
xmlcoll Documentation

Release 1.4.0

Clemson University

Apr 25, 2023

CONTENTS

1	Contents	1
2	Indices and tables	7
	Python Module Index	9
	Index	11

CONTENTS

1.1 Overview

xmlcoll is a python package for working with collections of items. The items have heterogeneous data stored as *properties* in a dictionary with keys given by a name and optional tags. The package API has routines to write data to and retrieve data from XML and to validate that XML against a schema.

1.1.1 Installation

Install from PyPI with pip by typing in your favorite terminal:

```
$ pip install xmlcoll
```

1.1.2 Usage

To get familiar with xmlcoll, please see the tutorial Jupyter [notebook](#).

1.1.3 Authors

- Bradley S. Meyer <mbradle@g.clemson.edu>

1.1.4 Documentation

The project documentation is available at <https://xmlcoll.readthedocs.io>.

1.1.5 Contribute

- Issue Tracker: <https://github.com/mbradle/xmlcoll/issues/>
- Source Code: <https://github.com/mbradle/xmlcoll/>

1.1.6 License

The project is licensed under the GNU Public License v3 (or later).

1.2 Changelog

All notable changes to this project will be documented in this file. This project adheres to [Semantic Versioning](#).

1.2.1 Version 1.4.0

New:

- Methods to store and retrieve collections as pandas dataframes are now part of the API.

Fix:

- Some documentation links and typos have been fixed.
- An extraneous import has been removed.

1.2.2 Version 1.3.0

New:

- A link to the readthedocs page has been added for more visibility from PYPI and GitHub.

Fix:

- An extraneous and non-functioning search link has been removed.

1.2.3 Version 1.2.0

Internal:

- Multiple xml catalogs are now allowed within python.

1.2.4 Version 1.1.0

Internal:

- The validator now uses XML Catalog. The appropriate schema files are included in the distribution.

1.2.5 Version 1.0.1

Internal:

- The PYPI classifiers have been updated.

1.2.6 Version 1.0.0

New:

- Initial release

1.3 Documentation

1.3.1 xmlcoll package

A package of python routines to work with data in XML format of samples.

xmlcoll.base module

class `xmlcoll.base.Properties`

Bases: `object`

A class for storing and retrieving properties.

get_properties()

Method to retrieve the properties.

Returns:

`dict`: The dictionary of current properties.

update_properties(properties)

Method to update the properties.

Args:

`properties (dict)`: A dictionary of properties. New properties are added. Old properties are updated. The keys for the dictionary entries are *name* and up to five optional tags labeled *tag1*, *tag2*, ..., *tag5*.

Returns:

On successful return, the properties have been updated.

xmlcoll.coll module

class `xmlcoll.coll.Collection(items=None)`

Bases: `Properties`

A class for storing and retrieving data about data items.

Args:

`items (list, optional)`: A list of individual `xmlcoll.coll.Item` objects.

add_item(item)

Method to add a item to a collection.

Args:

`item (xmlcoll.coll.Item)` The item to be added.

Return:

On successful return, the item has been added.

get()

Method to retrieve the item collection as a dictionary.

Returns:

`dict`: A dictionary of the items.

get_dataframe(*index_label='name', tag_delimiter='_'*)

Method to retrieve the collection data as a pandas dataframe.

Args:

`index_label` (`str`, optional): Index label for the dataframe.

`tag_delimiter` (`str`, optional): Delimiter used to separate tags in combined column names.

Returns:

`pandas.DataFrame`: A pandas dataframe containing the collection data. Columns are labeled by a string formed by concatenating property names and tags separated by the chosen delimiter.

remove_item(*item*)

Method to remove an item from a item collection.

Args:

`item` (`xmlcoll.coll.Item`) The item to be removed.

Return:

On successful return, the item has been removed.

update_from_dataframe(*df, index_label='name', tag_delimiter='_'*)

Method to update collection data from a pandas dataframe.

Args:

`df` (`pandas.DataFrame`): The pandas dataframe.

`index_label` (`str`, optional): Index label for the data frame.

`tag_delimiter` (`str`, optional): Delimiter used to separate tags in combined column names.

Returns:

On successful return, the collection has been updated with the data in the data frame.

update_from_xml(*file, xpath=""*)

Method to update a item collection from an XML file.

Args:

`file` (`str`) The name of the XML file from which to update.

`xpath` (`str`, optional): XPath expression to select items. Defaults to all items.

Returns:

On successful return, the item collection has been updated.

validate(*file*)

Method to validate a collection XML file.

Args:

`file` (`str`) The name of the XML file to validate.

Returns:

An error message if invalid and nothing if valid.

write_to_xml(*file*, *pretty_print=True*)

Method to write the collection to XML.

Args:

file (**str**) The output file name.

pretty_print (**bool**, optional): If set to True, routine outputs the xml in nice indented format.

Return:

On successful return, the item collection data have been written to the XML output file.

class xmlcoll.coll.**Item**(*name*, *properties=None*)

Bases: *Properties*

A class for storing and retrieving data about a data item.

Args:

properties (**dict**, optional): A dictionary of properties.

get_name()

Method to retrieve name of item.

Return:

str: The name of the item.

1.4 Acknowledgments

Original development of this package was funded by NASA's [Emerging Worlds Program](#).

INDICES AND TABLES

- genindex
- modindex

PYTHON MODULE INDEX

X

`xmlcoll`, 3
`xmlcoll.base`, 3
`xmlcoll.coll`, 3

A

`add_item()` (*xmlcoll.coll.Collection method*), 3

C

`Collection` (*class in xmlcoll.coll*), 3

G

`get()` (*xmlcoll.coll.Collection method*), 3
`get_dataframe()` (*xmlcoll.coll.Collection method*), 4
`get_name()` (*xmlcoll.coll.Item method*), 5
`get_properties()` (*xmlcoll.base.Properties method*), 3

I

`Item` (*class in xmlcoll.coll*), 5

M

module
 xmlcoll, 3
 xmlcoll.base, 3
 xmlcoll.coll, 3

P

`Properties` (*class in xmlcoll.base*), 3

R

`remove_item()` (*xmlcoll.coll.Collection method*), 4

U

`update_from_dataframe()` (*xmlcoll.coll.Collection method*), 4
`update_from_xml()` (*xmlcoll.coll.Collection method*), 4
`update_properties()` (*xmlcoll.base.Properties method*), 3

V

`validate()` (*xmlcoll.coll.Collection method*), 4

W

`write_to_xml()` (*xmlcoll.coll.Collection method*), 4

X

`xmlcoll`
 module, 3
`xmlcoll.base`
 module, 3
`xmlcoll.coll`
 module, 3