

# Exercise

## รูปแบบข้อมูลที่พบบ่อย

### สาระสำคัญ

1. โมดูล: ทำความรู้จักการเล่าเรื่องที่ซับซ้อนด้วยข้อมูล<sup>1</sup>
2. บทเรียน: รูปแบบข้อมูลที่พบบ่อย
3. หัวข้อย่อย: ไฟล์ PDF ที่มีโครงสร้างตาราง
4. วัตถุประสงค์: ค้นหาข้อมูล และเก็บข้อมูลที่อยู่ในไฟล์ PDF ที่มีโครงสร้างตาราง
5. เวลา: 60 นาที

### ขั้นตอน

#### บริบทการเรียนรู้

#### สไลด์นำเสนอที่เกี่ยวข้อง

1. ดาวน์โหลดไฟล์ PDF [ตารางที่ 1 พื้นที่ป่าไม้ของประเทศไทย ปี พ.ศ. 2516-2561](#) จากฐานข้อมูลสถิติกรมป่าไม้ปี 2561
2. ติดตั้งโปรแกรม Tabula (JavaScript) <https://tabula.technology/> โดยกดแตกไฟล์แล้วติดตั้ง

Tabula



---

<sup>1</sup> เอกสารนี้ดัดแปลงมาจากคู่มือการอบรม Introduction to Data Literacy ของธนาคารโลก โดย Eva Constantaras และปรับปรุงโดย Yan Naung Oak, Open Development Cambodia และ Open Development Initiative ซึ่งได้รับอนุญาตภายใต้ [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](#)

## Tabula



Tabula is a tool for liberating data tables locked inside PDF files.

[View the Project on GitHub](#)  
tabulapdf/tabula

Download for  
**Windows**

Download for  
**Mac**

View source on  
**GitHub**

Current Version: **1.2.1**

Other Versions: [pre-releases & archives](#)

Need help? Open an [issue on Github](#).

Donate: Help support this project by [backing us on OpenCollective](#).

We'd love to hear from you! Say hi on Twitter at [@TabulaPDF](#)

## Latest Version: Tabula 1.2.1

June 4, 2018

Tabula 1.2.1 fixes several bugs in the user interface and processing backend. (You can read about all the changes [in the release notes](#).)

Download Tabula below, or [on the release notes page](#).

Special thanks [to our OpenCollective backers](#) for supporting our work on Tabula; if you find Tabula useful in your work, please consider [a one-time or monthly donation](#).

### How Can Tabula Help Me?

If you've ever tried to do anything with data provided to you in PDFs, you know how painful it is — there's no easy way to copy-and-paste rows of data out of PDF files. Tabula allows you to extract that data into a CSV or Microsoft Excel spreadsheet using a simple, easy-to-use interface. Tabula works on Mac, Windows and Linux.

### Who Uses Tabula?

Tabula is used to power investigative reporting at news organizations of all sizes, including [ProPublica](#), [The Times of London](#), [Foreign Policy](#), [La Nación \(Argentina\)](#), [The New York Times](#) and the [St. Paul \(MN\) Pioneer Press](#).



[Download](#) [Help](#)

# JAVA + YOU, DOWNLOAD TODAY!

[Java Download](#)

[» What is Java?](#) [» Do I have Java?](#) [» Need Help?](#) [» Uninstall](#)



- Help Resources
  - » What is Java?
  - » Remove Older Versions
  - » Disable Java
  - » Error Messages
  - » Troubleshoot Java
  - » Other Help
- Windows 64-bit Users
  - Do you use both 32-bit and 64-bit browsers?
  - » FAQ about 64-bit Java for Windows
- Offline Installation
  - Trouble downloading?
  - Try the [offline installer](#)

## Download Java for Windows

**Recommended Version 8 Update 241 (filesize: 1.97 MB)**  
Release date January 14, 2020

**Important Oracle Java License Update**

**The Oracle Java License has changed for releases starting April 16, 2019.**

The new [Oracle Technology Network License Agreement for Oracle Java SE](#) is substantially different from prior Oracle Java licenses. The new license permits certain uses, such as personal use and development use, at no cost -- but other uses authorized under prior Oracle Java licenses may no longer be available. Please review the terms carefully before downloading and using this product. An FAQ is available [here](#).

Commercial license and support is available with a low cost [Java SE Subscription](#).

Oracle also provides the latest OpenJDK release under the open source [GPL License](#) at [jdk.java.net](#).

**!** In Windows 10, the Edge browser does not support plug-ins and therefore will not run Java. Switch to a different browser (Internet Explorer, for example) to run the Java plug-in. Select the More Actions option located at the top right of the Edge browser and click on Open with Internet Explorer. [More info](#)

**Agree and Start Free Download**

Java Setup - Welcome

**Welcome to Java - Updated License Terms**

The terms under which this version of the software is licensed have changed. [Updated License Agreement](#)

This version of the Java Runtime is licensed only for your personal (non-commercial) desktop and laptop use.

Commercial use of this software requires a separate license from Oracle or from your software vendor. Click Install to accept the license agreement and install Java now or click Remove to uninstall it from your system.

No personal information is gathered as part of our install process. [Details on the information we collect](#)

Change destination folder

**Install** Remove

3. หลังติดตั้ง เข้าที่โฟลเดอร์ Tabula และไปยัง <http://127.0.0.1:8080/>

Name	Date modified	Type	Size
AUTHORS	4/6/2018 22:18	Text Document	
LICENSE	4/6/2018 22:18	Text Document	
README	4/6/2018 22:18	Text Document	
tabula	4/6/2018 22:18	Application	
tabula	4/6/2018 22:18	Executable Jar File	

กด ดับเบิลคลิก ที่ tabula และรอสัญชีตัวหนังสือเยอะ ๆ ห้ามปิดหน้าต่างสีดำ

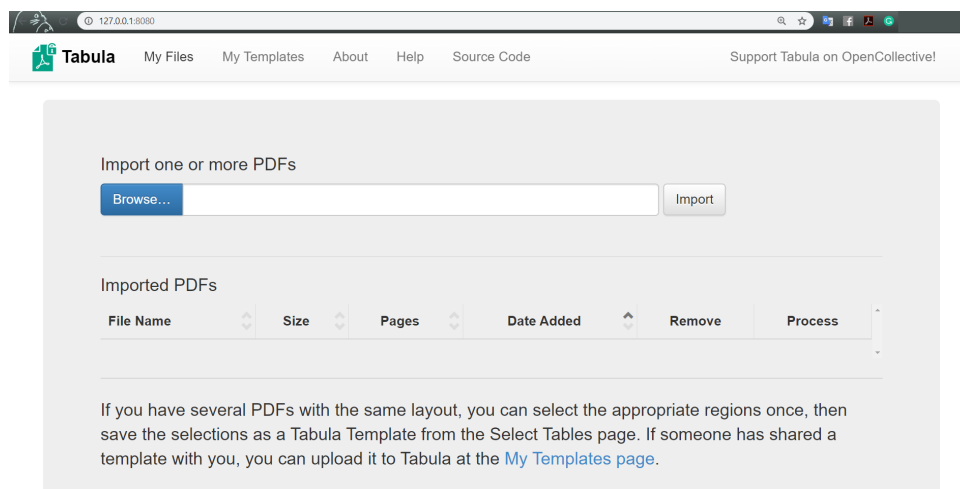
```

C:\Users\Admin\Desktop\tabula\tabula.exe
2563-03-10 10:33:53.297:INFO::main: Logging initialized @401ms
2563-03-10 10:33:53.310:INFO:oejr.Runner:main: Runner
2563-03-10 10:33:53.422:INFO:oejs.Server:main: jetty-9.2.9.v20150224
2563-03-10 10:33:59.145:WARN:oeja.AnnotationConfiguration:main: ServletContainerInitializers: detected. Class hierarchy:
empty
INFO: jrubby 9.1.15.0 (2.3.3) 2017-12-07 929fde8 Java HotSpot(TM) Client VM 25.241-b07 on 1.8.0_241-b07 +jit [mswin32-x86
]
INFO: using a shared (threadsafe!) runtime
DATA_DIR = C:\Users\Admin\AppData\Roaming\Tabula
DOCUMENTS_BASEPATH = C:\Users\Admin\AppData\Roaming\Tabula\pdfs
ENABLE_DEBUG_METHODS = false
running under / as root URI
=====
http://127.0.0.1:8080
=====
should we open browser?
java.lang.Boolean.getBoolean('tabula.openBrowser'): true
java.awt.Desktop.isDesktopSupported: true
java.awt.Desktop.getDesktop(): yes
=====
Launching web browser to http://127.0.0.1:8080

If it does not open in 10 seconds, you may manually open
a web browser to the above URL.
When you're done using the Tabula interface, you may
return to this window and press "Control-C" to close it.
=====
2563-03-10 10:34:05.676:INFO:oejsh.ContextHandler:main: Started o.e.j.w.WebAppContext@18ed237f/, file:/C:/Users/Admin/App

```

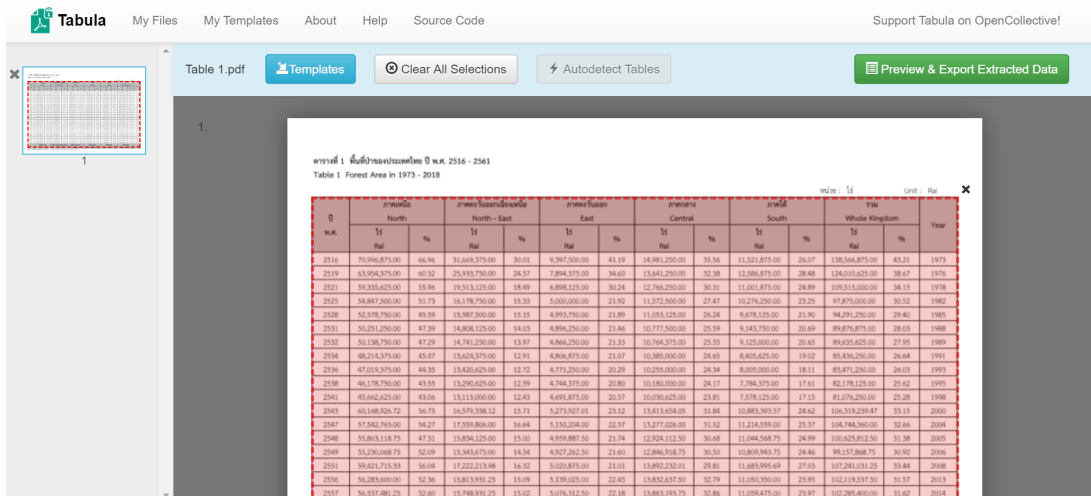
4. ถ้าหน้าจอไม่ขึ้น เข้า Web Browser และไปยัง <http://127.0.0.1:8080/>



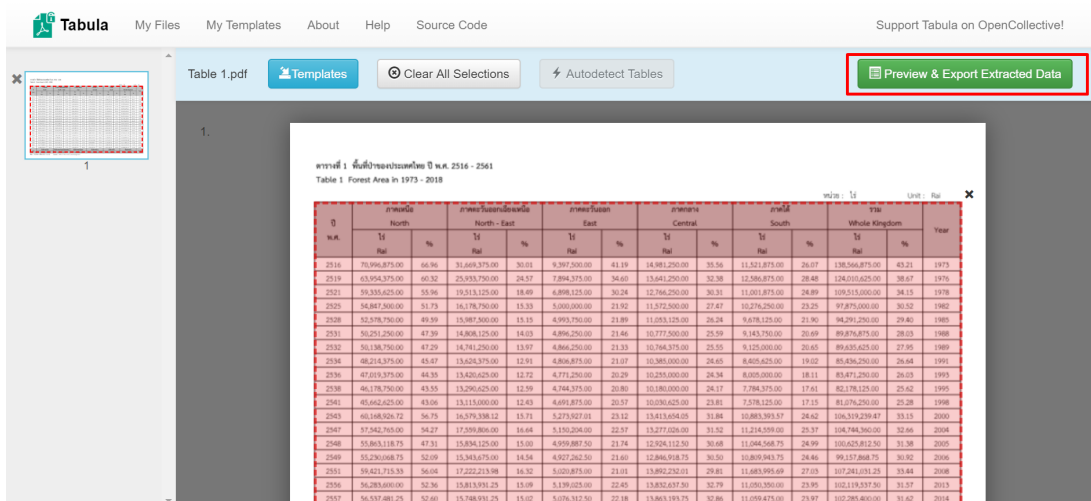
5. กดเลือกไฟล์ **Browse...** ที่ต้องการ และกดนำเข้าข้อมูล

Import

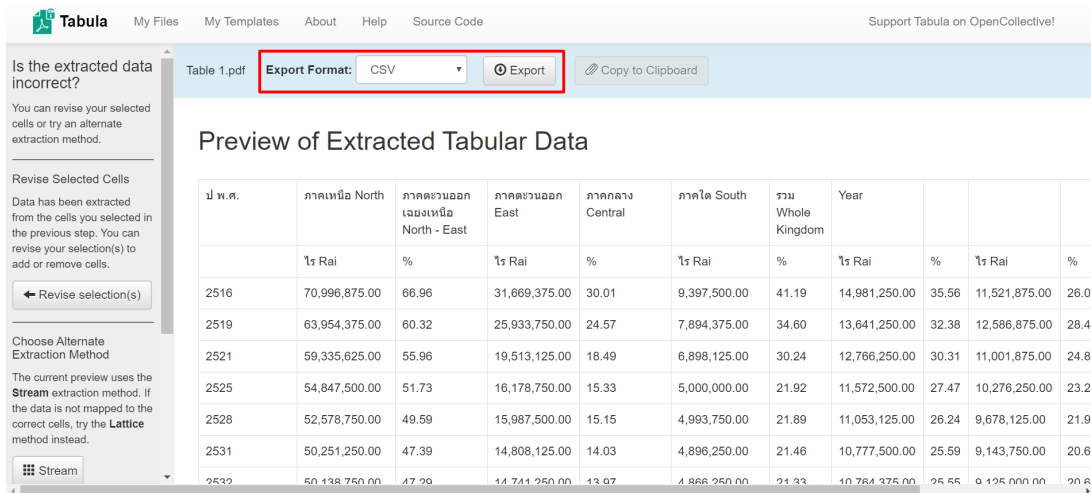
6. กดเลือกข้อมูลที่อยู่ในตาราง (คลิกและลากพื้นที่ด้วยเมาท์)



7. กดปุ่ม Preview & Export Extract Data เพื่อแปลงข้อมูล PDF เป็นตาราง



8. เลือกไฟล์ CSV และดึงข้อมูลที่ปุ่ม Export เพื่อบันทึกไฟล์



ตัวอย่างอื่น ๆ ในโฟลเดอร์

## การประเมิน

สามารถใช้เครื่องมือ Tabula ในการดึงข้อมูลจากไฟล์ PDF ที่อยู่ในรูปแบบของตารางได้ และเตรียมข้อมูลในรูปแบบของไฟล์ CSV เพื่อนำไปใช้ต่อในโปรแกรมคอมพิวเตอร์อื่น ๆ ได้

■ ■ ■