The Manhattan Project

7

Generated by Doxygen 1.8.13

# **Contents**

1	Clas	s Index			1
	1.1	Class	List		1
2	File	Index			3
	2.1	File Lis	st		3
3	Clas	s Docu	mentation	n	5
	3.1	Collec	tion Class	Reference	5
		3.1.1	Construc	ctor & Destructor Documentation	5
			3.1.1.1	Collection()	6
		3.1.2	Member	Function Documentation	6
			3.1.2.1	addDocument()	6
			3.1.2.2	addToIDF()	6
			3.1.2.3	compareSimilar()	6
			3.1.2.4	printSimilarTbl()	6
			3.1.2.5	sim()	6
			3.1.2.6	TFIDF()	7
		3.1.3	Member	Data Documentation	7
			3.1.3.1	Docs	7
			3.1.3.2	IDF	7
			3.1.3.3	library	7
			3.1.3.4	Novern	7
			3.1.3.5	NumDocs	7
			2126	aimularities	7

ii CONTENTS

3.2	ParseE	Box Class Refer	ence	 . 8
	3.2.1	Constructor &	Destructor Documentation	 . 8
		3.2.1.1 Pars	eBox()	 . 8
	3.2.2	Member Func	ion Documentation	 . 8
		3.2.2.1 isAl	phaNum()	 . 8
		3.2.2.2 isLe	gal()	 . 8
		3.2.2.3 long	estLegalIndex()	 . 8
		3.2.2.4 prop	Cap()	 . 9
		3.2.2.5 pun	cOnlyEOS()	 . 9
		3.2.2.6 pure	PuncCheck()	 . 9
3.3	Replac	e Class Refere	nce	 . 9
	3.3.1	Member Func	ion Documentation	 . 9
		3.3.1.1 add	Capital()	 . 9
		3.3.1.2 add	Exceptions()	 . 10
		3.3.1.3 con	ains()	 . 10
		3.3.1.4 swa	pString()	 . 10
3.4	Stemm	er Class Refere	nce	 . 10
	3.4.1	Constructor &	Destructor Documentation	 . 10
		3.4.1.1 Ster	nmer()	 . 10
	3.4.2	Member Func	ion Documentation	 . 10
		3.4.2.1 get	tem()	 . 11
		3.4.2.2 sho	ıldStem()	 . 11
3.5	Word (	Class Reference		 . 11
	3.5.1	Constructor &	Destructor Documentation	 . 11
		3.5.1.1 Wor	d()	 . 11
	3.5.2	Member Func	ion Documentation	 . 11
		3.5.2.1 get0	Count()	 . 12
		3.5.2.2 get\	<b>v</b> ord()	 . 12
		3.5.2.3 incr	ementWord()	 . 12
		3.5.2.4 setV	/ord()	 . 12

CONTENTS

3.6	WordC	ounts Class Reference
	3.6.1	Constructor & Destructor Documentation
		3.6.1.1 WordCounts()
	3.6.2	Member Function Documentation
		3.6.2.1 add()
		3.6.2.2 capitalFix()
		3.6.2.3 checkCaps()
		3.6.2.4 CLindex()
		3.6.2.5 countWords()
		3.6.2.6 getFirstLegalAlphIndex()
		3.6.2.7 getFirstLegalPuncIndex()
		3.6.2.8 implicitSentence()
		3.6.2.9 isAlphaChar()
		3.6.2.10 isAlphanum()
		3.6.2.11 L()
		3.6.2.12 printList()
		3.6.2.13 propCap()
		3.6.2.14 puncOnlyEOS()
		3.6.2.15 removeCapAmb()
		3.6.2.16 S()
		3.6.2.17 scanFileIn()
		3.6.2.18 sortList()
		3.6.2.19 split()
		3.6.2.20 splitAdd()
		3.6.2.21 stringLength()
		3.6.2.22 updateLevelInfo()
	3.6.3	Member Data Documentation
		3.6.3.1 addTmp
		3.6.3.2 capChecks
		3.6.3.3 capTmp
		3.6.3.4 exceptionTable
		3.6.3.5 letterCounter
		3.6.3.6 regBox
		3.6.3.7 sentenceCounter
		3.6.3.8 stemBox
		3.6.3.9 stringList
		3.6.3.10 tmp
		3.6.3.11 wordCount
		3.6.3.12 wordCounter
		3.6.3.13 wordList

iv CONTENTS

4	File I	Pocumentation (1997)	19
	4.1	Collection.cpp File Reference	19
	4.2	Collection.h File Reference	19
	4.3	main.cpp File Reference	19
		4.3.1 Function Documentation	20
		4.3.1.1 exceptionFail()	20
		4.3.1.2 fileFail()	20
		4.3.1.3 main()	20
		4.3.1.4 runCollection()	20
		4.3.1.5 runReadScore()	20
		4.3.1.6 Usage()	21
	4.4	ParseBox.cpp File Reference	21
	4.5	ParseBox.h File Reference	21
	4.6	Replace.cpp File Reference	21
	4.7	Replace.h File Reference	21
	4.8	Stemmer.cpp File Reference	22
	4.9	Stemmer.h File Reference	22
	4.10	Word.cpp File Reference	22
	4.11	Word.h File Reference	22
	4.12	WordCounts.cpp File Reference	22
	4.13	WordCounts.h File Reference	23
Inc	dex		25

Index

# **Chapter 1**

# **Class Index**

# 1.1 Class List

Here are the classes, structs, unions and interfaces with brief descriptions:

Collection							 										 								5
ParseBox							 																		8
Replace .							 																		ç
Stemmer							 																		10
Word							 																		11
WordCoun	ts																								12

2 Class Index

# Chapter 2

# File Index

# 2.1 File List

Here is a list of all files with brief descriptions:

Collection.cpp	19
Collection.h	
main.cpp	
ParseBox.cpp	
ParseBox.h	21
Replace.cpp	21
Replace.h	
Stemmer.cpp	22
Stemmer.h	
Word.cpp	
Word.h	22
WordCounts.cpp	22
WordCounts.h	23

File Index

# **Chapter 3**

# **Class Documentation**

### 3.1 Collection Class Reference

```
#include <Collection.h>
```

#### **Public Member Functions**

- Collection (int N)
- int addDocument (istream &fileName, Replace exceptions)
- bool compareSimilar ()
- void printSimilarTbl ()
- double sim (WordCounts &docA, WordCounts &docB)
- double TFIDF (Word &wordIn)

#### **Protected Member Functions**

• void addToIDF (const WordCounts &words)

#### **Protected Attributes**

- int NumDocs
- double Novern
- int Docs
- map< string, int > IDF
- vector< WordCounts > library
- vector< vector< double >> simularities

#### 3.1.1 Constructor & Destructor Documentation

#### 3.1.1.1 Collection()

```
Collection::Collection ( int N = 0 )
```

#### 3.1.2 Member Function Documentation

#### 3.1.2.1 addDocument()

#### 3.1.2.2 addToIDF()

#### 3.1.2.3 compareSimilar()

```
bool Collection::compareSimilar ( )
```

#### 3.1.2.4 printSimilarTbl()

```
void Collection::printSimilarTbl ( )
```

#### 3.1.2.5 sim()

#### 3.1.2.6 TFIDF()

#### 3.1.3 Member Data Documentation

#### 3.1.3.1 Docs

```
int Collection::Docs [protected]
```

#### 3.1.3.2 IDF

```
map<string, int> Collection::IDF [protected]
```

#### 3.1.3.3 library

```
vector<WordCounts> Collection::library [protected]
```

#### 3.1.3.4 Novern

```
double Collection::Novern [protected]
```

#### 3.1.3.5 NumDocs

```
int Collection::NumDocs [protected]
```

#### 3.1.3.6 simularities

```
vector<vector<double> > Collection::simularities [protected]
```

The documentation for this class was generated from the following files:

- · Collection.h
- Collection.cpp

### 3.2 ParseBox Class Reference

```
#include <ParseBox.h>
```

#### **Public Member Functions**

- ParseBox ()
- bool isLegal (const string &word)
- int longestLegalIndex (const string &word)
- bool propCap (const string &word)
- bool puncOnlyEOS (const string &word)
- bool isAlphaNum (const string &word)
- bool purePuncCheck (const string &word)

#### 3.2.1 Constructor & Destructor Documentation

#### 3.2.1.1 ParseBox()

```
ParseBox::ParseBox ( )
```

#### 3.2.2 Member Function Documentation

#### 3.2.2.1 isAlphaNum()

#### 3.2.2.2 isLegal()

#### 3.2.2.3 longestLegalIndex()

#### 3.2.2.4 propCap()

The documentation for this class was generated from the following files:

const string & word )

- ParseBox.h
- ParseBox.cpp

# 3.3 Replace Class Reference

bool ParseBox::purePuncCheck (

```
#include <Replace.h>
```

#### **Public Member Functions**

- bool addExceptions (istream &inExc)
- string swapString (const string &in)
- void addCapital (const string &cap)
- bool contains (const string &cap)

#### 3.3.1 Member Function Documentation

#### 3.3.1.1 addCapital()

#### 3.3.1.2 addExceptions()

The documentation for this class was generated from the following files:

const string & in )

- · Replace.h
- Replace.cpp

# 3.4 Stemmer Class Reference

```
#include <Stemmer.h>
```

#### **Public Member Functions**

- Stemmer ()
- bool shouldStem (const string &word)
- void getStem (string &word)

#### 3.4.1 Constructor & Destructor Documentation

#### 3.4.1.1 Stemmer()

```
Stemmer::Stemmer ( )
```

#### 3.4.2 Member Function Documentation

3.5 Word Class Reference

#### 3.4.2.1 getStem()

#### 3.4.2.2 shouldStem()

The documentation for this class was generated from the following files:

- · Stemmer.h
- Stemmer.cpp

#### 3.5 Word Class Reference

```
#include <Word.h>
```

#### **Public Member Functions**

- Word (string wordIn="")
- string getWord () const
- int getCount ()
- void incrementWord ()
- void setWord (string newWord)

#### 3.5.1 Constructor & Destructor Documentation

### 3.5.1.1 Word()

#### 3.5.2 Member Function Documentation

The documentation for this class was generated from the following file:

• Word.h

3.5.2.1 getCount()

#### 3.6 WordCounts Class Reference

```
#include <WordCounts.h>
```

#### **Public Member Functions**

- WordCounts (const Replace &exec)
- int scanFileIn (istream &inFile)
- void sortList ()
- void printList ()
- void countWords ()
- int stringLength ()
- void updateLevelInfo (const string &word)
- double L ()
- double S ()
- double CLindex ()
- void implicitSentence ()

Adds one to sentence if there is an implicit Sentence in the file.

- void splitAdd (string &word)
- void add (string word)
- void split (string word)
- bool isAlphanum (string word)
- int getFirstLegalPuncIndex (string word)
- int getFirstLegalAlphIndex (string word)
- bool isAlphaChar (char c)
- bool propCap (string word)
- bool puncOnlyEOS (string word)
- void capitalFix ()
- void removeCapAmb ()
- void checkCaps ()

#### **Public Attributes**

- vector< Word > wordList
- vector< string> stringList
- vector< int > capChecks
- int wordCount
- ParseBox regBox
- Stemmer stemBox
- Replace exceptionTable
- string tmp
- string addTmp
- string capTmp
- int letterCounter
- int wordCounter
- · int sentenceCounter

#### 3.6.1 Constructor & Destructor Documentation

### 3.6.1.1 WordCounts()

#### 3.6.2 Member Function Documentation

```
3.6.2.1 add()
```

```
void WordCounts::add (
     string word )
```

### 3.6.2.2 capitalFix()

```
void WordCounts::capitalFix ( )
```

#### 3.6.2.3 checkCaps()

```
void WordCounts::checkCaps ( )
```

```
3.6.2.4 CLindex()
```

```
double WordCounts::CLindex ( )
```

#### 3.6.2.5 countWords()

```
void WordCounts::countWords ( )
```

#### 3.6.2.6 getFirstLegalAlphIndex()

### 3.6.2.7 getFirstLegalPuncIndex()

#### 3.6.2.8 implicitSentence()

```
void WordCounts::implicitSentence ( )
```

Adds one to sentence if there is an implicit Sentence in the file.

Note

if the file ends in end of sentence token then nothing is changed

#### 3.6.2.9 isAlphaChar()

```
bool WordCounts::isAlphaChar ( {\tt char} \ c \ )
```

### 3.6.2.10 isAlphanum()

```
bool WordCounts::isAlphanum (
            string word )
3.6.2.11 L()
double WordCounts::L ( )
3.6.2.12 printList()
void WordCounts::printList ( )
3.6.2.13 propCap()
bool WordCounts::propCap (
           string word )
3.6.2.14 puncOnlyEOS()
bool WordCounts::puncOnlyEOS (
            string word )
3.6.2.15 removeCapAmb()
void WordCounts::removeCapAmb ( )
3.6.2.16 S()
```

double WordCounts::S ( )

```
3.6.2.17 scanFileIn()
```

```
int WordCounts::scanFileIn (
       istream & inFile )
3.6.2.18 sortList()
void WordCounts::sortList ( )
3.6.2.19 split()
void WordCounts::split (
            string word )
3.6.2.20 splitAdd()
void WordCounts::splitAdd (
            string & word )
3.6.2.21 stringLength()
int WordCounts::stringLength ( )
3.6.2.22 updateLevelInfo()
```

### 3.6.3 Member Data Documentation

#### 3.6.3.1 addTmp

string WordCounts::addTmp

#### 3.6.3.2 capChecks

vector<int> WordCounts::capChecks

#### 3.6.3.3 capTmp

string WordCounts::capTmp

#### 3.6.3.4 exceptionTable

Replace WordCounts::exceptionTable

#### 3.6.3.5 letterCounter

int WordCounts::letterCounter

#### 3.6.3.6 regBox

ParseBox WordCounts::regBox

#### 3.6.3.7 sentenceCounter

int WordCounts::sentenceCounter

#### 3.6.3.8 stemBox

Stemmer WordCounts::stemBox

#### 3.6.3.9 stringList

vector<string> WordCounts::stringList

#### 3.6.3.10 tmp

string WordCounts::tmp

#### 3.6.3.11 wordCount

int WordCounts::wordCount

#### 3.6.3.12 wordCounter

int WordCounts::wordCounter

#### 3.6.3.13 wordList

vector<Word> WordCounts::wordList

The documentation for this class was generated from the following files:

- WordCounts.h
- WordCounts.cpp

# **Chapter 4**

# **File Documentation**

# 4.1 Collection.cpp File Reference

```
#include <Collection.h>
#include <math.h>
#include <iostream>
```

### 4.2 Collection.h File Reference

```
#include <Collection.h>
#include <Replace.h>
#include <WordCounts.h>
#include <map>
```

#### **Classes**

class Collection

# 4.3 main.cpp File Reference

```
#include <iostream>
#include <fstream>
#include <WordCounts.h>
#include <Replace.h>
#include <Collection.h>
```

20 File Documentation

#### **Functions**

```
• int Usage (char *arg0)
```

- int fileFail (char \*arg2)
- int exceptionFail (char \*arg1)
- int runCollection (int argc, char \*argv[])
- int runReadScore (int argc, char \*argv[])
- int main (int argc, char \*argv[])

#### 4.3.1 Function Documentation

#### 4.3.1.1 exceptionFail()

```
int exceptionFail ( {\tt char} \ * \ {\tt arg1} \ )
```

#### 4.3.1.2 fileFail()

```
int fileFail ( {\tt char} \ * \ {\tt arg2} \ )
```

#### 4.3.1.3 main()

```
int main (
                      int argc,
                      char * argv[] )
```

#### 4.3.1.4 runCollection()

```
int runCollection (
                int argc,
                char * argv[] )
```

#### 4.3.1.5 runReadScore()

```
int runReadScore (
                int argc,
                char * argv[] )
```

#### 4.3.1.6 Usage()

```
int Usage ( {\tt char} \ * \ {\tt arg0} \ )
```

# 4.4 ParseBox.cpp File Reference

```
#include <ParseBox.h>
#include <regex>
#include <string>
#include <stdio.h>
#include <iostream>
```

#### 4.5 ParseBox.h File Reference

```
#include <regex>
#include <string>
#include <stdio.h>
#include <iostream>
```

#### **Classes**

• class ParseBox

### 4.6 Replace.cpp File Reference

```
#include <Replace.h>
```

# 4.7 Replace.h File Reference

```
#include <iostream>
#include <fstream>
#include <map>
#include <set>
#include <regex>
```

#### Classes

• class Replace

22 File Documentation

# 4.8 Stemmer.cpp File Reference

```
#include <Stemmer.h>
#include <iostream>
```

### 4.9 Stemmer.h File Reference

```
#include <regex>
#include <string>
```

#### Classes

• class Stemmer

# 4.10 Word.cpp File Reference

```
#include <Word.h>
```

### 4.11 Word.h File Reference

```
#include <string>
```

#### Classes

class Word

# 4.12 WordCounts.cpp File Reference

```
#include <iostream>
#include <fstream>
#include <vector>
#include <string>
#include <algorithm>
#include <WordCounts.h>
#include <regex>
```

### 4.13 WordCounts.h File Reference

```
#include <iostream>
#include <fstream>
#include <vector>
#include <algorithm>
#include <string>
#include <Word.h>
#include <regex>
#include <ParseBox.h>
#include <Stemmer.h>
#include <Replace.h>
```

#### Classes

class WordCounts

24 File Documentation

# Index

add	main.cpp, 20
WordCounts, 13	exceptionTable
addCapital Replace, 9	WordCounts, 17
addDocument	fileFail
Collection, 6	main.cpp, 20
addExceptions	
Replace, 9	getCount
addTmp	Word, 11
WordCounts, 16	getFirstLegalAlphIndex
addToIDF	WordCounts, 14
Collection, 6	getFirstLegalPuncIndex WordCounts, 14
CLindex	getStem Stemmer, 10
WordCounts, 13	getWord
capChecks	Word, 12
WordCounts, 16	77010, 72
capTmp	IDF
WordCounts, 17	Collection, 7
capitalFix	implicitSentence
WordCounts, 13	WordCounts, 14
checkCaps WordCounts, 13	incrementWord
Collection, 5	Word, 12
addDocument, 6	isAlphaChar
addToIDF, 6	WordCounts, 14
Collection, 5	isAlphaNum
compareSimilar, 6	ParseBox, 8
Docs, 7	isAlphanum WordCounts 14
IDF, 7	WordCounts, 14 isLegal
library, 7	ParseBox, 8
Novern, 7	Tarsobox, o
NumDocs, 7	L
printSimilarTbl, 6	WordCounts, 15
sim, 6	letterCounter
simularities, 7	WordCounts, 17
TFIDF, 6	library
Collection.cpp, 19	Collection, 7
Collection.h, 19	longestLegalIndex
compareSimilar	ParseBox, 8
Collection, 6	main
contains	main.cpp, 20
Replace, 10 countWords	main.cpp, 19
WordCounts, 14	exceptionFail, 20
Trordooding, 17	fileFail, 20
Docs	main, 20
Collection, 7	runCollection, 20
•	runReadScore, 20
exceptionFail	Usage, 20

26 INDEX

Novern	WordCounts, 16
Collection, 7	split
NumDocs	. WordCounts, 16
Collection, 7	splitAdd
,	WordCounts, 16
ParseBox, 8	stemBox
isAlphaNum, 8	WordCounts, 17
isLegal, 8	Stemmer, 10
longestLegalIndex, 8	getStem, 10
ParseBox, 8	
propCap, 8	shouldStem, 11
puncOnlyEOS, 9	Stemmer, 10
	Stemmer.cpp, 22
purePuncCheck, 9	Stemmer.h, 22
ParseBox.cpp, 21	stringLength
ParseBox.h, 21	WordCounts, 16
printList	stringList
WordCounts, 15	WordCounts, 17
printSimilarTbl	swapString
Collection, 6	Replace, 10
propCap	
ParseBox, 8	TFIDF
WordCounts, 15	Collection, 6
puncOnlyEOS	tmp
ParseBox, 9	WordCounts, 17
WordCounts, 15	
purePuncCheck	updateLevelInfo
ParseBox, 9	WordCounts, 16
	Usage
regBox	main.cpp, 20
WordCounts, 17	
removeCapAmb	Word, 11
WordCounts, 15	getCount, 11
Replace, 9	getWord, 12
addCapital, 9	incrementWord, 12
addExceptions, 9	setWord, 12
contains, 10	Word, 11
swapString, 10	Word.cpp, 22
Replace.cpp, 21	Word.h, 22
Replace.h, 21	wordCount
runCollection	WordCounts, 18
main.cpp, 20	wordCounter
runReadScore	WordCounts, 18
main.cpp, 20	WordCounts, 12
Пап.орр, 20	add, 13
S	addTmp, 16
WordCounts, 15	CLindex, 13
scanFileIn	capChecks, 16
WordCounts, 15	capTmp, 17
sentenceCounter	capitalFix, 13
WordCounts, 17	checkCaps, 13
setWord	countWords, 14
Word, 12	exceptionTable, 17
shouldStem	getFirstLegalAlphIndex, 14
Stemmer, 11	getFirstLegalPuncIndex, 14
sim	implicitSentence, 14
Collection, 6	isAlphaChar, 14
simularities	isAlphanum, 14
Collection, 7	L, 15
sortList	letterCounter, 17

INDEX 27

```
printList, 15
    propCap, 15
    puncOnlyEOS, 15
    regBox, 17
    removeCapAmb, 15
    S, 15
    scanFileIn, 15
    sentenceCounter, 17
    sortList, 16
    split, 16
    splitAdd, 16
    stemBox, 17
    stringLength, 16
    stringList, 17
    tmp, 17
    updateLevelInfo, 16
    wordCount, 18
    wordCounter, 18
    WordCounts, 13
    wordList, 18
WordCounts.cpp, 22
WordCounts.h, 23
wordList
    WordCounts, 18
```