

[How to Test if a File is PDF \(Java\)](#)

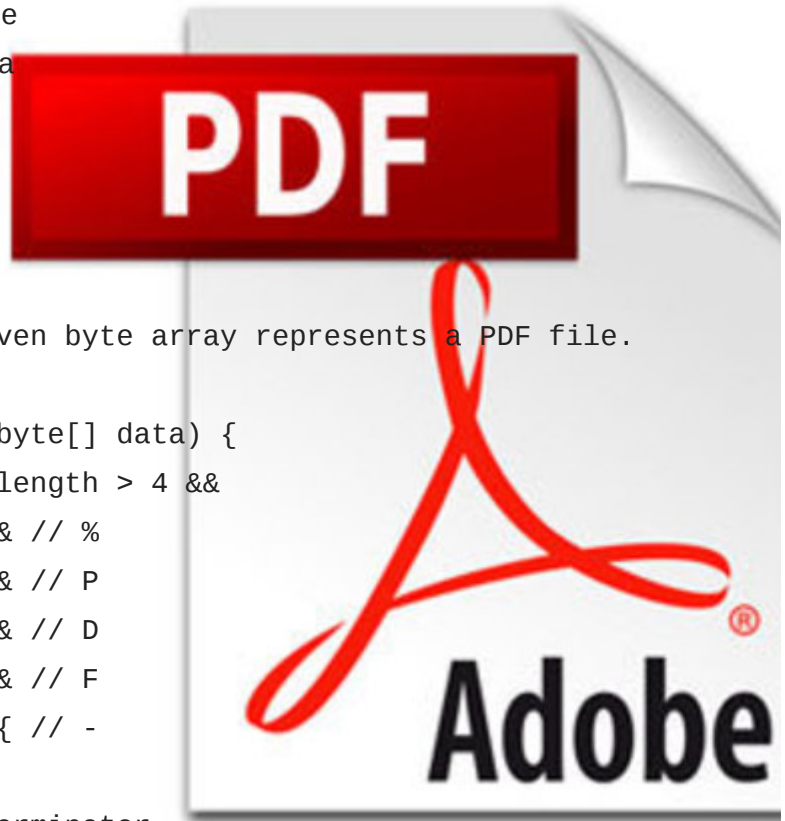
By [Roger Keays](#), 26 February 2016

Here is a simple java function to check if some data represents a PDF file. It is adapted from a [C# function posted on Stack Overflow by NinjaCross](#).

```
/**
 * Test if the data in the given byte array represents a PDF file.
 */
public static boolean is_pdf(byte[] data) {
    if (data != null && data.length > 4 &&
        data[0] == 0x25 && // %
        data[1] == 0x50 && // P
        data[2] == 0x44 && // D
        data[3] == 0x46 && // F
        data[4] == 0x2D) { // -

        // version 1.3 file terminator
        if (data[5] == 0x31 && data[6] == 0x2E && data[7] == 0x33 &&
            data[data.length - 7] == 0x25 && // %
            data[data.length - 6] == 0x25 && // %
            data[data.length - 5] == 0x45 && // E
            data[data.length - 4] == 0x4F && // O
            data[data.length - 3] == 0x46 && // F
            data[data.length - 2] == 0x20 && // SPACE
            data[data.length - 1] == 0x0A) { // EOL
            return true;
        }

        // version 1.3 file terminator
        if (data[5] == 0x31 && data[6] == 0x2E && data[7] == 0x34 &&
            data[data.length - 6] == 0x25 && // %
            data[data.length - 5] == 0x25 && // %
            data[data.length - 4] == 0x45 && // E
            data[data.length - 3] == 0x4F && // O
            data[data.length - 2] == 0x46 && // F
```



```

        data[data.length - 1] == 0x0A) { // EOL
            return true;
        }
    }
    return false;
}

```

The function takes a **byte[]** to make it possible to test data in streams and files. To read a file into a byte array you can use **java.nio.Files.readAllBytes**, e.g:

```

assertTrue(is_pdf(Files.readAllBytes(Paths.get("output.pdf")));

```

You'll have to handle the **IOExceptions** from those methods too. To test streamed output, just render to a **ByteArrayOutputStream** and call the **getBytes()** method after the output has rendered.

The unit tests for this function are below. Please feel free to contribute test cases, especially ones that fail.

```

@Test
public void test_valid_pdf_1_3_data_is_pdf() {
    assertTrue(is_pdf("%PDF-1.3 CONTENT %%EOF \n".getBytes()));
}

```

```

@Test
public void test_valid_pdf_1_4_data_is_pdf() {
    assertTrue(is_pdf("%PDF-1.4 CONTENT %%EOF\n".getBytes()));
}

```

```

@Test
public void test_invalid_data_is_not_pdf() {
    assertFalse(is_pdf("Hello World".getBytes()));
}

```

About Roger Keays



Roger Keays is an artist, an engineer, and a student of life. He has no fixed address and has left footprints on 40-something different countries around the world.

Roger is addicted to surfing. His other interests are music, psychology, languages, the proper use of semicolons, and finding good food.