hc Documentation

Release 0.1.1

Robin Schneider

Apr 28, 2020

Contents

1 hc introduction 3

1.1 Repositories . . . 3
1.2 Documentation . . . 3
1.3 Authors . . . 3
1.4 License . . . 4

2 Installation 5

2.1 Latest release . . . 5
2.2 Development version . . . 5

3 CLI interface 7

3.1 Positional Arguments . . . 7
3.2 Named Arguments . . . 7

4 History 9

5 Design principles 11

6 hc package 13

6.1 Submodules . . . 13
6.2 hc.cli module . . . 13
6.3 hc.datatypes module . . . 13
6.4 hc.defaults module . . . 13
6.5 hc.helpers module . . . 14
6.6 hc.opening_hours_js module . . . 14
6.7 hc.schulferien_org module . . . 14
6.8 hc.yaml module . . . 14
6.9 Module contents . . . 15

7 Contributing and issue reporting 17

8 FAQ 19

8.1 Am I allowed to use the data the script gathers? . . . 19

9 Changelog 21

9.1 hc master - unreleased . . . 21
9.2 hc v0.1.1 - 2017-06-04 . . . 21
9.3 hc v0.1.0 - 2017-03-02 . . . 21

Python Module Index 23

Index 25

Contents:

CHAPTER 1

hc introduction

(GitLab CI) - (Travis CI) - -

hc – holiday converter

Supports the following inputs and outputs:

- Input: German school holidays (HTML tables) from http://www.schulferien.org/
- Input and output: (School) holidays format used by opening_hours.js

Supports the following data serialization languages:

- YAML
- JSON

1.1 Repositories

- GitLab (primary repo with issue tracker)
- GitHub (mirror)

1.2 Documentation

- Read the Docs

1.3 Authors

- Robin Schneider

1.4 License

GNU Affero General Public License v3 (AGPL-3.0)

CHAPTER 2

Installation

2.1 Latest release

You can install hc by invoking the following commands:

```
gpg --recv-keys 'C505 B5C9 3B0D B3D3 38A1  B600 5FE9 2C12 EE88 E1F0'
mkdir --parent /tmp/hc && cd /tmp/hc
wget -r -nd -l 1 https://pypi.python.org/pypi/hc --accept-regex '^https://(test)?
↪pypi\.python\.org/packages/.*\.whl.*'
current_release="$(find . -type f -name '*.whl' | sort | tail -n 1)"
gpg --verify "${current_release}.asc" "${current_release}" && pip3 install --upgrade "
↪${current_release}"
```

Refer to Verifying PyPI and Conda Packages for more details. Note that this might pull down dependencies in an unauthenticated way! You might want to install the dependencies yourself beforehand.

Or if you feel lazy and agree that pip/issues/1035 should be fixed you can also install hc like this:

```
pip3 install hc
```

2.2 Development version

If you want to be more on the bleeding edge of hc development consider cloning the git repository and installing from it:

```
gpg --recv-keys 'EF96 BC32 AC57 CFC7 2DF0  1D8C 489A 4D5E C353 C98A'
git clone --recursive https://gitlab.com/ypid/hc.git
cd hc && git verify-commit HEAD
echo 'Check if the HEAD commit has a good signature and only proceed in that case!' &&
↪ read -r fnord
echo 'Then chose one of the commands below to install hc and its dependencies:'
pip3 install --upgrade .
```

(continues on next page)

(continued from previous page)

```
./setup.py develop --user
./setup.py install --user
./setup.py install
```

This will also get you the cache which is tracked in git as well to do integration testing over the whole dataset. Please be sure to use the cache by symlinking it to your user cache directory. The following should do the trick:

```
hc_cache="$(python3 -c 'from appdirs import user_cache_dir; print(user_cache_dir("hc
↪"))')"
ln -sT "$PWD/tests/cache/" "$hc_cache"
```

CHAPTER 3

CLI interface

Holiday converter tool

```
usage: hc [-h] [-V] [-d] [-v] [-q] [-n] [-c CACHE_DIR] [-i INPUT_FILE]
          [-f {schulferien_html}] [-F FROM_DATE] [-T TO_DATE] [-u]
          [-t {yaml,json}] [-s {opening_hours.js}] [-D]
          output-file
```

3.1 Positional Arguments

output-file Where to write the output file. '-' will write to STDOUT.

3.2 Named Arguments

-V, --version show program's version number and exit

-d, --debug Write debugging and higher to STDOUT|STDERR.

-v, --verbose Write information and higher to STDOUT|STDERR.

-q, --quiet, --silent Only write errors and higher to STDOUT|STDERR.

-n, --no-cache Do not cache intermediary files.

Default: True

-c, --cache-dir Cache directory, defaults to the default cache directory of your operating system.

-i, --input-file File path to the input file to process. '-' will read from STDIN.

-f, --input-format, --from Possible choices: schulferien_html

Format of the input file. Default: "schulferien_html".

Default: "schulferien_html"

-F, --from-date Process date range starting at given RFC 3339 date. Default: Current year and month “”2020-04”“.

Default: “2020-04”

-T, --to-date Process date range ending at given RFC 3339 date. Default: One year in the further “”2021-04”“.

Default: “2021-04”

-u, --update-output Update the output file instead of constructing it from scratch. Implementation incomplete.

Default: False

-t, --output-format, --to Possible choices: yaml, json

Format of the output file. Default: “yaml”.

Default: “yaml”

-s, --output-structure Possible choices: opening_hours.js

Structure of the output file. Default: “opening_hours.js”.

Default: “opening_hours.js”

-D, --dry-run Don’t write output.

Default: False

CHAPTER 4

History

This tool was created because Germany as of 2017 seems to be unable/unwilling to provide school holidays or holidays in general in a machine readable format. There are sites like http://www.schulferien.org/ which do a really good job in getting the data anyway through various sources and "providing them". Back in 2013, everything was great and schulferien.org just provided all iCal files they had for the school holidays of the current year and the following years as far as they are defined by the German Kultusministerkonferenz. A Perl Script has been used to parse all the iCal files and convert them (ref: convert_ical_to_json). Unfortunately, those days are over and after checking out all the available sources the least bad one was to go ahead and parse the HTML table of schulferien.org because the HTML version still provides all data. schulferien.org was contacted before to find a better solution but none has been found. One concern from schulferien.org are the use of (faulty) scripts which put load on their servers. It is therefore one key design goal of this tool to make the fewest requests to external resources possible and use extensive caching. This has been implemented see *Design principles*.

Refer to this issue for more details.

CHAPTER 5

Design principles

- Generic

  When you look around on the Internet, you find hundreds of public and/or school holiday APIs, libraries, websites providing HTML calenders, iCALs, PDFs, at least for Germany. Most of these have some artificial kind of limitation or restriction. This is an attempt to harvest holidays (which are generally not copyright protected), convert them and provide them without any limitations.

  The available tools where found useless for the use case of bundling holiday definitions for opening_hours.js which is why this tool has been written.

- Free Software

  All sources are provided under the GNU Affero General Public License v3 (AGPL-3.0). Resources such as holiday data is released under a Creative Commons Zero v1.0 Universal. Enjoy.

- Idempotent.

  The program can be run against it's output and should not make any changes to it. This property is checked by integration testing.

- Caching

  Make the fewest requests to external resources possible and use extensive caching. The cache is provided as separate git repository (hc-tests-cache) to also make use of the cache during CI testing which is done against a support matrix of Python versions and environments and therefore runs in parallel a number of times for each commit.

- Expendable

  Convert from anything to anything using a common internal data structure.

CHAPTER 6

hc package

6.1 Submodules

6.2 hc.cli module

Command line interface of hc

`hc.cli.`**`main`**`()`

6.3 hc.datatypes module

Data types definition

`class` `hc.datatypes.`**`MonthDayList`**

Bases: `list`

`class` `hc.datatypes.`**`PhData`**

Bases: `list`

`hc.datatypes.`**`fix_data_types`**`(`*dataset*`)`

`hc.datatypes.`**`fix_ph_data`**`(`*dataset*`)`

6.4 hc.defaults module

hc defaults

6.5 hc.helpers module

hc helpers

hc.helpers.**get_date_from_relative_month**(*relative_month*)

hc.helpers.**get_month_number**(*month_name*)

hc.helpers.**get_relative_month**(*date*)

6.6 hc.opening_hours_js module

OpenStreetMap opening_hours.js format. Refer to https://github.com/opening-hours/opening_hours.js/blob/master/holidays/README.md for the "spec".

class hc.opening_hours_js.**OpeningHoursJS**(*defs=None*)

Bases: object

FIRST_LEVEL_SORTING = {'PH': '20', 'SH': '30', '_nominatim_url': '10'}

SH_DATA_SORTING = {'name': '0'}

get_school_holidays(*out=None*)

read(*in_defs*)

static update_sh_format(*sh_data*)

hc.opening_hours_js.**find_ind**(*lst*, *key*, *value*)

6.7 hc.schulferien_org module

schulferien.org interface

class hc.schulferien_org.**SchulferienOrg**(*defs=None*, *cache=True*, *cache_dir=None*)

Bases: object

get_month_dataset(*date*)

get_school_holidays(*from_date*, *to_date*)

6.8 hc.yaml module

YAML representation

class hc.yaml.**PrettyHolidayYAMLDumper**(*stream*, *default_style=None*, *default_flow_style=None*, *canonical=None*, *indent=None*, *width=None*, *allow_unicode=None*, *line_break=None*, *encoding=None*, *explicit_start=None*, *explicit_end=None*, *version=None*, *tags=None*, *block_seq_indent=None*, *top_level_colon_align=None*, *prefix_colon=None*)

Bases: ruamel.yaml.dumper.RoundTripDumper

YAML dumper optimized human readability of the holiday format.

represent_dict(*data*)

write out tag if saved on loading

represent_list(*data*)

yaml_representers = {None: <unbound method SafeRepresenter.represent_undefined>, <typ

hc.yaml.**dump_holidays_as_yaml**(*unserialized_data*, *add_vspacing=True*)

hc.yaml.**get_clean_yaml**(*serialized_data*, *add_vspacing=False*)

6.9 Module contents

Holiday converter tool

CHAPTER 7

Contributing and issue reporting

You can contribute and report issues in the usual way as documented by GitHub. Unit and integration tests can be run locally and are automatically run in CI. Acceptable contributions need to pass all of them.

If you found a security vulnerability that might put users at risk please send your report/patch to ypid@riseup.net. Please consider using OpenPGP to encrypt your email.

CHAPTER 8

FAQ

8.1 Am I allowed to use the data the script gathers?

The author hopes so but keep in mind that he is not a lawyer. As this is about German law, the § 5 UrhG should apply after which content like school holidays in Germany are not copyright protected. The official source is https://www.kmk.org/service/ferien.html.

CHAPTER 9

Changelog

This project adheres to Semantic Versioning and human-readable changelog.

9.1 hc master - unreleased

9.2 hc v0.1.1 - 2017-06-04

9.2.1 Changed

- Pin `ruamel.yaml` on 0.14.X for now as there will be API changes in 0.15+. [ruamel]

9.3 hc v0.1.0 - 2017-03-02

9.3.1 Added

- Initial coding and design. [ypid]

Python Module Index

h

hc, 15

hc.cli, 13

hc.datatypes, 13

hc.defaults, 13

hc.helpers, 14

hc.opening_hours_js, 14

hc.schulferien_org, 14

hc.yaml, 14

Index

D

dump_holidays_as_yaml() (*in module hc.yaml*), 15

F

find_ind() (*in module hc.opening_hours_js*), 14
FIRST_LEVEL_SORTING (*hc.opening_hours_js.OpeningHoursJS attribute*), 14
fix_data_types() (*in module hc.datatypes*), 13
fix_ph_data() (*in module hc.datatypes*), 13

G

get_clean_yaml() (*in module hc.yaml*), 15
get_date_from_relative_month() (*in module hc.helpers*), 14
get_month_dataset() (*hc.schulferien_org.SchulferienOrg method*), 14
get_month_number() (*in module hc.helpers*), 14
get_relative_month() (*in module hc.helpers*), 14
get_school_holidays() (*hc.opening_hours_js.OpeningHoursJS method*), 14
get_school_holidays() (*hc.schulferien_org.SchulferienOrg method*), 14

H

hc (*module*), 15
hc.cli (*module*), 13
hc.datatypes (*module*), 13
hc.defaults (*module*), 13
hc.helpers (*module*), 14
hc.opening_hours_js (*module*), 14
hc.schulferien_org (*module*), 14
hc.yaml (*module*), 14

M

main() (*in module hc.cli*), 13
MonthDayList (*class in hc.datatypes*), 13

O

OpeningHoursJS (*class in hc.opening_hours_js*), 14

P

PhData (*class in hc.datatypes*), 13
PrettyHolidayYAMLDumper (*class in hc.yaml*), 14

R

read() (*hc.opening_hours_js.OpeningHoursJS method*), 14
represent_dict() (*hc.yaml.PrettyHolidayYAMLDumper method*), 14
represent_list() (*hc.yaml.PrettyHolidayYAMLDumper method*), 15

S

SchulferienOrg (*class in hc.schulferien_org*), 14
SH_DATA_SORTING (*hc.opening_hours_js.OpeningHoursJS attribute*), 14

U

update_sh_format() (*hc.opening_hours_js.OpeningHoursJS static method*), 14

Y

yaml_representers (*hc.yaml.PrettyHolidayYAMLDumper attribute*), 15