

# Read PDF Files

[SUGGEST EDITS](#)

You can read and separately extract the content of `.pdf` files using activities that can read all characters included in the document.

Depending on your needs, you can use a simple activity that can recognize the characters, or use one with an OCR engine. The benefits of using an OCR engine are that the document reading can be applied even on scanned, signed, or handwritten documents.

The example below presents two situations of reading a `.pdf` file:

- The first one explains how to read the `.pdf` file while using the [Read PDF Text](#) activity.
- The second one explains how to read the `.pdf` file while using the [Read PDF with OCR](#) activity.

The main difference between the two scenarios is that the second one is also using OCR engines, meaning that the details of extracted information are more accurate than in the first case if the analyzed file is an image, scanned, or includes signed or handwritten fields. You can find both activities in the **UiPath.PDF.Activities** package.

Only one workflow is required for both scenarios, common until the point of asking the user to choose the desired reading method.

This is how the automation process can be built:

1. Open Studio and create a new **Process**.
2. Drag a **Flowchart** container in the **Workflow Designer**.
  - Create the following variable:
3. Drag an **Input Dialog** activity and connect it to the **Start Node**.
  - In the **Properties** panel, add the expression `"Choose one option below:"` in the **Label** field.
  - Add the expression `{"Read PDF Text", "Read PDF With OCR"}` in the **Options** field.
  - Add the value `"Options"` in the **Title** field.
  - Add the variable `chooseOption` in the **Result** field.
4. Place a **Flow Decision** activity below the **Input Dialog** activity and connect it to it.
  - In the **Properties** panel, add the expression `chooseOption = "Read PDF Text"` in the **Condition** field.
5. Drag a **Sequence** container and connect it to the **True** branch of the **Flow Decision** activity. The name of the **Sequence** should be **Read PDF Text**. This activity extracts information by using regular expressions.
  - Create the following variables: