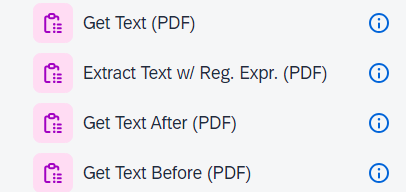
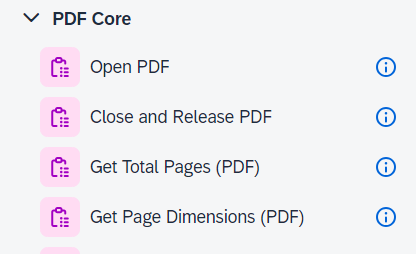
**PDF DATA EXTRACTION**

There are 2 SDK’s available in SAP Process Automation to Extract PDF data:

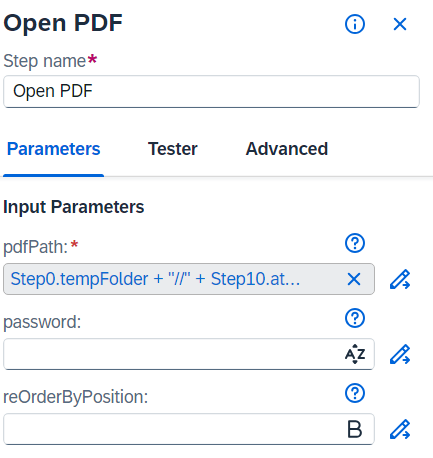
1. PDF SDK
2. DOX SDK

**PDF SDK**

This SDK has following methods:

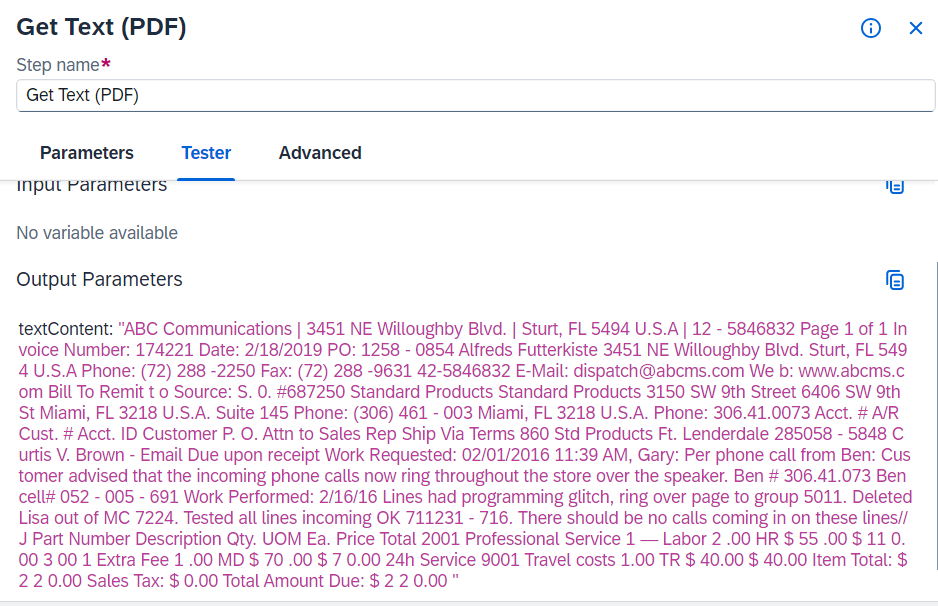


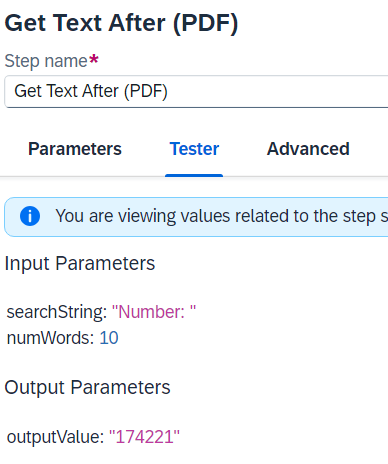
Open PDF method is used to open a pdf from specific path:



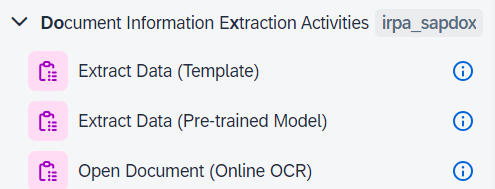
And close and release pdf is used to close an already open PDF. We usually use Get Text (PDF) to read the pdf and use either of Regular Expression or Get Text After or Before method to search for a particular string inside the Text. Once the match is found the value against it is saved in any variable.

Sample Output:

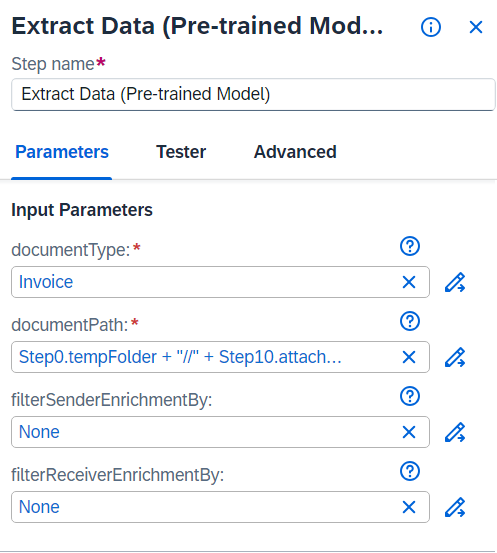


  
**DOX SDK**

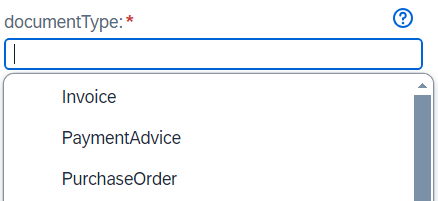
This SDK has following methods:



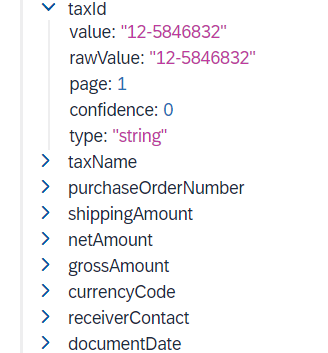
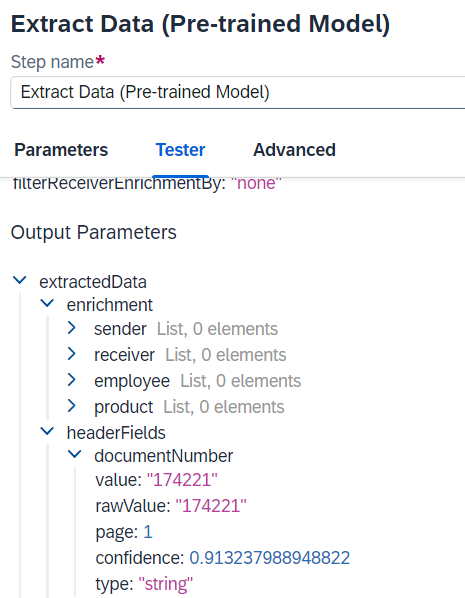
1. Extract Data (Pre-trained Model) -

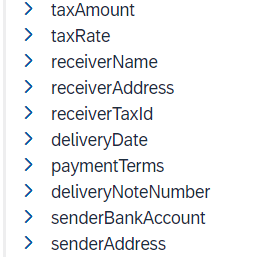


Pre-trained Model are only available for the following document type:



Sample Output:

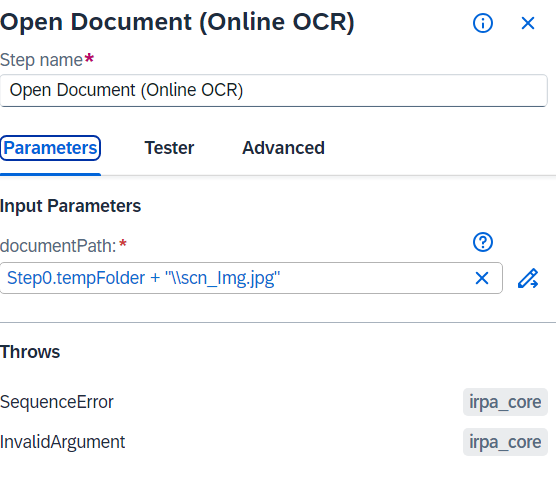




We have to write custom scripts on top of this standard output structure to fetch the data we want.

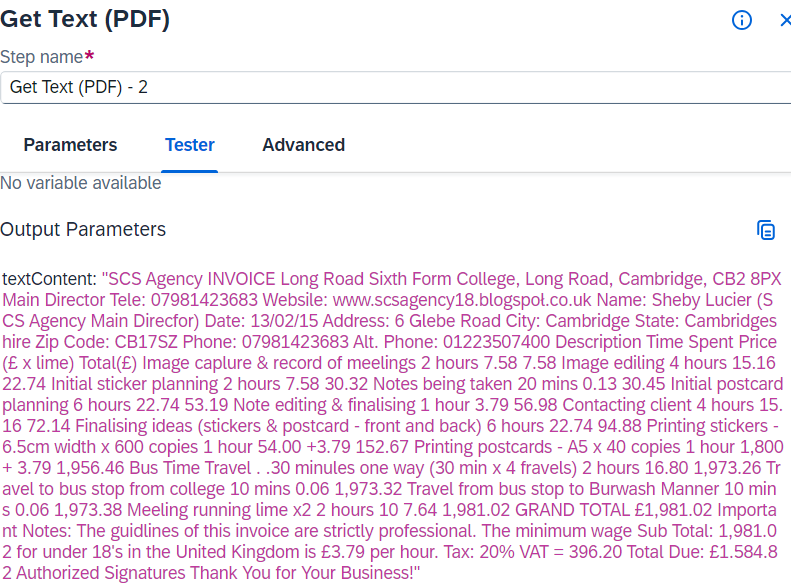
1. Open Document(Online OCR) –

This is like Open PDF instead it reads Image documents.



We can use the same Get Text PDF method and other methods from Open PDF to work with the extracted data.

Sample Output:



1. Extract Data(Template) -