

Data Extraction Rules

-

Functions description

DOUCMENT REVISION:

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Name | Description |
| 10/06/2019 | V1.0 | Romain Alexandre | First draft |

Contents

[Introduction 2](#_Toc11055496)

[Notes 2](#_Toc11055497)

[Assets / Logics 3](#_Toc11055498)

[ConfigFile metabot 3](#_Toc11055499)

[FormatUtility metabot 4](#_Toc11055500)

[Post - Process - IQ Bot 5](#_Toc11055501)

[Result Examples 6](#_Toc11055502)

# Introduction

The main purpose of this bot is to process IQ Bot CSV result file, but logic can apply to any CSV file.

Because our main sales speech is that we offer a solution which doesn’t require technical skills, IQ Bot design and training look good but most documents (if not all) require some cleaning process to get a result. The cleaning process is currently done by creating a task bot in AAE and can be a blocker for some customers, as it looks like a “development tool” / “scripting”.

The idea of the bot is to provide a ready to use solution, without having to “code” anything in AAE.

By only modifying a CSV file (Extraction\_Rules.csv) and a config file (Config.xml), a business user could “teach” a bot how to clean an IQ Bot raw CSV extraction.

Some available functionalities:

* Regex extraction
* Field format
* Replace values
* Lookup in external list (with confidence level)
* …

# Notes

1. No prerequisite for this bot, ready to use.
2. Using certain features of this metabot require knowledge of Regular Expression, what a Regular Expression Pattern is and what Regular Expression Matching Groups are. A great place to start exploring Regular Expressions (RegEx for short) is <https://regex101.com/>.
3. Date and numeric formats available [here](https://docs.microsoft.com/en-us/dotnet/standard/base-types/formatting-types).

# Assets / Logics

“Post - Process - IQ Bot” uses only two metabots to perform all cleaning operations.

* ConfigFile.mbot
* FormatUtility.mbot

## ConfigFile metabot

This metabot reads a config file with only 4 values:



* **IQ\_Bot\_Success\_Folder**: Your learning instance success folder that contains your IQ Bot success raw extractions
* **Post\_Process\_Results\_Folder**: Folder where will be stored your file after “cleaning” process
* **Extraction\_Rules\_Files**: File containing your extraction rules
* **Convert\_File\_To**: If you need to convert your CSV file to either an XML or JSON (or both) file

## FormatUtility metabot

FormatUtility metabot will perform rules below (in this order) on each extraction line to create a clean CSV file from an IQ Bot raw extraction.

* Column\_Name: Mandatory, same as header value in your IQ Bot success extraction file, bot will apply extraction logic for this field
* Column\_Type: Mandatory, can be either “Form”, same as Form type in IQ Bot, or “TableX” (e.g. Table1, Table2…) to handle multiple tables extraction
* New\_Column\_Name: Optional, if you need to rename a column value to match client’s requirements (“Column\_Name” will be replaced by this value)
* Remove\_Line\_If\_Regex: Optional, will remove all lines in IQ Bot extraction if field value matches this regex value
* Keep\_Line\_If\_Regex: Optional, will keep lines matching this regex value

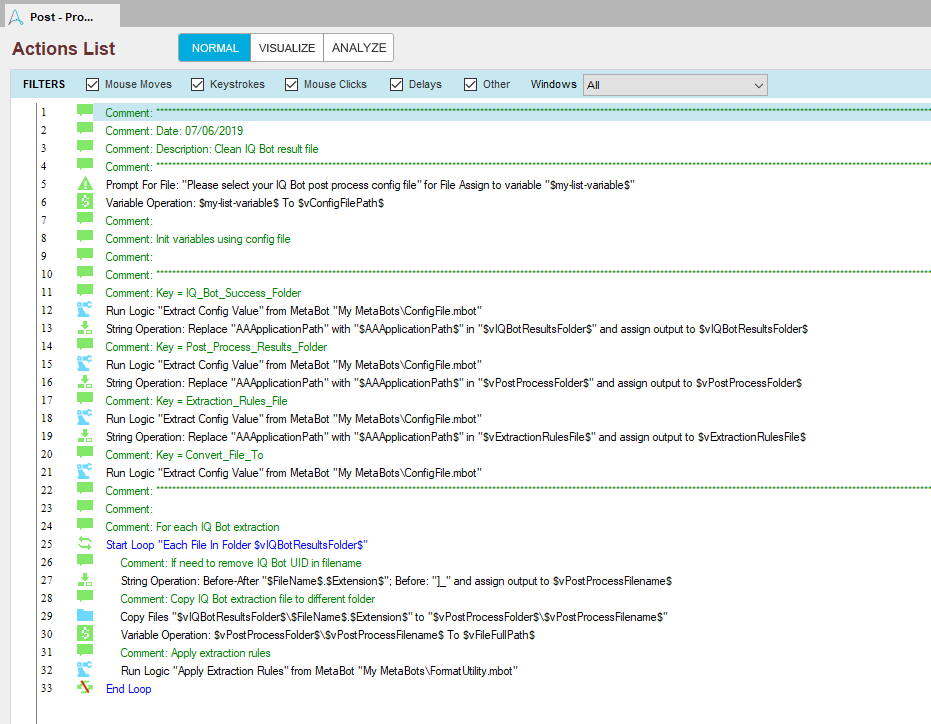
If line hasn’t been removed:

* Remove\_Spaces: Optional, can be either 1 or 0, if 1 all spaces from IQ Bot extraction will be removed
* Regex\_Extraction: Optional, regex extraction from your IQ Bot extraction (e.g **IN\d\*** will extract **IN123456** from **“invoice number: IN123456”**)
* Replace\_Values: Optional, replace values in your IQ Bot extraction
  + “:” = values separator (e.g. old value**”:”**new value)
  + “|” = iteration separator (e.g. old value1”:”new value1**”|”**old value2”:”new value2)
* Trim\_Value: Optional, can be either 1 or 0
* Format\_Type: Optional, can be either date or numeric
* Format\_Value: Optional, if your “Format\_Type” is date your “Format\_Value” can be dd.MM.yyyy, useful if client target system allows only specific format
* Best\_Match\_List\_File: Optional, if you have an external list to match your extraction value against (can be useful with company name extraction)
* Best\_Match\_Confidence: Mandatory if “Best\_Match\_List\_File” is not null, level of confidence in your lookup functionality (e.g. 0.8 means that your extraction needs to match at least 80% of one of the entries in the list, if more than one the highest confidence level is returned)

## Post - Process - IQ Bot

This bot will use a combination of config file and format utility metabots to perform cleaning operations.

Here is the bot code:



Only 15 active RPA commands, rest is comment…

Line # 5: bot will ask you to provide config file path.

From line # 12 to line # 21: bot will read and assign config file values to variables.

From line # 25 to line # 33: bot will loop on each file in IQ Bot results folder (LI success folder)

Lines # 27, 29 and 30: bot will copy and rename (remove IQ Bot CSV file UID) IQ Bot CSV extraction to your Post Process folder (defined in config file)

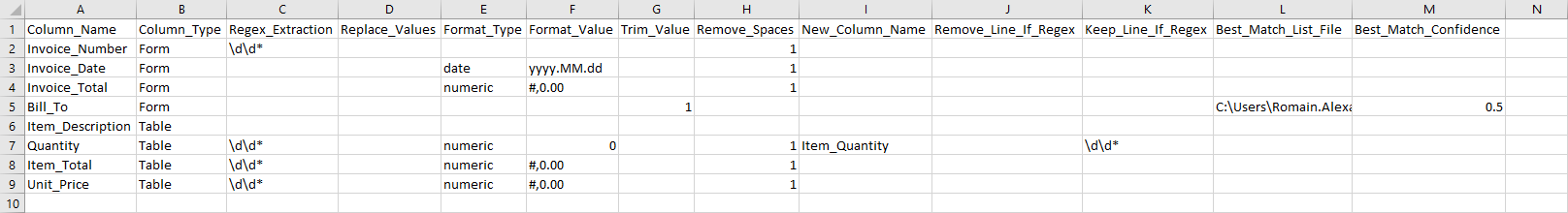
Line # 32: extraction rules application and creation of result file(s)

# Result Examples

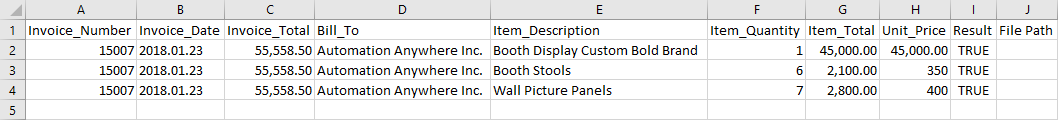
IQ Bot raw extraction:



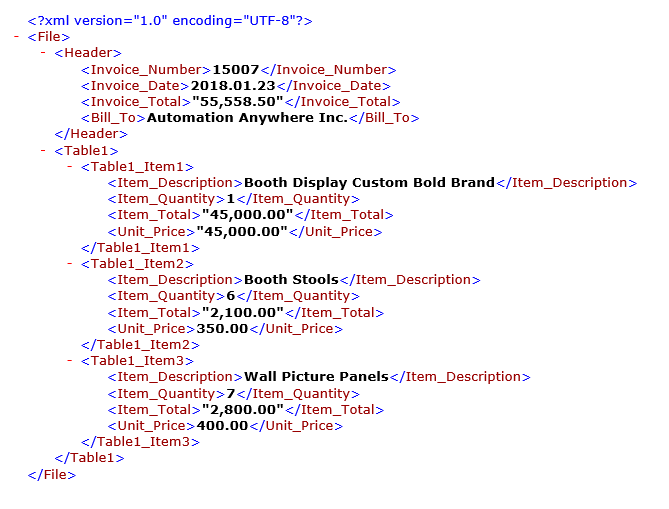
Extraction rules file:



CSV file after clean process:



XML file created:



JSON file created:

