



UWP



Extraction



WINDOWS &gt; GUIDES &gt; EXTRACTION

# Extract text from PDFs in Windows

This guide shows how to extract the full text content from a single page or a whole document.

For more granular control over text extraction, refer to our [parsing guide](#), which outlines the available text APIs in greater detail.

## Page text

The `TextParser` API offers a simple way to get the text from a given PDF page:

```
1 var textParser = await PDFView.Document.GetTextParserAsync(0);
2 var textBlocks = await textParser.GetTextAsync();
```



Note that the `GetTextAsync` method returns a list of `TextBlock`s. Each of these blocks contain the text found in a specific line (all of the continuous group of glyphs in that line).

Using the returned list of text lines from `GetTextAsync`, the page text can be unified to a single string:

```
1 var unifiedText = new StringBuilder();
2
3 for (var i = 0; i < pageCount; i++)
4 {
5     var textParser = await PDFView.Document.GetTextParserAsync(i);
6     var textBlocks = await textParser.GetTextAsync();
7
8     foreach (var textBlock in documentTextBlocks)
```



ASK AI

```
9      {
10         unifiedText.Append(textBlock.Contents);
11         unifiedText.Append(" ");
12     }
13 }
```

This will change, depending on your specific use case and document formatting, but it gives an idea of how to structure your `TextParser` usage. For a more in-depth look at the parser and how it interacts with glyphs and words and text blocks, see the [parsing guide](#).

## Document text

As each page has its own `TextParser`, the idea is similar to the above. Keep in mind that parsing can be performance intensive, especially for larger documents:

```
1  var pageCount = await PDFView.Document.GetTotalPageCountAsync();
2  var documentTextBlocks = new List<TextBlock>();
3  var unifiedText = "";
4
5  for (var i = 0; i < pageCount; i++)
6  {
7      var textParser = await PDFView.Document.GetTextParserAsync(i);
8      documentTextBlocks.AddRange(await textParser.GetTextAsync());
9  }
10
11 foreach (var textBlock in documentTextBlocks)
12 {
13     unifiedText += textBlock.Contents + " ";
14 }
```



Was this helpful?

✓ YES

✗ NO

Questions? [Contact us](#)