

PDF to HTML Conversion in PHP



This article will show you how to convert any PDF document to HTML format using PHP. This step by step tutorial will show you how to create html files from PDF with php.

You will need PDF to HTML Package to perform this task. Here is the url of the package - <http://sourceforge.net/projects/pdfhtml/>

After the installation of the package use the following command to execute it. You can also execute these commands from SSH or PHP script. We will focus at php script command execution. Let us consider the installed files are present in /usr/bin.

```
system('/usr/bin/pdfhtml /var/www/website/processed/example.pdf') // This will create a HTML file in the processed folder.
```

```
system('/usr/bin/pdfhtml /var/www/website/processed/example.pdf -') // This command will not create an HTML file but it will show the output of the file at screen.
```

Few Common Errors

- BAD Color Error: This error usually appear if the package doesn't install properly.
- Execution Error: It is also a common error that php doesn't execute the command and output file doesn't generate. To solve this error you need to confirm few things.
 - Confirm PHP is not running in the safe mode.
 - You must execute the command as root. So you also need to set the Apache Security Settings.

// //]]> //

PDF to TEXT Conversion in PHP



It is quite simple to calculate characters of a pdf document. To accomplish this task, we will use pdf2html library. Please download and install [pdf2html library](#).

PHP Code to execute PDF Conversion and Characters Calculation

Linux command execution to convert the pdf to text format.

'/usr/bin/pdftotext ' . \$file_path; //File path must be the absolute server path.

PHP command execution to convert the pdf to text format.

```
shell_exec('/usr/bin/pdftotext ' . $file_path);
```

Complete code to upload a file to the processed folder in your root directory.

PHP Code to execute PDF Conversion and Characters Calculation

```
if(move_uploaded_file($_FILES[$filen]['tmp_name'],'processed/'.$_FILES[$filen]['name'])
))
{
$file_name=$_FILES[$filen]['name'];
$file_path=$_SERVER['DOCUMENT_ROOT'].'processed/'.$_FILES[$filen]['name'];
$file_name=str_replace('.pdf','.txt',$file_name);
$output=shell_exec('/usr/bin/pdftotext ' . $file_path);
sleep(2);
$handle = fopen($file_name, "r");
$contents = fread($handle, filesize($file_name));
fclose($handle);
$file_count = strlen(str_replace(' ','',$contents));
}
```

Troubleshooting

// //]]> //

1. shell_exec function will not execute. If you don't have permission to run ssh commands and also if your php is running in the safe mode.
2. This script will generate a text file with same name and directory where you have placed the pdf file. So if the file isn't create in that directory and your program will work you will able to track the file in the root directory. This means you have to correct your file path.
3. Cannot count the calculation and upload the file. It is necessary to change the rights of processed folder to 777.

If you have further questions about this post, kindly post your comments.

Online URL: <https://www.articlediary.com/article/php-script-for-pdf-to-html-conversion-125.html>