

WTookits 的 API 文档

目录

Algorithm 模块	2
WTookits/src/com/algorithm/TDF.java	2
class IDF	2
WTookits/src/com/algorithm/TF.java	2
class TF	2
WTookits/src/com/TermWeighting.java	3
interface TermWeighting	3
Conf 模块	3
WTookits/src/com/conf/configuration.java	3
class Configuration	3
WTookits/src/com/conf/IDFFactory.java	4
class IDFFactory	4
WTookits/src/com/conf/TFFactory.java	5
class TFFactory	5
WTookits/src/com/conf/TermWeightingFactory.java	5
interface TermWeightingFactory	5
Input 模块	6
WTookits/src/com/input/ UnsupervisedDH.java	6
class UnsupervisedDH	6
WTookits/src/com/input/ SupervisedDH.java	7
class SupervisedDH	7
WTookits/src/com/main/ MainHolder.java	7
class MainHolder	7
WTookits/src/com/main/MainHolderNotInitializedException.java	8
class MainHolderNotInitializedException extends Exception	8
WTookits/src/com/main/ MapComparator.java	8
class MapComparator	8

Algorithm 模块

WTookits/src/com/algorithm/TDF.java

class IDF

public List<List<String>> split(List<Map.Entry<String, String>> data)

该函数是 class IDF 实现接口 TermWeighting 中的 split 函数，用于分词的作用

输入：List<Map.Entry<String, String>>类型的文本数据

输出：List<List<String>>对输入文本分好的词

public List<Map<String, Double>> calculate(List<List<String>> splitedWords)

该函数是 class IDF 实现接口 TermWeighting 中的 calculate 函数，用于计算每个词的权重

输入：List<List<String>>类型的已经分好的词的列表

输出：List<Map<String, Double>>计算好的每个词以及对应的权重的大小

WTookits/src/com/algorithm/TF.java

class TF

static List<List<String>> splitTermMap

分好的词的列表

static List<Map<String, Double>> termWeightMap

计算得到的权重和对应的词

public TF()

构造函数，初始化成员变量

public List<List<String>> split(List<Map.Entry<String, String>> data)

该函数是 class TF 实现 TermWeighting 接口的 split 函数，用于分词的作用

输入：List<Map.Entry<String, String>>类型的文本数据

输出：List<List<String>>对输入文本分好的词

```
public List<Map<String, Double>> calculate(List<List<String>> splitedWords)
```

该函数是 class TF 实现接口 TermWeighting 中的 calculate 函数，用于计算每个词的权重

输入：List<List<String>>类型的已经分好的词的列表

输出：List<Map<String, Double>>计算好的每个词以及对应的权重的大小

WTookits/src/com/TermWeighting.java

interface TermWeighting

```
List<List<String>> split(List<Map.Entry<String, String>> data)
```

接口 TermWeighting 中的 split 函数，分词作用，在实现接口的类中具体实现

```
List<Map<String, Double>> calculate(List<List<String>> splitedWords)
```

接口 TermWeighting 中的 calculate 函数，分词作用，在实现接口的类中具体实现

Conf 模块

WTookits/src/com/conf/configuration.java

class Configuration

```
private static Configuration configuration
```

配置类 class Configuration 的实例

```
private Properties properties
```

类 Configuration 的成员变量，配置文件的属性

```
private String configNowPath
```

类 Configuration 的成员变量，配置文件的路径

```
private String dirPath
```

类 Configuration 的成员变量，输入文本文件的路径

private String destPath

类 Configuration 的成员变量，输出结果文件的路径

private String termWeightStr

类 Configuration 的成员变量，配置文件的属性

public String getDestPath()

类 Configuration 的成员函数，用于获得保存结果文件的路径

输出：返回路径，String 字符串

public String getTermWeightStr()

类 Configuration 的成员函数，用于获得采用算法的名称

输出：返回使用算法的名字，String 字符串

public String getDirPath()

类 Configuration 的成员函数，用于获得输入文本文件的路径

输出：返回文本文件的路径，String 字符串

private Configuration(String configPath)

类 Configuration 的构造函数，用于从配置文件初始化配置

输入：配置文件的路径，String 字符串

public static Configuration loadConf(String configPath)

类 Configuration 的成员函数，用于加载配置文件

输入：配置文件路径，String 字符串

输出：返回类 Configuration 的实例，因为所有算法共用一个配置文件，采用单例模式

WTookits/src/com/conf/IDFFactory.java

class IDFFactory

private static IDFFactory idfFactory

类 IDFFactory 的成员变量，类 IDFFactory 的实例

private IDFFactory()

类 IDFFactory 的构造函数

public String getName()

成员函数，用于获得当前使用算法的名称标识符

输出：String 类型的标识符

public TermWeighting getTermWeightingInstance()

成员函数，用于构造由该工厂产生的当前算法的实例

输出：该工厂对应的构造算法实例，接口 TremWeighting 类型

WTookits/src/com/conf/TFFactory.java

class TFFactory

private static TFFactory tfFactory

成员变量，类 TFFactory 的实例

private TFFactory()

类 TFFactory 的构造函数

public static TFFactory getTFFactory()

类的成员函数，用于获得类的实例

输出：当前类的实例，TFFactory 的实例

public String getName()

类成员函数，用于返回当前使用的类的名称标识符

输出：类的名称标识符，String 类型

public TermWeighting getTermWeightingInstance()

成员函数，用于构造由该工厂产生的当前算法的实例

输出：该工厂对应的构造算法实例，接口 TremWeighting 类型

WTookits/src/com/conf/TermWeightingFactory.java

interface TermWeightingFactory

String getName()

接口函数，用于返回实现该接口的类的名称标识符

输出：类的名称标识符，String 类型

TermWeighting getTermWeightingInstance()

接口的函数，用于获得接口 TermWeighting 的实例，由实现该接口的类具体实现

输出：TermWeighting 接口的实例

List<TermWeightingFactory> termWeightingFactories

接口函数，用于获得不同算法的构造工厂

输出：TermWeightingFactory 工厂的实例列表

Input 模块

WToolkits/src/com/input/ UnsupervisedDH.java

class UnsupervisedDH

protected List<Map.Entry<String, String>> data

类 UnsupervisedDH 的成员变量，用于存储所有待分析的数据，Map.Entry<String,String>是文件名到文本内容的映射。

protected Map<Integer, String> fileIndex

类 UnsupervisedDH 的成员变量，用于存储 key 为文档编号，value 为文件名

public UnsupervisedDH(String dirPath)

类 UnsupervisedDH 的构造函数，用于对 UnsupervisedDH 对象创建的初始化操作。

输入：dirPath 字段指定要加载的数据文件路径；为 String 类型。

输出：返回一个 UnsupervisedDH 的实例对象

public List<Map.Entry<String, String>> getData()

类 UnsupervisedDH 的成员函数，用于获取已加载的待分析数据。

输入：无

输出：返回 List<Map.Entry<String,String>>数据类型的数据对象。

private void loadFromFile(String dirPath)

类 UnsupervisedDH 的成员函数，用于加载指定目录下的所有待分析的文本。

输入：dirPath 字段指定数据文件路径，该函数根据路径加载该文件；为 String 类型。

输出：无

public void writeResult(List<Map<String, Double>> result, String outFile)

类 UnsupervisedDH 的成员函数，用于对已进行分词以及词频计算操作后的 result

进行按权重排序输出最终结果到指定输出文件 outFile。

输入：List<Map<String, Double>>类型的 result 以及 String 类型的 outFile。

输出：无

WTookits/src/com/input/ SupervisedDH.java

class SupervisedDH

protected Map<String, String> fileToCato

类 SupervisedDH 的成员变量，用于存储待分析文件的类型，key 为文件名，value 为文本类型。

public SupervisedDH(String dirPath, String categoryPath)

类 SupervisedDH 的构造函数，用于对 SupervisedDH 对象创建的初始化操作。

输入：dirPath 字段指定要加载的数据文件的路径，为 String 类型；categoryPath 指定文本类型加载的路径。

输出：返回一个 SupervisedDH 的实例对象

private void loadCatogoryMap(String categoryPath)

类 SupervisedDH 的构造函数，用于对 SupervisedDH 对象创建的初始化操作。

输入：categoryPath 字段指定文本类型加载的路径；为 String 类型。

输出：无

WTookits/src/com/main/ MainHolder.java

class MainHolder

private static TermWeighting termWeighting;

类 MainHolder 的静态私有成员变量，用于存储算法分析对象，类型为 TermWeighting。

public static TermWeighting getTermWeighting()

类 MainHolder 的公有静态成员函数，用于获取算法分析对象。

输入：无。

输出：返回一个 TermWeighting 的衍生子类实例对象

public static void setUp(Configuration configuration)

类 MainHolder 的静态公有成员函数，用于加载配置信息并初始化算法分析对象 termWeighting。

输入：configuration 存储相关配置信息；为 Configuration 类型。
输出：无

WTookits/src/com/main/MainHolderNotInitializedException.java

class MainHolderNotInitializedException extends Exception

WTookits/src/com/main/ MapComparator.java

class MapComparator

public static Map sortByValue(Map map)

类 MapComparator 的公有静态成员函数，用于 Map 的按 value 排序。

输入：Map 类型，value 必须为 int、 double 、 float。

输出：返回一个排序后的 Map 对象