

bookget 使用说明

目录

- 简介.....1
- 使用说明.....1
- 支持的数字图书馆2
- URL 格式示例.....3
- IIIF 自动检测下载5
- 通用批量文件下载6

简介

本软件编写之初衷，是用于下载《钦定古今图书集成》5341 册。是个人自用工具，功能是按作者喜好编写，未能尽满足于众人，还请见谅。目前共支持世界 27 个数字图书馆。

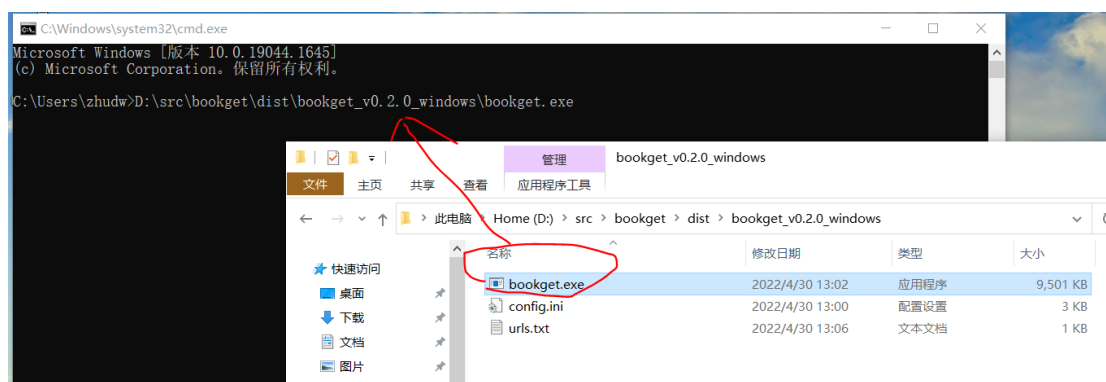
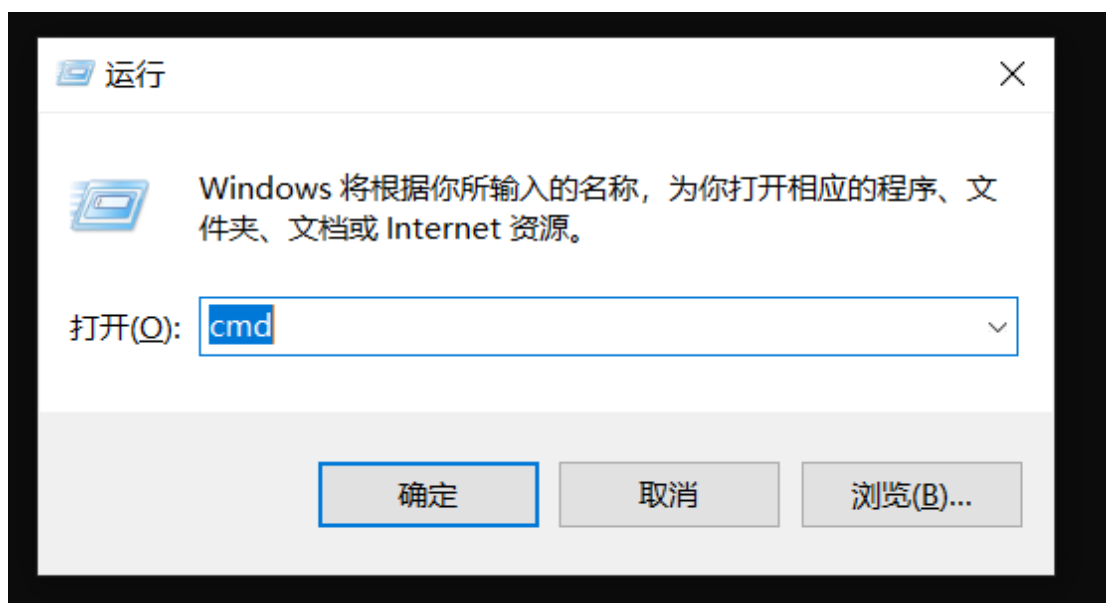
文件名	说明
bookget.exe	Windows 7/10/11 64 位系统（不支持 32 位）
bookget	Linux 64 位系统（文件名虽相同，压缩包不同）
bookget	MacOSX 系统 Intel 处理器（文件名虽相同，压缩包不同）
bookget_m2	MacOSX 系统 M1/M2 处理器
config.ini	配置文件（记事本可修改）
urls.txt	下载地址（记事本可修改）

使用说明

- A. 编写 urls.txt 每行一个下载地址，即是你能在线阅读的书籍网址，可参考后文格式。
- B. 双击运行，或者先打开终端，再把 bookget 拖拽到终端窗口，回车即可。
- C. 等待下载完成，程序会自动退出。请到 Downloads 目录下查看已下载文件。

Windows 用户：

- 1、 点击你的电脑屏幕左下角【开始】菜单，在菜单中找【命令提示符】或【PowerShell】
- 2、 【可选】使用快捷键：Windows 图标 + R，“运行”中输入 cmd 或 powershell，回车）。
- 3、 如下图所示，找到 bookget.exe 拖拽到黑色窗口中，回车即可运行。



MacOSX 用户：打开系统自带应用【终端】，拖拽 bookget 到终端黑色窗口中，回车运行。
Linux 用户： 无需介绍。

支持的数字图书馆

1. 中国国家图书馆 <http://read.nlc.cn/thematDataSearch/toGujiIndex>
2. 哈佛大学图书馆 <https://hollis.harvard.edu/>
3. 中国台北图书馆 <http://rbook.ncl.edu.tw/>
4. hathitrust 数字图书馆 <https://www.hathitrust.org/>
5. 普林斯顿大学图书馆 <https://library.princeton.edu/>
6. 日本京都大学图书馆 <https://rmda.kulib.kyoto-u.ac.jp/>
7. 美国国会图书馆 <https://www.loc.gov/collections/chinese-rare-books/>
8. 日本国立国会图书馆 <http://dl.ndl.go.jp/>
9. 日本 E 国宝 e-Museum <https://emuseum.nich.go.jp/>
10. 日本宫内厅书陵部 https://db2.sido.keio.ac.jp/kanseki/T_bib_search.php
11. 日本东京大学东洋文化研究所 <http://shanben.ioc.u-tokyo.ac.jp/list.php>
12. 中国香港中文大学图书馆 <https://repository.lib.cuhk.edu.hk/sc/collection>
13. 牛津大学博德利图书馆

- <https://digital.bodleian.ox.ac.uk/collections/chinese-digitization-project/>
14. 日本国立公文书馆（内库文库） <https://www.digital.archives.go.jp/>
 15. 日本东洋文库 <http://dsr.nii.ac.jp/toyobunko/index.html.ja>
 16. 日本早稻田大学图书馆 <https://www.wul.waseda.ac.jp/kotenseki/search.php>
 17. 韩国国家图书馆 <https://lod.nl.go.kr/> 或 <https://www.dlibrary.go.kr/>
 18. 新日本古典籍综合数据库 <https://kotenseki.nijl.ac.jp/>
 19. 德国柏林国立图书馆 <https://digital.staatsbibliothek-berlin.de>
 20. 日本京都大学人文科学研究所 - 东方学数字图书博物馆
<http://kanji.zinbun.kyoto-u.ac.jp/db-machine/toho/html/top.html>
 21. 英国图书馆（藏有手稿本） <http://www.bl.uk/manuscripts/>
 22. 中国香港科技大学图书馆 <https://lbezone.ust.hk/>
 23. 中国台北故宫博物院 - 善本古籍 <https://rbk-doc.npm.edu.tw/>
 24. 日本国立历史民俗博物馆 <https://khirin-a.rekihaku.ac.jp/>
 25. 日本本市立米泽图书馆 <https://www.library.yonezawa.yamagata.jp/dg/zen.html>
 26. 日本庆应义塾大学图书馆 <https://dcollections.lib.keio.ac.jp/ja/kanseki>
 27. 日本关西大学图书馆 <https://www.iif.ku-orcas.kansai-u.ac.jp/books>

URL 格式示例

要下载的图书网址存在 urls.txt 文件中，每行一个 URL，回车换行，可以有多个 URL。
原则上，以你能在浏览器中【在线阅读】书籍正文的 URL 为下载地址。

中国国家图书馆-整书多册 URL（白天很慢，可夜间或清晨下载）：

http://read.nlc.cn/allSearch/searchDetail?searchType=1002&showType=1&indexName=data_892&fid=411999021002

中国国家图书馆-图书单册 URL：

<http://read.nlc.cn/OutOpenBook/OpenObjectBook?aid=403&bid=70621.0>

hathitrust 数字图书馆-图书单册 URL

<https://babel.hathitrust.org/cgi/pt?id=ucl.c087423515&view=lup&seq=1&skin=2021>

哈佛大学图书馆-图书在线阅读（分享）URL

<https://iif.lib.harvard.edu/manifests/view/drs:53262215>

日本京都大学图书馆-图书在线阅读 URL

<https://rmda.kulib.kyoto-u.ac.jp/item/rb00024956>

日本京都大学人文科学研究所-图书在线阅读 URL

<http://kanji.zinbun.kyoto-u.ac.jp/db-machine/toho/ShiSanJingZhuShu/html/A002menu.html>

美国国会图书馆 - （大陆需自备海外 VPN，免 VPN 方法在后文批量文件下载有提及。）

<https://www.loc.gov/item/2014514163/>

普林斯顿大学图书馆 - 图书在线阅读 URL

<https://catalog.princeton.edu/catalog/9940468523506421>

日本国立国会图书馆 - 部分图书在线阅读 URL（其它的可以手动打印下载）

<https://dl.ndl.go.jp/info:ndljp/pid/8929985>

中国台北图书馆古典与特藏文献 - （白天很慢，可夜间或清晨下载）

<https://rbook.ncl.edu.tw/NCLSearch/Search/SearchDetail?item=422a7598bd0046aebf2684ae0f945d25fDcyODIz0&image=1&page=&whereString=&sourceWhereString=&SourceID=>

日本 E 国宝 - 画册在线阅读 URL（部分单图有误，暂未修复）

https://emuseum.nich.go.jp/detail?content_base_id=100168&content_part_id=009&languageId=zh&webView=

日本宫内厅书陵部 - 图书在线阅读 URL

https://db2.sido.keio.ac.jp/kanseki/T_bib_frame.php?id=006754

日本东京大学东洋文化研究所 汉籍善本 - 图书在线阅读 URL

http://shanben.ioc.u-tokyo.ac.jp/main_p.php?nu=C5613401&order=rn_no&no=00870

中国香港中文大学图书馆 - 图书在线阅读 URL（需自备 VPN，从海外访问）

<https://repository.lib.cuhk.edu.hk/sc/item/cuhk-412225#page/1/mode/2up>

牛津大学博德利图书馆 - 图书在线阅读 URL

<https://digital.bodleian.ox.ac.uk/objects/310cb04e-6bce-44e3-85b5-03417c9644a8/>

日本国立公文书馆（内库文库） - 图书在线阅读 URL

<https://www.digital.archives.go.jp/DAS/meta/listPhoto?LANG=default&ID=F1000000000000095447&ID=&NO=&TYPE=>

日本早稻田大学图书馆 - 图书在线阅读 URL

https://archive.wul.waseda.ac.jp/kosho/ri08/ri08_01899/

日本东洋文库（丝绸之路项目） - 图书在线阅读 URL

<http://dsr.nii.ac.jp/toyobunko/XI-6-A-16/V-1/>

韩国国家图书馆（特殊情况，详见独立 PDF 说明文档）

<http://lod.nl.go.kr/page/CNTS-00076977176>

新日本古典籍综合数据库（详见独立 PDF 说明文档）

<https://kotenseki.nijl.ac.jp/biblio/100270332/viewer/1>

<https://kotenseki.nijl.ac.jp/biblio/100270332>

德国柏林图书馆 URL

<https://digital.staatsbibliothek-berlin.de/werkansicht?PPN=PPN3343671770>

https://digital.staatsbibliothek-berlin.de/werkansicht?PPN=PPN3343671770&PHYSID=PHYS_0001

英国图书馆 URL（只生成 dezoomify-rs.urls 文件，生成后，请双击它下载）

http://www.bl.uk/manuscripts/Viewer.aspx?ref=or_6814!1_fs001r

中国香港科技大学图书馆 URL

<https://lbezone.ust.hk/bib/b1129168>

中国台北故宫博物院-善本古籍 URL（详见独立 PDF 说明文档）

日本国立历史民俗博物馆

单册 URL:

<https://khirin-a.rekihaku.ac.jp/sohanshiki/h-172-1>

<https://khirin-a.rekihaku.ac.jp/sohanshiki/h-173-1>

多册 URL:

使用和“批量下载”相同格式，但是无需修改 config.ini 中配置。

如：第 1-9 册，第 10-90 册。用圆括号包围数字。

[https://khirin-a.rekihaku.ac.jp/sohanshiki/h-172-\(1-90\)](https://khirin-a.rekihaku.ac.jp/sohanshiki/h-172-(1-90))

[https://khirin-a.rekihaku.ac.jp/sohankanjo/h-173-\(1-61\)](https://khirin-a.rekihaku.ac.jp/sohankanjo/h-173-(1-61))

日本本市立米泽图书馆

https://www.library.yonezawa.yamagata.jp/dg/AA001_view.html

https://www.library.yonezawa.yamagata.jp/dg/AA002_view.html

日本庆应义塾大学图书馆

<https://dcollections.lib.keio.ac.jp/ja/kanseki/110x-24-1>

日本关西大学图书馆

<https://www.iiif.ku-orcas.kansai-u.ac.jp/books/210185040#?page=1>

IIIF 自动检测下载

IIIF 是一种业界标准，凡使用此标准的网站，都有提供 **IIIF Manifest** 链接，如哈佛大学、牛津大学等图书馆。适用性更广，理论上所有支持 IIIF 的图书馆都可以下载。

使用方法:

启用自动检测功能, 需要在 config.ini 中找到 AutoDetect = 0 改为 2, 保存文件。

```
2= IIIF manifest.json 自动检测下载图片
```

```
AutoDetect = 2
```

设置为 2 以后, 就关闭了内置支持的二十多个图书馆。用完以后, 不要忘记再改为 0。

复制包含 **IIIF Manifest** 链接的网页 URL, 或者直接复制 **IIIF Manifest** 的 URL, 粘贴到 urls.txt 中保存, 运行 bookget 即可自动检测识别。

如果检测成功, 你会看到 **【Auto Detect】**, 会在下一行出现 xxxx.json 的提示。

```
Auto Detect 0001 https://dcollections.lib.keio.ac.jp/ja/kanseki/110x-24-1
Get 0001 https://dcollections.lib.keio.ac.jp/sites/default/files/iiif/KAN/110X-24-1/manifest.json
A total of 46 pages.
```

附: 已通过测试的 URL 如下

第一种: **IIIF Manifest** 的 URL (推荐手动找 URL)

<https://iiif.lib.harvard.edu/manifests/drs:53262215>
<https://digicoll.lib.berkeley.edu/nanna/iiif/91514/manifest>
<https://khirin-a.rekihaku.ac.jp/iiif/rekihaku/H-173-1/manifest.json>
https://khirin-a.rekihaku.ac.jp/manifests/sohan_shiki/H-172-01.json
<https://dcollections.lib.keio.ac.jp/sites/default/files/iiif/KAN/110X-24-1/manifest.json>
<https://iiif.bodleian.ox.ac.uk/iiif/manifest/310cb04e-6bce-44e3-85b5-03417c9644a8.json>

第二种: 包含 **IIIF Manifest** 链接的网页 URL

<https://dcollections.lib.keio.ac.jp/ja/kanseki/110x-24-1>
<https://khirin-a.rekihaku.ac.jp/sohanshiki/h-172-1>
<https://khirin-a.rekihaku.ac.jp/sohankanjo/h-173-1>

通用批量文件下载

因考虑到 bookget 不可能支持无穷数量的网站, 特别提供通用批量下载功能。当然, 这个功能在很多下载工具中都有了, bookget 只是提供自动生成 0001/0002 这样的顺序下载, 以保证批量下载时文件名不乱。

启用批量下载功能, 需要在 config.ini 中找到 AutoDetect = 0 改为 1, 保存文件。

```
1=通用批量下载 (类似 IDM、迅雷);
```

```
AutoDetect = 1
```

例如：

第 1 页网址是 <https://lbezone.ust.hk/obj/6/o/b1129168/ebook/pg00001.jpg>

第 2 页网址是 <https://lbezone.ust.hk/obj/6/o/b1129168/ebook/pg00002.jpg>

.....

第 84 页网址是 <https://lbezone.ust.hk/obj/6/o/b1129168/ebook/pg00084.jpg>

那么，你可以在【urls.txt】文件中填写以下 URL，即可下载全部 84 页。

[https://lbezone.ust.hk/obj/6/o/b1129168/ebook/pg000\(01-84\).jpg](https://lbezone.ust.hk/obj/6/o/b1129168/ebook/pg000(01-84).jpg)

注解：(01-100)、(1-100)、(001-100)支持三种通配符写法。

如果你想下载第 31-40 页，可以使用以下 URL

[https://lbezone.ust.hk/obj/6/o/b1129168/ebook/pg000\(31-40\).jpg](https://lbezone.ust.hk/obj/6/o/b1129168/ebook/pg000(31-40).jpg)

如果你只想下载第 1 页，可以使用以下 URL：

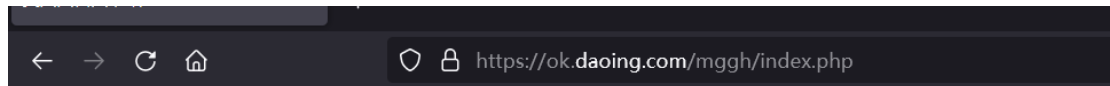
<https://lbezone.ust.hk/obj/6/o/b1129168/ebook/pg00001.jpg>

运行 bookget 后，会在 Downloads 目录下生成 f97611dc51431c96505f9aa399fe29d7 这样的文件夹名称，这是为保证批量 URL 时避免错乱而做的处理。

又或者，美国国会在中国大陆，免 VPN 的方式。

可以借助【书格】提供的在线工具，解析 jp2 图片网址，保存为 urls.txt

<https://ok.daoing.com/mggh/index.php>



以下表单链接为必填

书籍页面网址:

请填写书籍总页面的网址（不是分册的网址）可自动识别的链接格式可以是（类似）：
<https://www.loc.gov/item/2012402424/>

已获取页面链接数: 78

图片类型:

不使用CDN链接 ☒

如果在国内下载，因为国会图书馆默认使用了CloudFlare的CDN会导致下载失败，**建议勾选**：不使用CDN链

获取

当前已获取的是：书籍链接

累计已获取 2 册，共78 / 78页（文件格式：jp2）

如果斜杠（/）前后两个数字不相等，表示获取文件可能不完整。

成功：已获取完整数据

```
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/00001a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/01b03a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/03b04a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/04b05a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/05b06a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/06b07a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/07b08a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/08b09a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/09b10a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/10b11a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/11b12a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/12b13a.jp2
http://140.147.239.202/service/asian/lcnclscd/2012402424/1A001/13b14a.jp2
```

（本文完）