

---

# **sqlitebiter Documentation**

*Release 0.11.0*

**Tsuyoshi Hombashi**

**Aug 19, 2017**



---

# Table of Contents

---

<b>1</b>	<b>sqlitebiter</b>	<b>1</b>
1.1	Summary . . . . .	1
1.2	Features . . . . .	1
<b>2</b>	<b>Installation</b>	<b>3</b>
2.1	Install via pip (recommended) . . . . .	3
2.1.1	Mandatory Python package dependencies . . . . .	3
2.1.2	Google Sheets dependencies . . . . .	3
2.1.3	Test dependencies . . . . .	4
2.1.4	Misc . . . . .	4
2.2	Installing executable files in Windows . . . . .	4
<b>3</b>	<b>Usage</b>	<b>5</b>
3.1	sqlitebiter command help . . . . .	5
3.2	Create a SQLite database from CSV/Excel/JSON/SQLite/etc. files . . . . .	6
3.2.1	Examples . . . . .	6
3.2.2	sqlitebiter file subcommand help . . . . .	7
3.2.3	Supported data formats . . . . .	7
3.2.4	Table naming convention . . . . .	7
3.3	Create a SQLite database from URL . . . . .	8
3.3.1	Example . . . . .	8
3.3.2	sqlitebiter url subcommand help . . . . .	10
3.4	Create a SQLite database from Google Sheets . . . . .	10
3.4.1	Requirements . . . . .	10
3.4.2	Example . . . . .	11
3.4.3	sqlitebiter gs subcommand help . . . . .	11
<b>4</b>	<b>Indices and tables</b>	<b>13</b>
<b>5</b>	<b>Links</b>	<b>15</b>
<b>6</b>	<b>Indices and tables</b>	<b>17</b>



## Summary

A CLI tool to convert CSV/Excel/HTML/JSON/LTSV/Markdown/SQLite/TSV/Google-Sheets to a SQLite database file.

## Features

- **Create a SQLite database file from:**
  - **File(s):**
    - \* CSV
    - \* Microsoft Excel™
    - \* HTML
    - \* JSON
    - \* Labeled Tab-separated Values (LTSV)
    - \* Markdown
    - \* SQLite
    - \* Tab separated values (TSV)
  - Google Sheets
  - URL (scrape data from web pages)
- Multi-byte character support



### Install via pip (recommended)

`sqlitebiter` can be installed via `pip` (Python package manager).

```
pip install sqlitebiter
```

Python package dependencies are as follows.

### Mandatory Python package dependencies

Following mandatory Python packages are automatically installed during `sqlitebiter` installation process:

- `appconfigpy`
- `click`
- `logbook`
- `path.py`
- `pytablereader`
- `SimpleSQLite`
- `sqliteschema`
- `typepy`

### Google Sheets dependencies

Following Python packages are required to [manual installation](#) when you use Google Sheets feature:

- `oauth2client`
- `pyOpenSSL`

## Test dependencies

- pytest
- pytest-runner
- tox
- XlsxWriter

## Misc

- lxml (Faster HTML convert if installed)

## Installing executable files in Windows

1. Navigate to <https://github.com/thombashi/sqlitebiter/releases>
2. Download the latest version of the `sqlitebiter_win_x64.zip`
3. Unzip the file
4. Execute `sqlitebiter.exe` in either Command Prompt or PowerShell

```
>cd sqlitebiter_win_x64
>sqlitebiter.exe -h
Usage: sqlitebiter.exe [OPTIONS] COMMAND [ARGS]...

Options:
  --version          Show the version and exit.
  -a, --append       append table(s) to existing database.
  -i, --index TEXT   comma separated attribute names to create indices.
  -v, --verbose
  --debug           for debug print.
  --quiet          suppress execution log messages.
  -h, --help        Show this message and exit.

Commands:
  configure  Configure the following application settings:...
  file       Convert tabular data within...
  gs         Convert a spreadsheet in Google Sheets to a...
  url        Scrape tabular data from a URL and convert...
```



## sqlitebiter command help

sqlitebiter has following subcommands:

- **file:** Convert tabular data within CSV/Excel/HTML/JSON/LTSV/Markdown/SQLite/TSV file(s) to a SQLite database file.
  - *Create a SQLite database from CSV/Excel/JSON/SQLite/etc. files*
- **url:** Scrape tabular data from a URL and convert data to a SQLite database file.
  - *Create a SQLite database from URL*
- **gs:** Convert a spreadsheet in Google Sheets to a SQLite database file.
  - *Create a SQLite database from Google Sheets*
- **configure:** Configure the application settings

```
Usage: sqlitebiter [OPTIONS] COMMAND [ARGS]...

Options:
  --version          Show the version and exit.
  -a, --append      append table(s) to existing database.
  -i, --index TEXT  comma separated attribute names to create indices.
  -v, --verbose
  --debug           for debug print.
  --quiet           suppress execution log messages.
  -h, --help       Show this message and exit.

Commands:
  configure  Configure the following application settings:...
  file      Convert tabular data within...
  gs       Convert a spreadsheet in Google Sheets to a...
  url     Scrape tabular data from a URL and convert...
```

## Create a SQLite database from CSV/Excel/JSON/SQLite/etc. files

sqlitebiter file is a subcommand to convert tabular data file(s) to a SQLite database file.

### Examples

Using wildcard to convert multiple files:

#### Example

```
$ ls
sample_data.csv  sample_data.xlsx  sample_data_multi.json  sample_data_
↳single.json
$ sqlitebiter file * -o sample.sqlite
[INFO] sqlitebiter file: convert 'sample_data.csv' to 'sample_data' table
[INFO] sqlitebiter file: convert 'sample_data.xlsx' to 'samplesheet1'
↳table
[INFO] sqlitebiter file: convert 'sample_data.xlsx' to 'samplesheet3'
↳table
[INFO] sqlitebiter file: convert 'sample_data_multi.json' to 'table_b'
↳table
[INFO] sqlitebiter file: convert 'sample_data_multi.json' to 'table_a'
↳table
[INFO] sqlitebiter file: convert 'sample_data_single.json' to 'sample_
↳data_single' table
```

#### Output

```
$ sqlite3 sample.sqlite
sqlite> .schema
CREATE TABLE 'sample_data' ("attr_a" INTEGER, "attr_b" REAL, "attr_c"
↳TEXT);
CREATE TABLE 'samplesheet1' (a INTEGER, b REAL, c TEXT);
CREATE TABLE 'samplesheet3' (aa INTEGER, ab TEXT, ac TEXT);
CREATE TABLE 'table_b' (a INTEGER, b REAL);
CREATE TABLE 'table_a' ("attr_a" INTEGER, "attr_b" REAL, "attr_c" TEXT);
CREATE TABLE 'sample_data_single' ("attr_a" INTEGER, "attr_b" REAL,
↳"attr_c" TEXT);
```

Designate multiple file path to convert:

#### Example

```
$ sqlitebiter file hoge.csv sample_excel.xlsx
[INFO] sqlitebiter file: convert 'hoge.csv' to 'hoge' table
[INFO] sqlitebiter file: convert 'sample_excel.xlsx' to 'sheet3' table
[INFO] sqlitebiter file: convert 'sample_excel.xlsx' to 'sheet1' table
```

---

#### Note:

- Available JSON Schema is limited. Acceptable format described in [here](#)
  - Wildcard characters cannot use in Windows environments
-

## sqlitebiter file subcommand help

```
Usage: sqlitebiter file [OPTIONS] [FILES]...
```

Convert tabular data within CSV/Excel/HTML/JSON/LTSV/Markdown/SQLite/TSV file(s) to a SQLite database file.

Options:

```
-o, --output-path PATH  Output path of the SQLite database file. Defaults to
                        'out.sqlite'.
--encoding ENCODING    Encoding to load files. Defaults to 'utf-8'.
-h, --help             Show this message and exit.
```

## Supported data formats

Following table shows that the supported data formats:

Table 3.1: Available data formats

Format	File Extension	Remarks
CSV	.csv	
Excel	.xlsx/.xls	Create table for each sheet in the Excel workbook.
HTML	.html/.htm	Scrape tabular data from <table> tags in the HTML file. And create table for each <table> tag data.
JSON	.json	
LTSV	.ltsv	
Markdown	.md	Extract tabular data in the Markdown file. And create a table for each <table> tabular data.
SQLite	.sqlite/.sqlite3	
TSV	.tsv	

## Table naming convention

Table name automatically decided as follows for each format:

Format	Table Name
CSV	<filename>
Excel	<Sheet name>
HTML	<p>&lt;title&gt;_&lt;key&gt;. &lt;title&gt; replaced with a title tag.</p> <p>&lt;key&gt; replaced with: <b>(1)</b> id attribute of the table tag.</p> <p><b>(2)</b> unique string if id attribute not present in the table tag.</p>
JSON	
LTSV	<filename>
Markdown	<filename>
TSV	<filename>

• **Common behavior**

– **<filename> replaced with filename of converting file (without extensions)**

\* e.g. If the input file name is `sample.csv`, <filename> is `sample`

If a created table name already exists in the database, the behavior differs depending on the existing table (after this referred to as A) and create table (after this referred to as B) structure:

1. **A and B has the same table name and table structure**

- Append creating table data to the existing table data

2. **A and B has the same table name, but different table structure**

- `sqlitebiter` try to create unique table name for B by appending suffix id number

## Create a SQLite database from URL

`sqlitebiter url` is a subcommand to fetch table data from the Internet and convert to a SQLite database file.

### Example

Following is an example that convert HTML table tags within a web page to SQLite tables.

**Example**

```
$ sqlitebiter -v url "https://en.wikipedia.org/wiki/Comparison_of_
↳firewalls"
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html1_
↳(Firewall TEXT, License TEXT, Costandusagelimits TEXT, OS TEXT)' table
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html2_
↳(Firewall TEXT, License TEXT, Cost TEXT, OS TEXT)' table
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html3_
↳(CanTarget TEXT,
↳Changingdefaultpolicytoacceptrejectbyissuingasinglerule TEXT,
↳IPdestinationaddresses TEXT, IPsourceaddresses TEXT,
↳TCPUDPdestinationports TEXT, TCPUDPsourceports TEXT,
↳EthernetMACdestinationaddress TEXT, EthernetMACsourceaddress TEXT,
↳Inboundfirewallingress TEXT, Outboundfirewallegress TEXT)' table
```

```
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html4_
↳(Can TEXT, workatOSILayer4statefulfirewall TEXT,
↳workatOSILayer7applicationinspection TEXT,
↳ChangeTTLtransparenttotraceroute TEXT, ConfigureREJECTwithanswer TEXT,
↳DMZdemilitarizedzoneallowsforsingleseveralhostsnottobefirewalled TEXT,
↳Filteraccordingtotimeofday TEXT, RedirectTCPUDPportsportforwarding_
↳TEXT, RedirectIPaddressesforwarding TEXT,
↳FilteraccordingtoUserAuthorization TEXT, TrafficratelimitQoS TEXT,
↳Tarpit TEXT, Log TEXT)' table
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html5_
↳(Features TEXT, ConfigurationGUItextorbothmodes TEXT,
↳RemoteAccessWebHTTPTelnetSSHRDPSerialCOMRS232 TEXT,
↳Changeruleswithoutrequiringrestart TEXT,
↳Abilitytocentrallymanageallfirewallstogether TEXT)' table
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html6_
↳(Features TEXT,
↳Modularitysupportstthirdpartymodulestoextendfunctionality TEXT, IPS :
↳Intrusion prevention system] TEXT, OpenSourceLicense TEXT, supports_
↳IPv6 ?] TEXT, ClassHomeProfessional TEXT,
↳OperatingSystemsonwhichitruns TEXT)' table
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html7_
↳(Can TEXT, NAT44staticdynamicwoportsPAT TEXT, NAT64NPTv6 TEXT,
↳IDSIntrusionDetectionSystem TEXT, VPNVirtualPrivateNetwork TEXT,
↳AVAntiVirus TEXT, Sniffer TEXT, Profileselection TEXT)' table
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html9_
↳(A TEXT, B TEXT)' table
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html10_
↳(A TEXT, B TEXT)' table
[INFO] sqlitebiter url: convert 'https://en.wikipedia.org/wiki/
↳Comparison_of_firewalls' to 'Comparison_of_firewalls_Wikipedia_html11_
↳(A TEXT, B TEXT)' table
```

## Output

```
sqlite> .schema
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html1' (Firewall TEXT,
↳License TEXT, Costandusagelimits TEXT, OS TEXT);
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html2' (Firewall TEXT,
↳License TEXT, Cost TEXT, OS TEXT);
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html3' (CanTarget TEXT,
↳Changingdefaultpolicytoacceptrejectbyissuingasinglerule TEXT,
↳IPdestinationaddresses TEXT, IPsourceaddresses TEXT,
↳TCPUDPdestinationports TEXT, TCPUDPsourceports TEXT,
↳EthernetMACdestinationaddress TEXT, EthernetMACsourceaddress TEXT,
↳Inboundfirewallingress TEXT, Outboundfirewallegress TEXT);
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html4' (Can TEXT,
↳[workatOSILayer4statefulfirewall] TEXT,
↳[workatOSILayer7applicationinspection] TEXT,
↳ChangeTTLtransparenttotraceroute TEXT, ConfigureREJECTwithanswer TEXT,
↳DMZdemilitarizedzoneallowsforsingleseveralhostsnottobefirewalled TEXT,
↳Filteraccordingtotimeofday TEXT, RedirectTCPUDPportsportforwarding_
↳TEXT, RedirectIPaddressesforwarding TEXT,
↳FilteraccordingtoUserAuthorization TEXT, TrafficratelimitQoS TEXT,
↳Tarpit TEXT, Log TEXT);
```

```
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html5' (Features TEXT,
↳ ConfigurationGUItextorbothmodes TEXT,
↳ [RemoteAccessWebHTTPTelnetSSHRDPSerialCOMRS232] TEXT,
↳ Changeruleswithoutrequiringrestart TEXT,
↳ Abilitytocentrallymanageallfirewallstogether TEXT);
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html6' (Features TEXT,
↳ Modularitysupportsthirdpartymodulestoextendfunctionality TEXT, [IPS :
↳ Intrusion prevention system] TEXT, OpenSourceLicense TEXT, [supports
↳ IPv6 ?] TEXT, ClassHomeProfessional TEXT,
↳ OperatingSystemsonwhichitruns TEXT);
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html7' (Can TEXT,
↳ [NAT44staticdynamicwoportsPAT] TEXT, [NAT64NPTv6] TEXT,
↳ IDSIntrusionDetectionSystem TEXT, VPNVirtualPrivateNetwork TEXT,
↳ AVAntiVirus TEXT, Sniffer TEXT, Profileselection TEXT);
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html9' (A TEXT, B TEXT);
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html10' (A TEXT, B TEXT);
CREATE TABLE 'Comparison_of_firewalls_Wikipedia_html11' (A TEXT, B TEXT);
```

## sqlitebiter url subcommand help

Usage: sqlitebiter url [OPTIONS] URL

Scrape tabular data **from a URL and** convert data to a SQLite database file.

Options:

```
-f, --format [csv|excel|html|json|ltsv|markdown|mediawiki|sqlite|tsv]
Data format to loading (defaults to html).
-o, --output-path PATH
Output path of the SQLite database file.
Defaults to 'out.sqlite'.
-e, --encoding ENCODING
HTML page read encoding. Defaults to utf-8.
-p, --proxy PROXY
Specify a proxy in the form
[user:passwd@]proxy.server:port.
-h, --help
Show this message and exit.
```

## Create a SQLite database from Google Sheets

sqlitebiter gs is a subcommand to convert [Google Sheets](#) to a SQLite database file.

### Requirements

Following python packages are required to use Google Sheets feature.

- [oauth2client](#)
- [pyOpenSSL](#)

Dependency Python package installation:

```
pip install oauth2client
pip install pyopenssl
```

## Example

```
$ sqlitebiter gs credentials-xxxxxxxxxxxx.json samplebook -o sample.sqlite
[INFO] sqlitebiter gs: convert 'google sheets' to 'sheet3' table
[INFO] sqlitebiter gs: convert 'google sheets' to 'sheet1' table
$ sqlite3 sample.sqlite
$ sqlite> .schema
CREATE TABLE 'sheet3' (a INTEGER, b REAL, c TEXT);
CREATE TABLE 'sheet1' (a INTEGER, b REAL, c TEXT);
```

## sqlitebiter gs subcommand help

```
Usage: sqlitebiter gs [OPTIONS] CREDENTIALS TITLE
```

Convert a spreadsheet **in** Google Sheets to a SQLite database file.

CREDENTIALS: OAuth2 Google credentials file. TITLE: Title of the Google Sheets to convert.

Options:

```
-o, --output-path PATH  Output path of the SQLite database file. Defaults to
                        'out.sqlite'.
-h, --help              Show this message and exit.
```





## CHAPTER 4

---

### Indices and tables

---

- `genindex`



## CHAPTER 5

---

### Links

---

- [pip](#): A tool for installing python packages
- [GitHub repository](#)
- [Issue tracker](#)
- [PyPI](#)



## CHAPTER 6

---

### Indices and tables

---

- `genindex`