Extract Text from PDF in Node.js

Contents [Hide]

- Extract Text From PDF Document

Extract Text From PDF Document

Extracting text from the PDF document is a very common and useful task. Extracting text from PDFs serves a variety of purposes, from improving search and availability to enabling analysis and automation of data in various fields such as business, research, and information management.

In case you want to extract text from PDF document, you can use AsposePdfExtractText function. Please check following code snippet in order to extract text from PDF file using Node.js via C++.

Check the code snippets and follow the steps to extract text from your PDF:

CommonJS:

1. Call `require` and import `asposepdfnodejs` module as `AsposePdf` variable.
2. Specify the name for the PDF file from which the text will be extracted.
3. Call `AsposePdf` as Promise and perform the operation for extracting text. Receive the object if successful.
4. Call the function AsposePdfExtractText.
5. Extracted text is stored in the JSON object. Thus, if 'json.errorCode' is 0, the extracted text is displayed using console.log. If the json.errorCode parameter is not 0 and, accordingly, an error appears in your file, the error information will be contained in 'json.errorText'.

```javascript
const AsposePdf = require('asposepdfnodejs');
const pdf_file = 'Aspose.pdf';
AsposePdf().then(AsposePdfModule => {
    /*Extract text from a PDF-file*/
    const json = AsposePdfModule.AsposePdfExtractText(pdf_file);
    console.log("AsposePdfExtractText => %O", json.errorCode == 0 ? json.extractText : json.errorText);
});
```

ECMAScript/ES6:

1. Import the `asposepdfnodejs` module.
2. Specify the name for the PDF file from which the text will be extracted.
3. Initialize the AsposePdf module. Receive the object if successful.
4. Call the function AsposePdfExtractText.
5. Extracted text is stored in the JSON object. Thus, if 'json.errorCode' is 0, the extracted text is displayed using console.log. If the json.errorCode parameter is not 0 and, accordingly, an error appears in your file, the error information will be contained in 'json.errorText'.

```javascript
import AsposePdf from 'asposepdfnodejs';
const AsposePdfModule = await AsposePdf();
const pdf_file = 'Aspose.pdf';
/*Extract text from a PDF-file*/
const json = AsposePdfModule.AsposePdfExtractText(pdf_file);
console.log("AsposePdfExtractText => %O", json.errorCode == 0 ? json.extractText : json.errorText);
```

‹ Extract Images from PDF in Node.js

Subscribe to Aspose Product Updates

Get monthly newsletters & offers directly delivered to your mailbox.

Enter email address   Submit

Home	Products	New Releases
Pricing	Docs	Live Demos
Free Support	Free Consulting	Paid Support
Paid Consulting	Blog	Websites
About		

© Aspose Pty Ltd 2001-2024. All Rights Reserved.

Privacy Policy Terms of use Contact