

# Introduction

PDF is a widely used format for sharing documents. In this module we will take a look at the types of PDF documents and the methods you have at your disposal when trying to extract data from a PDF document. We will also take a look at the anchor functionality which is used when dealing with unstable selectors.

## Requirements:

Previous modules should be completed.

## Things you will learn:

- How to use Read PDF Activity
- How to use Read PDF with OCR Activity
- How to use anchors to get data from a field inside a PDF

Note 1: If the PDF is opened with Adobe Reader DC Acrobat, there might be a few steps to take before you can extract specific elements using UiPath studio methods. Start Acrobat and press Ctrl+k. That opens the Preferences pop-up. Select Reading, out of the categories on the left panel. Verify that the drop down Reading Order options is set to the Acrobat recommended option, 'Infer reading order from document (recommended)', 'Page vs. Document': should be set to 'Read the entire document' and 'Confirm before tagging documents' should be unchecked. Then on the left panel, click Accessibility. In the Other Accessibility Options section, check the first two boxes if they are not already checked: 'Use document structure for tab order when no explicit tab order is specified', 'Enable assistive technology support', and click OK.

Note 2: If you still have problems with extracting specific elements from the PDF file opened with Acrobat Reader DC, you could try an older version of Acrobat DC (any version starting with 18 should be fine <https://www.adobe.com/devnet-docs/acrobatetk/tools/ReleaseNotesDC/index.html#> ). Acrobat DC is updated automatically on computer to the last available version. In some of the latest versions (starting with 19) there could be problems with accessibility, Adobe Reader is slowly dropping support for untagged documents. Steps to follow:

- uninstall the current version of Acrobat Reader DC
- install the base release of Acrobat Reader DC <https://www.adobe.com/devnet-docs/acrobatetk/tools/ReleaseNotesDC/continuous/dccontinuous.html#dccontinuous>
- install a patch to any of the versions starting with 18
- disable Adobe Reader auto update <https://www.wintips.org/how-to-disable-automatic-update-in-adobe-reader-dc/>