MyProject

1

Generated by Doxygen 1.8.11

Contents

1 Namespace Index **1**

1.1 Namespace List 1

2 File Index **3**

2.1 File List 3

3 Namespace Documentation **5**

3.1 diff Namespace Reference 5

3.1.1 Function Documentation 6

3.1.1.1 calculate_distance(document_one_dictionary, document_two_dictionary) 6

3.1.1.2 create_dictionary(file_path) 6

3.1.1.3 ensure_dir(file_path) 6

3.1.1.4 format_text(text_line) 6

3.1.1.5 inner_product(document_one_dictionary, document_two_dictionary) 6

3.1.1.6 remove_delimiters(text_line) 7

3.1.1.7 text_formatting(my_dictionary) 7

3.1.2 Variable Documentation 7

3.1.2.1 delimiters 7

3.1.2.2 document_dictionary 7

3.1.2.3 document_one_dictionary 7

3.1.2.4 document_two_dictionary 7

3.1.2.5 document_word_count 7

3.1.2.6 end 7

3.1.2.7 file 8

3.1.2.8 file_index_one . . . 8
3.1.2.9 folder_path . . . 8
3.1.2.10 idf_key . . . 8
3.1.2.11 output_folder_path . . . 8
3.1.2.12 result . . . 8
3.1.2.13 result_dictionary . . . 8
3.1.2.14 result_file . . . 8
3.1.2.15 result_file_name . . . 8
3.1.2.16 sorted_result . . . 8
3.1.2.17 start . . . 8
3.1.2.18 temp_dictionary . . . 8
3.1.2.19 tf_key . . . 8

4 File Documentation **9**

4.1 diff.py File Reference . . . 9

Index **11**

Chapter 1

Namespace Index

1.1 Namespace List

Here is a list of all namespaces with brief descriptions:

diff 5

Chapter 2

File Index

2.1 File List

Here is a list of all files with brief descriptions:

diff.py 9

Chapter 3

Namespace Documentation

3.1 diff Namespace Reference

Functions

- def remove_delimiters (text_line)
- def format_text (text_line)
- def create_dictionary (file_path)
- def inner_product (document_one_dictionary, document_two_dictionary)
- def calculate_distance (document_one_dictionary, document_two_dictionary)
- def text_formatting (my_dictionary)
- def ensure_dir (file_path)

Variables

- string delimiters = "\'\"/<?>,.:∗-+\\=‘∼!@#^&()_ ;"
- dictionary document_word_count = {}
- start = time.clock()
- folder_path = raw_input("Please input the path to the folder")
- dictionary document_dictionary = {}
- output_folder_path = raw_input("Please input the path to the output folder")
- dictionary result_dictionary = {}
- dictionary temp_dictionary = {}
- idf_key = math.log(len(document_dictionary) / (document_word_count[key] ∗ 1.0))
- tf_key = document_dictionary[filename][key]
- document_one_dictionary = document_dictionary[file_index_one]
- document_two_dictionary = document_dictionary[file_index_two]
- result = calculate_distance(document_one_dictionary, document_two_dictionary)
- file_index_one
- end
- sorted_result = sorted(result_dictionary.items(), key=operator.itemgetter(1))
- result_file_name = raw_input("Please enter the name of file where you wish to save the result.")
- result_file = open(result_file_name ,'w')
- file

3.1.1 Function Documentation

3.1.1.1 def diff.calculate_distance (*document_one_dictionary, document_two_dictionary*)

```
Function name : inner_product
Input arguments :
    1. dictionary document_one_dictionary: The dictionary corresponding to document one.
    2. dictionary document_two_dictionary: The dictionary corresponding to document two.
Purpose : To find the distance between two documents
Return Value : The cosine distance of two vectors
```

Definition at line 98 of file diff.py.

3.1.1.2 def diff.create_dictionary (*file_path*)

```
Function name : create_dictionary
Input arguments :
    1. string file_path : The path to the file to be read.
Purpose : To create the word frequency table/ vector
Return Value : The dictionary containing the word frequenct vector
```

Definition at line 50 of file diff.py.

3.1.1.3 def diff.ensure_dir (*file_path*)

```
Function name : inner_product
Input arguments :
    1. string file_path : The path of the folder.
Purpose : To create a directory if it does not exists.
```

Definition at line 130 of file diff.py.

3.1.1.4 def diff.format_text (*text_line*)

```
Function name : format_text
Input arguments :
    1. string text_line : Piece of text
Purpose : To format the text by removing delimiters and trailing whitespaces
Return Value : The formatted string.
```

Definition at line 37 of file diff.py.

3.1.1.5 def diff.inner_product (*document_one_dictionary, document_two_dictionary*)

```
Function name : inner_product
Input arguments :
    1. dictionary document_one_dictionary : The dictionary corresponding to document one.
    2. dictionary document_two_dictionary : The dictionary corresponding to document two.
Purpose : To find the inner product of two vectors
Return Value : The inner product of two vectors
```

Definition at line 82 of file diff.py.

3.1.1.6 def diff.remove_delimiters (*text_line*)

```
Function name : remove_delimiters
Input arguments :
    1. string text_line : Piece of text
Purpose : To remove unnecessary characters
Return Value : The string after removing characters.
```

Definition at line 24 of file diff.py.

3.1.1.7 def diff.text_formatting (*my_dictionary*)

```
Function name : inner_product
Input arguments :
    1. dictionary my_dictionary : The dictionary to be formatted
Purpose : To remove the words which have very less importance to the meaning.
Return Value : The modified dictionary.
```

Definition at line 113 of file diff.py.

3.1.2 Variable Documentation

3.1.2.1 string diff.delimiters = "\'\"/<?>,.:*-+\\=‘∼!@#^&()_ ;"

Definition at line 17 of file diff.py.

3.1.2.2 dictionary diff.document_dictionary = {}

Definition at line 144 of file diff.py.

3.1.2.3 diff.document_one_dictionary = document_dictionary[file_index_one]

Definition at line 169 of file diff.py.

3.1.2.4 diff.document_two_dictionary = document_dictionary[file_index_two]

Definition at line 171 of file diff.py.

3.1.2.5 dictionary diff.document_word_count = {}

Definition at line 21 of file diff.py.

3.1.2.6 diff.end

Definition at line 175 of file diff.py.

3.1.2.7 diff.file

Definition at line 188 of file diff.py.

3.1.2.8 diff.file_index_one

Definition at line 175 of file diff.py.

3.1.2.9 diff.folder_path = raw_input("Please input the path to the folder")

Definition at line 143 of file diff.py.

3.1.2.10 diff.idf_key = math.log(len(document_dictionary) / (document_word_count[key] * 1.0))

Definition at line 159 of file diff.py.

3.1.2.11 diff.output_folder_path = raw_input("Please input the path to the output folder")

Definition at line 145 of file diff.py.

3.1.2.12 diff.result = calculate_distance(document_one_dictionary, document_two_dictionary)

Definition at line 173 of file diff.py.

3.1.2.13 dictionary diff.result_dictionary = {}

Definition at line 149 of file diff.py.

3.1.2.14 diff.result_file = open(result_file_name ,'w')

Definition at line 186 of file diff.py.

3.1.2.15 diff.result_file_name = raw_input("Please enter the name of file where you wish to save the result.")

Definition at line 185 of file diff.py.

3.1.2.16 diff.sorted_result = sorted(result_dictionary.items(), key=operator.itemgetter(1))

Definition at line 181 of file diff.py.

3.1.2.17 diff.start = time.clock()

Definition at line 141 of file diff.py.

3.1.2.18 dictionary diff.temp_dictionary = {}

Definition at line 155 of file diff.py.

3.1.2.19 diff.tf_key = document_dictionary[filename][key]

Definition at line 160 of file diff.py.

Chapter 4

File Documentation

4.1 diff.py File Reference

Namespaces

- diff

Functions

- def diff.remove_delimiters (text_line)
- def diff.format_text (text_line)
- def diff.create_dictionary (file_path)
- def diff.inner_product (document_one_dictionary, document_two_dictionary)
- def diff.calculate_distance (document_one_dictionary, document_two_dictionary)
- def diff.text_formatting (my_dictionary)
- def diff.ensure_dir (file_path)

Variables

- string diff.delimiters = "\'\"/<?>,.:*-+\\=‘~!@#^&()_ ;"
- dictionary diff.document_word_count = {}
- diff.start = time.clock()
- diff.folder_path = raw_input("Please input the path to the folder")
- dictionary diff.document_dictionary = {}
- diff.output_folder_path = raw_input("Please input the path to the output folder")
- dictionary diff.result_dictionary = {}
- dictionary diff.temp_dictionary = {}
- diff.idf_key = math.log(len(document_dictionary) / (document_word_count[key] * 1.0))
- diff.tf_key = document_dictionary[filename][key]
- diff.document_one_dictionary = document_dictionary[file_index_one]
- diff.document_two_dictionary = document_dictionary[file_index_two]
- diff.result = calculate_distance(document_one_dictionary, document_two_dictionary)
- diff.file_index_one
- diff.end
- diff.sorted_result = sorted(result_dictionary.items(), key=operator.itemgetter(1))
- diff.result_file_name = raw_input("Please enter the name of file where you wish to save the result.")
- diff.result_file = open(result_file_name ,'w')
- diff.file

Index

calculate_distance
    diff, 6
create_dictionary
    diff, 6

delimiters
    diff, 7
diff, 5
    calculate_distance, 6
    create_dictionary, 6
    delimiters, 7
    document_dictionary, 7
    document_one_dictionary, 7
    document_two_dictionary, 7
    document_word_count, 7
    end, 7
    ensure_dir, 6
    file, 7
    file_index_one, 8
    folder_path, 8
    format_text, 6
    idf_key, 8
    inner_product, 6
    output_folder_path, 8
    remove_delimiters, 6
    result, 8
    result_dictionary, 8
    result_file, 8
    result_file_name, 8
    sorted_result, 8
    start, 8
    temp_dictionary, 8
    text_formatting, 7
    tf_key, 8
diff.py, 9
document_dictionary
    diff, 7
document_one_dictionary
    diff, 7
document_two_dictionary
    diff, 7
document_word_count
    diff, 7

end
    diff, 7
ensure_dir
    diff, 6

file
    diff, 7
file_index_one
    diff, 8
folder_path
    diff, 8
format_text
    diff, 6

idf_key
    diff, 8
inner_product
    diff, 6

output_folder_path
    diff, 8

remove_delimiters
    diff, 6
result
    diff, 8
result_dictionary
    diff, 8
result_file
    diff, 8
result_file_name
    diff, 8

sorted_result
    diff, 8
start
    diff, 8

temp_dictionary
    diff, 8
text_formatting
    diff, 7
tf_key
    diff, 8