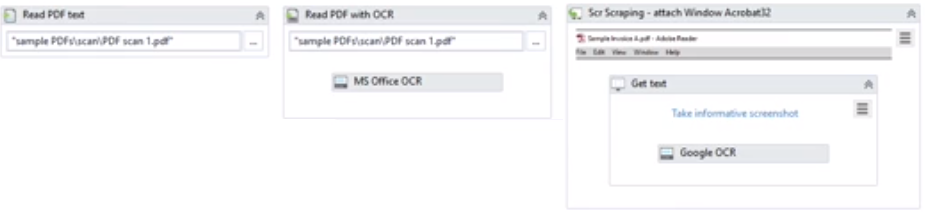
2 Actividades de PDF

* Procesar grandes fragmentos de textos (Read PDF Text)
* Elementos de texto concretos (Read PDF With OCR) porque el texto esta en imagen

Para Read PDF with OCR se debe escoger cual motor, si Google OCR o Microsoft OCR



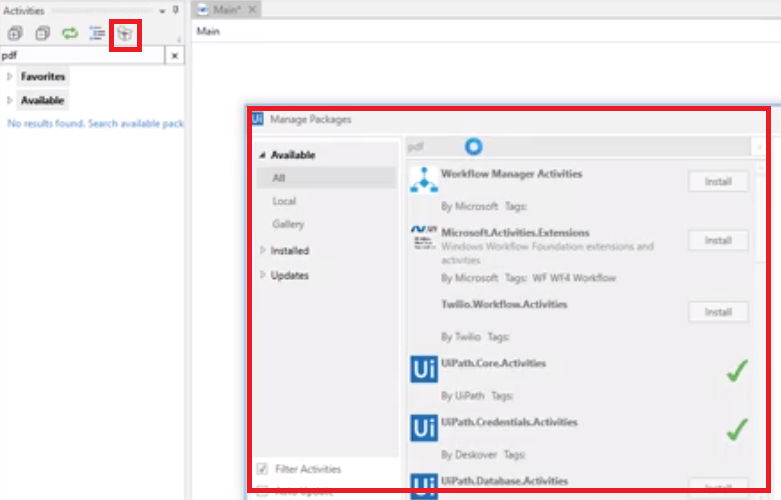
Se pueden ejecutar en background

Otro metodo para extraer texto es con Screen Scraping

Tambien para extraer texto concreto tambien se puede user Anchor Base

RECOMENDABLE USAR Read PDF Text en vez de Read PDF With OCR porque el OCR tiene tendencias a error

El metodo Anchor Base pude ser mas fiable que los otros porque admite cambien estructurales en el archive



**Si queremos extraer información específica de una serie de archivos PDF con una estructura similar pero el flujo de trabajo solo funciona con uno de los archivos de la serie, ¿qué deberíamos investigar?**



Ninguna de las opciones.



La propiedad TimeoutMS.



La propiedad Selector.



La propiedad ContinueOnError

**Multiple Choice**

2)

***Seleccione todas las afirmaciones siguientes que sean verdaderas respecto a la actividad Read PDF with OCR. (Seleccione todas las respuestas que correspondan).***

Ninguna de las opciones

Funciona con archivos PDF nativos

Permite utilizar distintos motores de OCR (Microsoft, Google)

Permite especificar el intervalo de páginas que se va a leer

**Single choice**

3)

**¿Cuál de estas actividades requiere que el archivo PDF esté abierto con Acrobat Reader para poder leerlo?**



Get Text



Read PDF Text



Read Image



Read PDF with OCR

**Multiple Choice**

4)

***Tenemos una factura en PDF nativo y tenemos que leer la cifra en USD que está junto a la etiqueta AMOUNT. ¿Qué métodos podemos aplicar para extraer el valor deseado? (Seleccione todas las respuestas que correspondan).***

Utilizar una actividad Read PDF Text y después utilizar los métodos necesarios de manejo de String para recuperar solo la cifra

Abrir el archivo en Acrobat Reader o cualquier lector de PDF compatible y utilizar Anchor Base con la etiqueta como anclaje

Utilizar la actividad Get Text con un Selector fiable (si está disponible) para extraer solo la cifra del archivo PDF.

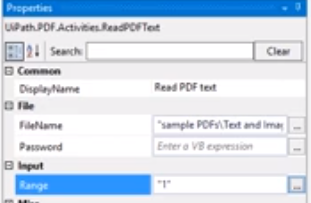
**Single choice**

5)

***¿Cómo puede un robot leer sólo la primera página de un archivo PDF utilizando actividades de PDF? (Seleccione todas las respuestas que correspondan).***



Definiendo la propiedad Range con el valor: “1”



For range 1-5

80%

**Si queremos extraer información específica de varios archivos PDF nativos con la misma estructura, ¿qué actividad debemos utilizar?**



Read PDF with OCR



No hay ninguna actividad para hacer esto



Get Text



Get Text with OCR

2) **¿Puede la actividad Read PDF with OCR abrir el documento PDF en la pantalla para leerlo?**



No



Sí

**Multiple Choice**

3)

***¿Cuál de los siguientes métodos se puede utilizar para leer texto de un documento .pdf nativo? (Seleccione todas las respuestas que correspondan).***

La actividad Read PDF Tex

tLa actividad Read PDF with OCR

La automatización de la interfaz de usuario (abrir el documento pdf en Adobe Acrobat Reader y utilizar Get Text)

**Single choice**

4)

**¿Cuál es el modo más sencillo de obtener el número de factura de un archivo PDF nativo?**



Utilizar la actividad Read PDF with OCR y obtener el valor mediante manejo de String.



Utilizar la actividad Read PDF Text y obtener el valor mediante manejo de String.



Abrir el archivo PDF con Adobe Acrobat Reader y extraer solo la información relevante.

**Single choice**

5)

**Si queremos extraer información específica de una serie de archivos PDF con una estructura similar pero el flujo de trabajo solo funciona con uno de los archivos de la serie, ¿qué deberíamos investigar?**



La propiedad Selector.



La propiedad TimeoutMS.



La propiedad ContinueOnError



Ninguna de las opciones.

20%