MyProject

1

Generated by Doxygen 1.8.11

Contents

Chapter 1

Namespace Index

1.1 Namespace List

Here is a list of all namespaces with brief descriptions:

Chapter 2

File Index

2.1 File List

Here is a list of all files with brief descriptions:

Chapter 3

Namespace Documentation

3.1 create_dict Namespace Reference

Functions

- def remove_delimiters (text_line)
- def format_text (text_line)
- def create_dictionary (file_path)
- def text_formatting (my_dictionary)

Variables

- string delimiters = "\'\"/<?>,.:∗-+\\=‘∼!@#^&()_ ;{}[ ]|"
- dictionary document_word_count = {}
- folder_path = raw_input("Please input the path to the folder\t")
- dictionary document_dictionary = {}
- dict_file = open('dict.txt','w')
- key
- end
- file

3.1.1 Function Documentation

3.1.1.1 def create_dict.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 create_dict.py.

3.1.1.2 def create_dict.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 create_dict.py.

3.1.1.3 def create_dict.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 create_dict.py.

3.1.1.4 def create_dict.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 79 of file create_dict.py.

3.1.2 Variable Documentation

3.1.2.1 string create_dict.delimiters = "\'\"/<?>,.:∗-+\\=‘~!@#^&()_ ;{}[ ]|"

Definition at line 17 of file create_dict.py.

3.1.2.2 create_dict.dict_file = open('dict.txt','w')

Definition at line 107 of file create_dict.py.

3.1.2.3 dictionary create_dict.document_dictionary = {}

Definition at line 101 of file create_dict.py.

3.1.2.4 dictionary create_dict.document_word_count = {}

Definition at line 21 of file create_dict.py.

3.1.2.5 create_dict.end

Definition at line 109 of file create_dict.py.

3.1.2.6 create_dict.file

Definition at line 109 of file create_dict.py.

3.1.2.7 create_dict.folder_path = raw_input("Please input the path to the folder\t")

Definition at line 100 of file create_dict.py.

3.1.2.8 create_dict.key

Definition at line 109 of file create_dict.py.

3.2 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\t")
- dictionary document_dictionary = {}
- output_folder_path = raw_input("Please input the path to the output folder\t")
- dictionary result_dictionary = {}
- dict_file = open('dict.txt','r')
- text_line = text_line.rstrip()
- text_line_divided = text_line.split(' ')
- dictionary temp_dictionary = {}
- idf_key = math.log(len(document_dictionary) / (document_word_count[key] ∗ 1.0))
- tf_key = document_dictionary[filename][key]
- file_index = os.listdir(folder_path)
- file_index_one = file_index[index_one]
- file_index_two = file_index[index_two]
- 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)
- 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.\t")
- result_file = open(result_file_name ,'w')
- file

3.2.1 Function Documentation

3.2.1.1 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 94 of file diff.py.

3.2.1.2 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.

3.2.1.3 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 126 of file diff.py.

3.2.1.4 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.

3.2.1.5 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 78 of file diff.py.

3.2.1.6 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.

3.2.1.7 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 109 of file diff.py.

3.2.2 Variable Documentation

3.2.2.1 string diff.delimiters = "\'\"/<?>,.:∗-+\\=‘∼!@#^&()_ ;{}[ ]|"

Definition at line 17 of file diff.py.

3.2.2.2 diff.dict_file = open('dict.txt','r')

Definition at line 150 of file diff.py.

3.2.2.3 dictionary diff.document_dictionary = {}

Definition at line 140 of file diff.py.

3.2.2.4 diff.document_one_dictionary = document_dictionary[file_index_one]

Definition at line 178 of file diff.py.

3.2.2.5 diff.document_two_dictionary = document_dictionary[file_index_two]

Definition at line 180 of file diff.py.

3.2.2.6 dictionary diff.document_word_count = {}

Definition at line 21 of file diff.py.

3.2.2.7 diff.end

Definition at line 184 of file diff.py.

3.2.2.8 diff.file

Definition at line 197 of file diff.py.

3.2.2.9 diff.file_index = os.listdir(folder_path)

Definition at line 170 of file diff.py.

3.2.2.10 diff.file_index_one = file_index[index_one]

Definition at line 175 of file diff.py.

3.2.2.11 diff.file_index_two = file_index[index_two]

Definition at line 176 of file diff.py.

3.2.2.12 diff.folder_path = raw_input("Please input the path to the folder\t")

Definition at line 139 of file diff.py.

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

Definition at line 163 of file diff.py.

3.2.2.14 diff.output_folder_path = raw_input("Please input the path to the output folder\t")

Definition at line 141 of file diff.py.

3.2.2.15 diff.result = calculate_distance(document_one_dictionary, document_two_dictionary)

Definition at line 182 of file diff.py.

3.2.2.16 dictionary diff.result_dictionary = {}

Definition at line 145 of file diff.py.

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

Definition at line 195 of file diff.py.

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

Definition at line 194 of file diff.py.

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

Definition at line 190 of file diff.py.

3.2.2.20 diff.start = time.clock()

Definition at line 137 of file diff.py.

3.2.2.21 dictionary diff.temp_dictionary = {}

Definition at line 159 of file diff.py.

3.2.2.22 diff.text_line = text_line.rstrip()

Definition at line 152 of file diff.py.

3.2.2.23 diff.text_line_divided = text_line.split(' ')

Definition at line 153 of file diff.py.

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

Definition at line 164 of file diff.py.

3.3 result Namespace Reference

Functions

- def calculate_nearest (file_name)

Variables

- folder_path = raw_input("Please input the path to the folder\t")
- cur_home_dir = os.getcwd()
- file_name = raw_input("Please input the name of result file.\t")
- result_dict = calculate_nearest(file_name)
- res_out_file = open('eval_res-nearest', 'w')
- key
- end
- file

3.3.1 Function Documentation

3.3.1.1 def result.calculate_nearest (*file_name*)

```
Function name : calculate_nearest
Input arguments :
    1. string file_name : Name of file that conatins result
Purpose : To find the nearest file of each and every file
Return Value : The dictionary containing the nearest file index of each file
```

Definition at line 16 of file result.py.

3.3.2 Variable Documentation

3.3.2.1 result.cur_home_dir = os.getcwd()

Definition at line 49 of file result.py.

3.3.2.2 result.end

Definition at line 59 of file result.py.

3.3.2.3 result.file

Definition at line 59 of file result.py.

3.3.2.4 result.file_name = raw_input("Please input the name of result file.\t")

Definition at line 52 of file result.py.

3.3.2.5 result.folder_path = raw_input("Please input the path to the folder\t")

Definition at line 48 of file result.py.

3.3.2.6 result.key

Definition at line 59 of file result.py.

3.3.2.7 result.res_out_file = open('eval_res-nearest', 'w')

Definition at line 57 of file result.py.

3.3.2.8 result.result_dict = calculate_nearest(file_name)

Definition at line 53 of file result.py.

Chapter 4

File Documentation

4.1 create_dict.py File Reference

Namespaces

- create_dict

Functions

- def create_dict.remove_delimiters (text_line)
- def create_dict.format_text (text_line)
- def create_dict.create_dictionary (file_path)
- def create_dict.text_formatting (my_dictionary)

Variables

- string create_dict.delimiters = "\'\"/<?>,.:∗-+\\=‘∼!@#^&()_ ;{}[ ]|"
- dictionary create_dict.document_word_count = {}
- create_dict.folder_path = raw_input("Please input the path to the folder\t")
- dictionary create_dict.document_dictionary = {}
- create_dict.dict_file = open('dict.txt','w')
- create_dict.key
- create_dict.end
- create_dict.file

4.2 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\t")
- dictionary diff.document_dictionary = {}
- diff.output_folder_path = raw_input("Please input the path to the output folder\t")
- dictionary diff.result_dictionary = {}
- diff.dict_file = open('dict.txt','r')
- diff.text_line = text_line.rstrip()
- diff.text_line_divided = text_line.split(' ')
- 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.file_index = os.listdir(folder_path)
- diff.file_index_one = file_index[index_one]
- diff.file_index_two = file_index[index_two]
- 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.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.\t")
- diff.result_file = open(result_file_name ,'w')
- diff.file

4.3 result.py File Reference

Namespaces

- result

Functions

- def result.calculate_nearest (file_name)

Variables

- result.folder_path = raw_input("Please input the path to the folder\t")
- result.cur_home_dir = os.getcwd()
- result.file_name = raw_input("Please input the name of result file.\t")
- result.result_dict = calculate_nearest(file_name)
- result.res_out_file = open('eval_res-nearest', 'w')
- result.key
- result.end
- result.file

Index