English-Russian character OCR text recognition

README

VERSION 1.0

Table of Contents

1. Introduction

This document contains all necessary information for the user to make use of the Russian-English Character OCR Text Recognition bot. This manual contains a description of the functions, capabilities and configurations of the Bot.

1.1 Overview

A user may often come across a task of dual language text extraction from a scanned document. Customer account numbers in Russia, for example, would often contain Latin (English) characters and Cyrillic (Russian) characters in them. Various documents may include text in several languages (bills of lading, for example). The bot presents an easy way to extract text written in several languages.

1.2 Use Cases

The key cases include:

- Multiple language text extraction from scanned documents using OCR technology.
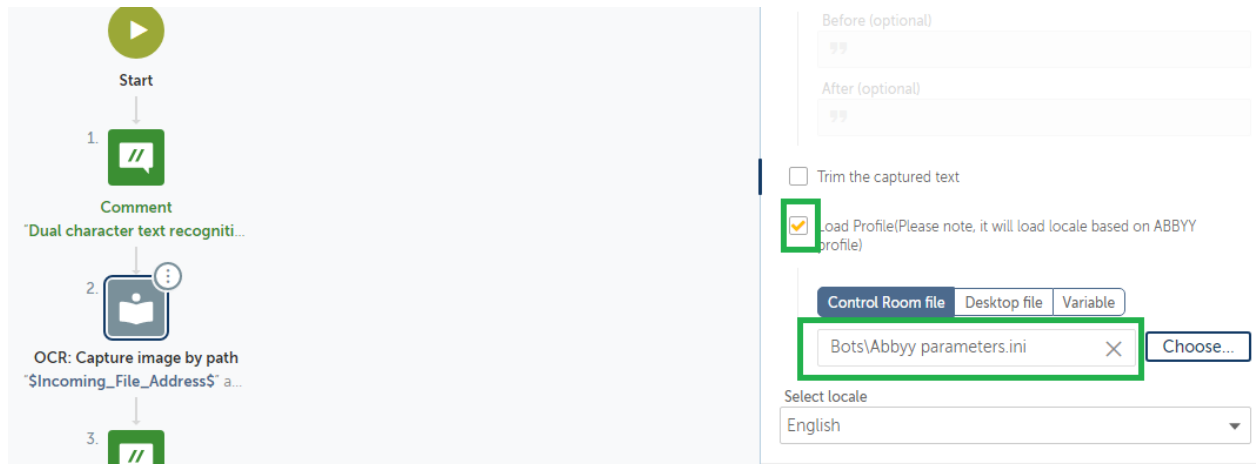
2. Requirements & Prerequisites

There are no system requirements.

2.1 Prerequisites

Automation Anywhere A2019 (A360).

3. Description

This simple bot supports dual character English/Russian OCR text recognition. The process utilizes a file "**Abbyy parameters.ini**", which is used to set a custom OCR locale (set to English and Russian presently) in Automation Anywhere **OCR: Capture image by path** tool.

Picture 1.

The user may modify the **Abbyy parameters.ini** file to enable OCR recognition of any mix of languages supported by Abbyy, for example, French and Russian, etc.

The bot contains two variables used as Input: Incoming_File_Address – must be an image file type, supported by Automation Anywhere tools (.bmp, .jpeg, .jpg, .gif, .png); Recognized_File_Address – the address of the text file, where the recognized text string (Recognized_Text_String) will be saved, the file must be a .txt file and must be saved in the UTF8 format.

The **Abbyy parameters.ini** file may be edited in Notepad. The contents of the file are as follows:

[RecognizerParams]

TextLanguage=English,Russian

[SynthesisParamsForDocument]

DetectDocumentStructure=True

Picture 2.

The second line may be edited to include the desired mix of languages, for example, "French,Russian", etc., any mix of languages supported by Abbyy OCR.

4. Getting started

4.1 Configuration

Configurable variable list with details as mentioned in the Description

Vaiables			
Variable Name	**Type**	**Direction**	**Additional Info**
Incoming_File_Address	String	Input	Address of an image file type, supported by Automation Anywhere tools (.bmp, .jpeg, .jpg, .gif, .png)
Recognized_File_Address	String	Input	Address of the text file, where the recognized text string (Recognized_Text_String) will be saved

Other variables.

Vaiables			
Variable Name	**Type**	**Direction**	**Additional Info**
Recognized_Text_String	String	None	Stores the extracted from an image text

5. Support

Free bots are not officially supported. Questions may be addressed to info@adtb2b.ru.

Company's website is www.adtb2b.ru.