pathsummary

Release 0.1.0a6

David Scheliga

Sep 23, 2021

CONTENTS

Path summary is a helper module to summarize the content of a folder based on the mimetypes (file types and sub types).

CHAPTER

ONE

INSTALLATION

Install the latest release from pip.

```
$ pip install pathsummary
```

1.1 API reference

pathsummary	
pathsummary.summarize_folder_files(folder_path)	Summarizes the content of the *folder path*.
pathsummary.count_type_occurrences(summary_table)	Counts the occurences of file- and subtypes.

1.1.1 pathsummary

Module Attributes

FILETYPE_IMAGE	Deprecated; use FileTypes.IMAGE instead.

Functions

count_type_occurrences(summary_table)	Counts the occurences of file- and subtypes.
summarize_folder_files(folder_path)	Summarizes the content of the *folder path*.

Classes

FileMetaData(filepath, mimetype)	
FileTypes()	
PathSummary(summary_table)	

1.1.2 summarize_folder_files

pathsummary.**summarize_folder_files**(*folder_path: Union[str, pathlib.Path]*) → pandas.core.frame.DataFrame

Summarizes the content of the *folder path*.

Parameters **folder_path** – Path which content is summarized.

Returns DataFrame

Examples

```
>>> from pathsummary import summarize_folder_files
>>> from doctestprinter import print_pandas
>>> sample_summary = summarize_folder_files("tests/resources/images")
>>> print_pandas(sample_summary)
                          file_rootpath      filename  encoding
filetype  subtype
   image      bmp  tests/resources/images   image00.bmp      None
              bmp  tests/resources/images   image01.BMP      None
             jpeg  tests/resources/images  image00.jpeg      None
             jpeg  tests/resources/images   image00.jpg      None
             jpeg  tests/resources/images  image01.JPEG      None
             jpeg  tests/resources/images   image01.JPG      None
              png  tests/resources/images   image00.png      None
              png  tests/resources/images   image01.PNG      None
             tiff  tests/resources/images   image00.tif      None
             tiff  tests/resources/images  image00.tiff      None
             tiff  tests/resources/images   image01.TIF      None
             tiff  tests/resources/images  image01.TIFF      None
```

```
>>> counts_of_equal_types = sample_summary.index.value_counts()
>>> counts_of_equal_types.sort_index(inplace=True)
>>> print_pandas(counts_of_equal_types, formats="{:<}#{:>}")
('image', 'bmp')   2
('image', 'jpeg')  4
('image', 'png')   2
('image', 'tiff')  4
```

1.1.3 count_type_occurrences

pathsummary.**count_type_occurrences**(*summary_table: pandas.core.frame.DataFrame*) → pandas.core.frame.DataFrame

Counts the occurences of file- and subtypes.

Parameters **summary_table** – The summarized content of a path.

Returns DataFrame

Examples

```
>>> from pathsummary import summarize_folder_files, count_type_occurrences
>>> from doctestprinter import print_pandas
>>> sample_summary = summarize_folder_files("tests/resources/images")
>>> sample_counts = count_type_occurrences(sample_summary)
>>> print_pandas(sample_counts, formats="{}#{:}{:.0%}")
                   count  ratio
filetype  subtype
   image      bmp      2    17%
             jpeg      4    33%
              png      2    17%
             tiff      4    33%
```

1.1.4 pathsummary.PathSummary

pathsummary.PathSummary.empty	States if the path summary is empty.
pathsummary.PathSummary.summarize_path(...)	Summarizes the content of the given *folder path*.
pathsummary.PathSummary.from_file_paths(...)	Creates a *path summary* based on the given file paths.
pathsummary.PathSummary.iter_by_filetype(...)	Iterates the filepaths releated to the *file type* and optionally *sub type*.
pathsummary.PathSummary.count_type_occurrences()	Counts the occurences of file- and subtypes.
pathsummary.PathSummary.table	

pathsummary.PathSummary.empty

property PathSummary.**empty: bool**

States if the path summary is empty.

Returns bool

pathsummary.PathSummary.summarize_path

static PathSummary.**summarize_path**(*folder_path: Union[str, pathlib.Path]*) → pathsummary.PathSummary

Summarizes the content of the given *folder path*.

Parameters **folder_path** – The folder path which content is summarized.

Returns PathSummary

Examples

```
>>> from pathsummary import PathSummary
>>> from doctestprinter import print_pandas
>>> from pathlib import Path
>>> test_path = "tests/resources/images"
>>> sample_summary = PathSummary.summarize_path(test_path)
>>> print_pandas(sample_summary.table)
                          file_rootpath      filename  encoding
filetype  subtype
   image      bmp  tests/resources/images   image00.bmp      None
              bmp  tests/resources/images   image01.BMP      None
             jpeg  tests/resources/images  image00.jpeg      None
             jpeg  tests/resources/images   image00.jpg      None
             jpeg  tests/resources/images  image01.JPEG      None
             jpeg  tests/resources/images   image01.JPG      None
              png  tests/resources/images   image00.png      None
              png  tests/resources/images   image01.PNG      None
             tiff  tests/resources/images   image00.tif      None
             tiff  tests/resources/images  image00.tiff      None
             tiff  tests/resources/images   image01.TIF      None
             tiff  tests/resources/images  image01.TIFF      None
```

pathsummary.PathSummary.from_file_paths

static PathSummary.**from_file_paths**(*file_paths: List[pathlib.Path]*) → pathsummary.PathSummary

Creates a *path summary* based on the given file paths.

Parameters **file_paths** – File paths of the path summary.

Returns PathSummary

Examples

```
>>> from pathsummary import summarize_folder_files
>>> from doctestprinter import print_pandas
>>> from pathlib import Path
>>> test_file_paths = [
...     Path("tests/resources/images/image01.BMP"),
...     Path("tests/resources/images/image00.jpeg"),
...     Path("tests/resources/images/image00.jpg"),
...     Path("tests/resources/images/image01.JPEG"),
...     Path("tests/resources/images/image01.JPG"),
... ]
>>> sample_summary = PathSummary.from_file_paths(test_file_paths)
>>> print_pandas(sample_summary.table)
                          file_rootpath      filename  encoding
filetype  subtype
   image      bmp  tests/resources/images   image01.BMP      None
             jpeg  tests/resources/images  image00.jpeg      None
             jpeg  tests/resources/images   image00.jpg      None
```

(continues on next page)

(continued from previous page)

```
            jpeg  tests/resources/images  image01.JPEG      None
            jpeg  tests/resources/images   image01.JPG      None
```

```
>>> test_file_paths = [
...     Path("tests/resources/images/image01.BMP"),
... ]
>>> sample_summary = PathSummary.from_file_paths(test_file_paths)
>>> print_pandas(sample_summary.table)
                          file_rootpath     filename  encoding
filetype  subtype
   image      bmp  tests/resources/images  image01.BMP      None
```

pathsummary.PathSummary.iter_by_filetype

PathSummary.**iter_by_filetype**(*file_type: str*, *sub_type: Optional[str] = None*) → pathlib.Path

Iterates the filepaths releated to the *file type* and optionally *sub type*.

Parameters

- **file_type** – Files based on the file type are yielded.
- **sub_type** – Optional sub type to narrow the selection down.

Returns Path

Examples

```
>>> from pathsummary import summarize_folder_files, FileTypes
>>> from doctestprinter import print_pandas, doctest_iter_print
>>> from pathlib import Path
>>> test_file_paths = [
...     Path("tests/resources/images/image01.BMP"),
...     Path("tests/resources/images/image00.jpeg"),
...     Path("tests/resources/images/image00.jpg"),
...     Path("tests/resources/images/image01.JPEG"),
...     Path("tests/resources/images/image01.JPG"),
... ]
>>> sample_summary = PathSummary.from_file_paths(test_file_paths)
>>> doctest_iter_print(sample_summary.iter_by_filetype(FileTypes.IMAGE))
tests/resources/images/image01.BMP
tests/resources/images/image00.jpeg
tests/resources/images/image00.jpg
tests/resources/images/image01.JPEG
tests/resources/images/image01.JPG
```

```
>>> doctest_iter_print(
...     sample_summary.iter_by_filetype(FileTypes.IMAGE, "bmp")
... )
tests/resources/images/image01.BMP
```

```
>>> test_file_paths = [Path("a/path/text.csv")]
>>> sample_summary = PathSummary.from_file_paths(test_file_paths)
>>> doctest_iter_print(sample_summary.iter_by_filetype("text", "csv"))
a/path/text.csv
```

pathsummary.PathSummary.count_type_occurrences

PathSummary.**count_type_occurrences**()

Counts the occurences of file- and subtypes.

Returns DataFrame

Examples

```
>>> from pathsummary import PathSummary
>>> from doctestprinter import print_pandas
>>> sample_summary = PathSummary.summarize_path("tests/resources/images")
>>> sample_counts = sample_summary.count_type_occurrences()
>>> print_pandas(sample_counts, formats="{}#{:}{:.0%}")
                   count  ratio
filetype  subtype
   image      bmp      2    17%
             jpeg      4    33%
              png      2    17%
             tiff      4    33%
```

pathsummary.PathSummary.table

property PathSummary.**table: pandas.core.frame.DataFrame**

CHAPTER

TWO

INDICES AND TABLES

- genindex

PYTHON MODULE INDEX

INDEX