run on a cloud orchestrator ("Robocloud Control Room").

The process is to be developed using Python and RPA frameworks provided by Robocloud. It should modular and reusable so other news sources can be easily implemented.

It starts by reading the parameters in Work Items inside task "**Capture**". The parameters are:

1. Search term
2. Category
3. Months
4. News source (Optional, defaults to L.A .Times)

It then follows the steps described in the As-is, via the Selenium framework, while performing error validations during each step. When done, it starts the second task, "**Excel**". This task performs some validations and generates the excel file inside the "**output**" folder, located in the orchestrator container for this Task.