

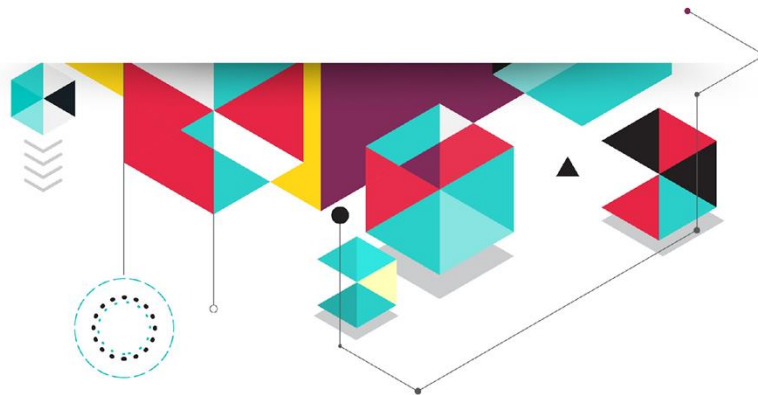


## LECCIÓN 10: AUTOMATIZACIÓN DE PDF, RESUMEN

# Información general



En esta lección hemos visto los tipos de documentos PDF y los métodos disponibles para extraer datos de estos archivos. También hemos analizado los anclajes, una manera de tratar con selectores inestables.



## Enseñanzas clave

- Podemos separar las actividades de PDF en 2 categorías: una para procesar grandes fragmentos de texto o documentos completos y otra para centrarnos en elementos de texto concretos.
- Cuando buscamos extraer datos de un PDF, según el archivo tendremos que elegir una de estas 2 actividades: **Read PDF Text** y **Read PDF With OCR**.
- Estas actividades pueden ejecutarse en segundo plano.
- Otro método para extraer bloques de texto es la herramienta **Screen Scraping**.
- Cuando buscamos extraer un valor concreto de archivos PDF, también se puede utilizar **Anchor Base**.

## Procedimientos recomendados



Utilice **Read PDF Text** en lugar de **Read PDF With OCR** siempre que sea posible porque el OCR tiene tendencia a dar errores.



El método **Anchor Base** puede ser más fiable que los otros porque admite cambios estructurales importantes en el archivo.

# Enlaces útiles



[PDF Data Extraction \(Extracción de datos de PDF\)](#)

