

Lectura de un pdf mediante código java

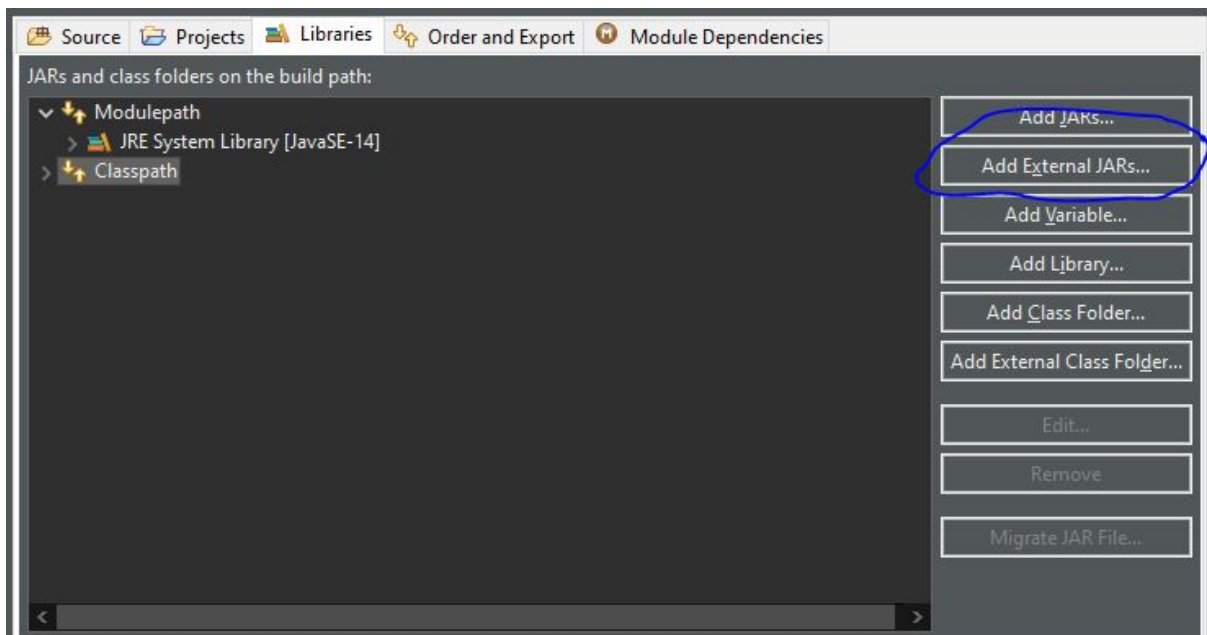
Instalación

Para leer un contenido de un PDF con java usaremos una librería llamada jPDFText que nos permitirá extraer el texto y guardarlo en un String.

Comenzamos descargando la librería desde la siguiente página

<https://www.qoppa.com/files/pdf/text/jars/jPDFText.jar>

Una vez descargado lo metemos en una carpeta de librerías dentro de nuestro proyecto y abrimos eclipse. Hacemos click derecho > configure build path y en la pestaña de librerías seleccionamos classpath y add external Jar



Una vez lo hayamos añadido le damos a apply and close y ya tendremos la librería importada

Utilización

Una vez tengamos la librería importada en el build path del proyecto debemos importar esa misma librería para poder usarla con la siguiente línea

```
import com.qoppa.pdfText.PDFText;
```

Posteriormente deberemos crear una instancia a una clase PDFText que nos ofrece la librería que hemos importado, que reciba la ruta del archivo y una contraseña pero como esta no se va a usar se establecerá como null

```
String ruta = sc.next();  
PDFText pdfText = new PDFText (ruta, null);
```

Esto guardará el contenido del PDF en una variable en la clase PDFText así que para sacar el texto deberemos llamar al método getText y guardarlo en una variable de tipo String

```
String docText = pdfText.getText();
```

Con el contenido de un PDF en la variable solo quedaría escribirlo en la consola mediante un println o si estamos creando una clase que devuelva el texto, añadir la variable al return

```
System.out.printf(docText);  
  
return(docText);
```

Nota: Este método deberá estar dentro de un try catch para controlar las excepciones, si no el programa tiende a dar problemas.

Metodo completo:

```
public String leerPDF() {
    Scanner sc = new Scanner(System.in);
    try {

        System.out.println("Escribe la ruta del archivo que quieres leer");
        String ruta = sc.next();
        PDFText pdfText = new PDFText (ruta, null);

        String docText = pdfText.getText();

        |

        System.out.printf(docText);

        return(docText);

    } catch (Throwable t) {
        // TODO Auto-generated catch block
        t.printStackTrace();
    }
    return null;
}
```