
Catafolk

Release 0

Oct 04, 2021

Contents:

1	Dataset	1
2	File	3
3	Index	5
4	Sources	7
5	Transformer	9
6	Operations	11
7	Utilities	13
8	Indices and tables	15
	Python Module Index	17
	Index	19

CHAPTER 1

Dataset

CHAPTER 2

File

Most of the metadata included in Catafolk comes from music files, such as kern files. This metadata is extracted by the *File* classes: there is one for every file format. Currently supported are *kern* and *xml*.

Files can be instantiated using `get_file()` which automatically determines the file format based on the extension.

```
>>> data_dir = 'tests/datasets/bronson-child-ballads/data'
>>> file = get_file(f'{data_dir}/child01.krn')
>>> file
<KernFile name=child01 format=kern>
>>> file.metadata.keys()
dict_keys(['ONM', 'PPG', 'AMT', 'YOR', 'OVM', 'PPE', 'PSR', 'PSP', 'PSD', 'ENC', 'EMD', 'EEV'])
>>> file.metadata['ONM']
'Child Ballad No. 1, Tune No. 1'
```

Note that metadata is collected only once and then stored internally. In the code above, we accessed `file.metadata` twice, but the file was only read out once. You can reset the stored metadata using `File.reset()`.

class `catafolk.file.File` (*filepath*, *encoding*='utf-8')

Bases: object

The base class for music files.

checksum

An md5 checksum of the file.

```
>>> file = get_file('tests/datasets/bronson-child-ballads/data/child01.krn')
>>> file.checksum
'350fc2b9839d7d7669d83f77efdc03c2'
```

format = None

The file format, e.g. 'kern' or 'xml'

metadata

A dictionary with metadata collected from the file. Which properties are included in the metadata

depends completely on the contents of the file. Properties like the path are *not* included in the metadata.

Metadata is collected only once, and then stored in the class. You can reset the file using `reset()`.

relpath (*root*)

Return the relative path with respect to some root directory.

```
>>> file = get_file('tests/datasets/bronson-child-ballads/data/child01.krn')
>>> file.path
'tests/datasets/bronson-child-ballads/data/child01.krn'
>>> file.relpath(root='tests/datasets/bronson-child-ballads')
'data/child01.krn'
```

Parameters *root* (*str*) – The root directory

Returns The relative path

Return type *str*

reset ()

Reset the file instance: remove stored metadata and checksum. This only affects the class instance, the file itself is (as always) left untouched. This might be useful if the file has changed and you want to refresh the metadata.

class catafolk.file.**KernFile** (*filepath*, *encoding*='utf-8')

Bases: `catafolk.file.File`

A class for loading ****kern**** files.

When extracting metadata, the class looks for kern *reference records* of the form `!!![key]: [value]`. These are collected in a dictionary. If a key is encountered multiple times, all values are collected in a list.

format = 'kern'

class catafolk.file.**XMLFile** (*filepath*, *encoding*='utf-8')

Bases: `catafolk.file.File`

format = 'xml'

`catafolk.file.get_file` (*filepath*, *format*='infer', ****kwargs**)

File factory that returns a File instance of the right type

Parameters

- **filepath** (*str*) – Path to the file
- **format** (*str*, *optional*) – File format. Currently supported are *kern* and *xml*. When *format*='infer' the format is inferred from the extension. By default 'infer'

Returns The File instance

Return type *File*

CHAPTER 3

Index

CHAPTER 4

Sources

CHAPTER 5

Transformer

CHAPTER 6

Operations

Some useful utilities

`catafolk.utils.checksum_iterable(iterable)`
Return an md5 checksum of an iterable

```
>>> checksum_iterable(['hello', 'world'])  
'fc5e038d38a57032085441e7fe7010b0'
```

Parameters `iterable` (*iterable*) – An iterable of strings or byte-likes

Returns An md5 checksum of the iterable

Return type `str`

`catafolk.utils.file_checksum(path: str)`
Return an md5 checksum of a given file

```
>>> path = 'tests/datasets/bronson-child-ballads/data/child01.krn'  
>>> file_checksum(path)  
'350fc2b9839d7d7669d83f77efdc03c2'
```

Parameters `path` (*str*) – The filepath

Returns A md5 checksum of the file

Return type `str`

CHAPTER 8

Indices and tables

- `genindex`
- `modindex`
- `search`

C

`catafolk.file`, 3
`catafolk.utils`, 13

C

`catafolk.file` (*module*), 3
`catafolk.utils` (*module*), 13
`checksum` (*catafolk.file.File* attribute), 3
`checksum_iterable()` (*in module catafolk.utils*),
13

F

`File` (*class in catafolk.file*), 3
`file_checksum()` (*in module catafolk.utils*), 13
`format` (*catafolk.file.File* attribute), 3
`format` (*catafolk.file.KernFile* attribute), 4
`format` (*catafolk.file.XMLFile* attribute), 4

G

`get_file()` (*in module catafolk.file*), 4

K

`KernFile` (*class in catafolk.file*), 4

M

`metadata` (*catafolk.file.File* attribute), 3

R

`relpath()` (*catafolk.file.File* method), 4
`reset()` (*catafolk.file.File* method), 4

X

`XMLFile` (*class in catafolk.file*), 4