

Module **search**

Functions

def AggregateDataFrameSearch(searchTerms: list, dataframes: list, colName: str, uniqueOnly=False)

Description:

Used to aggregate the results of searching multiple (i.e File and Folder) dataframes that share a common column name.

Args:

```
searchTerms : list[str]
    #: The terms to individually search for
dataframes : list[pd.DataFrame]
    #: The dataframes to search within
colName : str
    #: The name of the dataframe column to search within
```

Kwargs:

```
uniqueOnly : bool : default = False
    #: If true only count each search match once, otherwise sort matches by # of hits
```

Outputs:

```
allResults : list
    #: List of dataframe indices (absolute paths) corresponding to the search results
termResultsDict : dict
    #: Dict containing each search term -> matching results list
resultTermsDict : dict
    #: Dict containing each result -> matching search terms
```

def ExactNameSearch(name: str, tree: src.fileorganiser.general.Tree, ignoreCase=True, phrase=False)

Description:

Search the dataframe names column for exact matches. If phrase=True this matches any string that contains the passed name. Otherwise only full words are matched if phrase=False.

Args:

```
name : str
    #: The name to search for
tree : Tree
    #: The tree class containing the dataframes
```

Kwargs:

```
ignoreCase : bool : default = True
    #: Case sensitive/insensitive search
phrase : bool : default = True
    #: Match whole word only or any string containing phrase
```

Outputs:

```
output : tuple(list, list)
#: Tuple containing the list of paths (df index) of the found search items
in the following form: (folderResults, fileResults)
```

Notes:

The `Search_MultipleNamesInclusive` method produces results that (inclusively) match against a list of names using repeated application of this method for each name.

```
def FindSimilarSearchTerms(searchTerm: str)
```

Description:

Find the unique lemmatized synonymns of the given search term.

Args:

```
searchTerm : str
#: The term to find synonyms of
```

Outputs:

```
synonyms : list[str]
#: List of synonymns found
```

```
def FuzzyColumnSearch(searchTerm: str, df: pandas.core.frame.DataFrame, colName: str, threshold=0.9)
```

Description:

Search the specified dataframe column for fuzzy matches above the threshold. Uses the `rapidfuzz` process function to determine the similarity between strings.

Args:

```
searchTerm : str
#: The term to search for
df : pd.DataFrame
#: The dataframe to search within
colName : str
#: The name of the dataframe column to search within
```

Kwargs:

```
threshold : float : default = 0.9
#: Minimum similarity threshold
```

Outputs:

```
resultIndices : list
#: List of dataframe indices corresponding to the search results
```

```
def MultiTermFuzzySearch(searchTerms: list, df: pandas.core.frame.DataFrame, colName: str, uniqueOnly=False)
```

Description:

Aggregates the results of multiple fuzzy searches within the same dataframe column. If `uniqueOnly=False` the results are ordered by the number of occurrences (if multiple search terms matched to the same item).

Args:

```
searchTerms : list[str]
    #: The terms to individually search for
df : pd.DataFrame
    #: The dataframe to search within
colName : str
    #: The name of the dataframe column to search within
```

Kwargs:

```
uniqueOnly : bool : default = False
    #: If true only count each search match once, otherwise sort matches by # of hits
```

Outputs:

```
allResults : list
    #: List of dataframe indices corresponding to the search results
termResultsDict : dict
    #: Dict containing each search term -> matching results list
resultTermsDict : dict
    #: Dict containing each result -> matching search terms
```

```
def RankedResultsByOccurences(allResults: list, termResultsDict: dict, resultTermDict: collections.defaultdict)
```

Description:

Used to sort the search results based on the number of matched results. Results with multiple matches (eg from several of the included search terms) are ranked higher.

Args:

```
allResults : list
    #: List of dataframe indices (absolute paths) corresponding to the search results
termResultsDict : dict
    #: Dict containing each search term -> matching results list
resultTermsDict : dict
    #: Dict containing each result -> matching search terms
```

Outputs:

```
sortedResults : list
    #: List of dataframe indices (absolute paths) corresponding to the sorted search results
```

Index

Super-module

`fileorganiser`

Functions

`AggregateDataframeSearch`

`ExactNameSearch`

`FindSimilarSearchTerms`

`FuzzyColumnSearch`

`MultiTermFuzzySearch`

`RankedResultsByOccurences`