PLSA (baseline) 1.0.1

Generated by Doxygen 1.5.6

Sun Mar 7 16:16:27 2010

# **Contents**

1	Data	a Struct	ure Index		1
	1.1	Data S	tructures		 1
2	File	Index			3
	2.1	File L	st		 3
3	Data	a Struct	ure Docui	mentation	5
	3.1	coocci	ır Struct R	eference	 5
		3.1.1	Field Do	cumentation	 5
			3.1.1.1	x	 5
			3.1.1.2	column	 5
	3.2	info S	ruct Refer	ence	 6
		3.2.1	Field Do	cumentation	 7
			3.2.1.1	verbose	 7
			3.2.1.2	debug	 7
			3.2.1.3	textio	 7
			3.2.1.4	rounding	 7
			3.2.1.5	no_output	 7
			3.2.1.6	seed	 7
			3.2.1.7	num_clusters	 7
			3.2.1.8	base_fn	 7
			3.2.1.9	maxiter	 7
			3.2.1.10	m	 7
			3.2.1.11	n	 7
			3.2.1.12	co_fn	 8
			3.2.1.13	cos	 8
			3.2.1.14	row_ids	 8
			3.2.1.15	column_ids	 8
			2 2 1 16	Maria Cara Cara Cara Cara Cara Cara Cara	0

ii CONTENTS

			3.2.1.17	probw1_z	8
			3.2.1.18	probw2_z	8
			3.2.1.19	probz	8
			3.2.1.20	probz_w1w2	8
			3.2.1.21	world_id	8
			3.2.1.22	world_size	8
			3.2.1.23	block_start	9
			3.2.1.24	block_end	9
			3.2.1.25	block_size	9
			3.2.1.26	sigfpe_count	9
			3.2.1.27	program_start	9
			3.2.1.28	run_time	9
			3.2.1.29	readCO_time	9
			3.2.1.30	initEM_time	9
			3.2.1.31	calculateML_time	9
			3.2.1.32	applyEStep_time	9
			3.2.1.33	applyMStep_time	9
			3.2.1.34	normalizeProbs_time	9
			3.2.1.35	printCoProbs_time	9
			3.2.1.36	program_end	9
	3.3	wmstru	act Struct l	Reference	10
		3.3.1	Field Do	cumentation	10
			3.3.1.1	ptr	10
			3.3.1.2	size	10
			3.3.1.3	file	10
			3.3.1.4	line	10
			3.3.1.5	next	10
4	File	Dooum	entation		11
•	4.1			erence	11
	4.1	4.1.1		ocumentation	12
		7.1.1	4.1.1.1	HAVE_BOOL	12
			4.1.1.2	HAVE_FLOAT_H	12
			4.1.1.3	HAVE_INTTYPES_H	12
			4.1.1.3	HAVE_LIMITS_H	12
			4.1.1.4	HAVE_MALLOC	12
			4.1.1.6	HAVE_MEMORY_H	12
			7.1.1.0	TIAVE_MILMORT_IT	1.4

CONTENTS

		4.1.1.7	HAVE_REALLOC	12
		4.1.1.8	HAVE_STDBOOL_H	12
		4.1.1.9	HAVE_STDINT_H	12
		4.1.1.10	HAVE_STDLIB_H	12
		4.1.1.11	HAVE_STRING_H	12
		4.1.1.12	HAVE_STRINGS_H	12
		4.1.1.13	HAVE_SYS_STAT_H	12
		4.1.1.14	HAVE_SYS_TYPES_H	12
		4.1.1.15	HAVE_UNISTD_H	12
		4.1.1.16	PACKAGE	12
		4.1.1.17	PACKAGE_BUGREPORT	12
		4.1.1.18	PACKAGE_NAME	12
		4.1.1.19	PACKAGE_STRING	12
		4.1.1.20	PACKAGE_TARNAME	12
		4.1.1.21	PACKAGE_VERSION	12
		4.1.1.22	STDC_HEADERS	12
		4.1.1.23	VERSION	12
4.2	debug.	c File Refe	erence	13
	4.2.1	Function	Documentation	13
		4.2.1.1	checkCoProb	13
		4.2.1.2	debugCheckCo	13
		4.2.1.3	handler_sigfpe	13
4.3	debug.	h File Ref	erence	14
	4.3.1	Function	Documentation	14
		4.3.1.1	checkCoProb	14
		4.3.1.2	debugCheckCo	14
		4.3.1.3	handler_sigfpe	14
4.4	em-est	ep.c File F	Reference	15
	4.4.1	Function	Documentation	15
		4.4.1.1	applyEStep	15
		4.4.1.2	calculateML	15
		4.4.1.3	initEM	15
4.5	em-est	ep.h File F	Reference	16
	4.5.1	Function	Documentation	16
		4.5.1.1	applyEStep	16
		4.5.1.2	calculateML	16

iv CONTENTS

		4.5.1.3	initEM	 16
4.6	em-ms	step.c File	Reference	 17
	4.6.1	Function	Documentation	 17
		4.6.1.1	applyMStep	 17
		4.6.1.2	normalizeProbs	 17
4.7	em-ms	step.h File	Reference	 18
	4.7.1	Function	Documentation	 18
		4.7.1.1	applyMStep	 18
		4.7.1.2	normalizeProbs	 18
4.8	input.c	File Refe	erence	 19
	4.8.1	Function	Documentation	 19
		4.8.1.1	initializePostInput	 19
		4.8.1.2	readCO	 19
4.9	input.l	r File Refe	erence	 20
	4.9.1	Function	Documentation	 20
		4.9.1.1	initializePostInput	 20
		4.9.1.2	readCO	 20
4.10	) main.c	File Refe	erence	 21
	4.10.1	Function	Documentation	 21
		4.10.1.1	main	 21
4.1	output	.c File Ref	ference	 22
	4.11.1	Function	Documentation	 22
		4.11.1.1	printCoProb	 22
4.12	2 output	.h File Re	ference	 23
	4.12.1	Function	Documentation	 23
		4.12.1.1	printCoProb	 23
4.13	B param	eters.c File	e Reference	 24
	4.13.1	Define D	Documentation	 24
		4.13.1.1	_GNU_SOURCE	 24
	4.13.2	Function	Documentation	 24
		4.13.2.1	checkSettings	 24
		4.13.2.2	processOptions	 24
		4.13.2.3	usage	 24
4.14	param	eters.h Fil	e Reference	 25
	4.14.1	Function	Documentation	 25
		4.14.1.1	checkSettings	 25

CONTENTS

4.14.1.2 processOptions	
4.14.1.3 usage	25
4.15 plsa-defn.h File Reference	26
4.15.1 Define Documentation	27
4.15.1.1 BLOCK_SIZE	27
4.15.1.2 DBL_LESS	27
4.15.1.3 DOEXP	27
4.15.1.4 DOLOG	27
4.15.1.5 DOLOG1PEXP	27
4.15.1.6 DOLOGONE	27
4.15.1.7 FCLOSE	27
4.15.1.8 FOPEN	27
4.15.1.9 GET_COS	27
4.15.1.10 GET_COS_POSITION	27
4.15.1.11 GET_PROBW1_Z	27
4.15.1.12 GET_PROBW2_Z	28
4.15.1.13 GET_PROBZ	28
4.15.1.14 GET_PROBZ_W1W2	28
4.15.1.15 LN_LIMIT	28
4.15.1.16 logSumsInline	28
4.15.1.17 MAINPROC	28
4.15.1.18 MIN_PROB	28
4.15.1.19 ML_DELTA	28
4.15.1.20 PROGRESS_MSG	28
4.15.1.21 RANDOM_FLOAT	29
4.15.1.22 ROUND_DIGITS	29
4.15.1.23 SET_COS	29
4.15.2 Typedef Documentation	29
4.15.2.1 COOCCUR	29
4.15.2.2 INFO	29
4.15.2.3 PROBNODE	29
4.16 run.c File Reference	30
4.16.1 Function Documentation	30
4.16.1.1 initialize	30
4.16.1.2 run	30
4.16.1.3 uninitialize	30

vi CONTENTS

4.17 run.h File Reference	
4.17.1 Function Documentation	
4.17.1.1 initialize	
4.17.1.2 run	
4.17.1.3 uninitialize	
4.18 wmalloc.c File Reference	
4.18.1 Function Documentation	
4.18.1.1 countFree	
4.18.1.2 countMalloc	
4.18.1.3 hash	
4.18.1.4 initWMalloc	
4.18.1.5 printInUseWMalloc	
4.18.1.6 printWMalloc	
4.18.1.7 wfree	
4.18.1.8 wmalloc	
4.18.1.9 wrealloc	
4.18.2 Variable Documentation	
4.18.2.1 inuse_malloc	
4.18.2.2 max_malloc	
4.18.2.3 tempstr	
4.18.2.4 wm_array	
4.19 wmalloc.h File Reference	
4.19.1 Define Documentation	
4.19.1.1 TEMPSTRLEN	
4.19.1.2 WM_SIZE	
4.19.2 Typedef Documentation	
4.19.2.1 WMSTRUCT	
4.19.3 Function Documentation	
4.19.3.1 countFree	
4.19.3.2 countMalloc	
4.19.3.3 initWMalloc	
4.19.3.4 printInUseWMalloc	
4.19.3.7 wmalloc	

# **Chapter 1**

# **Data Structure Index**

## 1.1 Data Structures

Here are	the data	structures	with	brief	descri	ptions

cooccur								•								•		•								5
info																										6
wmstruct							 							 	 										1	10

2 Data Structure Index

# **Chapter 2**

# **File Index**

## 2.1 File List

Here is a list of all files with brief descriptions:

config.n	I
debug.c	3
debug.h	4
em-estep.c	5
em-estep.h	6
em-mstep.c	7
em-mstep.h	8
input.c	9
input.h	0
main.c	1
output.c	2
output.h	3
parameters.c	4
parameters.h	5
plsa-defn.h	6
run.c	0
run.h	1
wmalloc.c	2
yymelloo h	4

4 File Index

## **Chapter 3**

## **Data Structure Documentation**

#### 3.1 cooccur Struct Reference

#include <plsa-defn.h>

#### **Data Fields**

- PROBNODE x
- unsigned int column

#### 3.1.1 Field Documentation

#### 3.1.1.1 PROBNODE cooccur::x

The co-occurrence count, as a log value

#### 3.1.1.2 unsigned int cooccur::column

Column position of this value

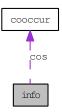
The documentation for this struct was generated from the following file:

• plsa-defn.h

#### 3.2 info Struct Reference

#include <plsa-defn.h>

Collaboration diagram for info:



#### **Data Fields**

- bool verbose
- bool debug
- bool textio
- bool rounding
- bool no\_output
- unsigned int seed
- unsigned int num\_clusters
- char \* base\_fn
- unsigned int maxiter
- unsigned int m
- unsigned int n
- char \* co\_fn
- COOCCUR \*\* cos
- unsigned int \* row\_ids
- unsigned int \* column\_ids
- unsigned int iter
- PROBNODE \* probw1\_z
- PROBNODE \* probw2\_z
- PROBNODE \* probz
- PROBNODE \*\* probz\_w1w2
- signed int world\_id
- signed int world\_size
- unsigned int block\_start
- unsigned int block\_end
- unsigned int block\_size
- unsigned int sigfpe\_count
- time\_t program\_start
- double run\_time
- double readCO\_time
- double initEM\_time
- double calculateML\_time
- double applyEStep\_time
- double applyMStep\_time
- double normalizeProbs\_time
- double printCoProbs\_time
- time\_t program\_end

3.2 info Struct Reference 7

#### 3.2.1 Field Documentation

#### 3.2.1.1 bool info::verbose

Verbose output?

#### 3.2.1.2 bool info::debug

Debugging output?

#### 3.2.1.3 bool info::textio

Text I/O

#### 3.2.1.4 bool info::rounding

Should the output values be rounded?

#### 3.2.1.5 bool info::no\_output

Suppress output

#### 3.2.1.6 unsigned int info::seed

Random seed

#### 3.2.1.7 unsigned int info::num\_clusters

Number of clusters

#### 3.2.1.8 char\* info::base\_fn

Base filename for the output file

#### 3.2.1.9 unsigned int info::maxiter

Maximum number of iterations

#### 3.2.1.10 unsigned int info::m

Number of unique query terms

#### 3.2.1.11 unsigned int info::n

Number of terms in the document collection

#### 3.2.1.12 char\* info::co\_fn

Co-occurrence filename

#### 3.2.1.13 COOCCUR\*\* info::cos

Co-occurrence counts in a COOCCUR data structure

#### 3.2.1.14 unsigned int\* info::row\_ids

List of row identifiers (m of them)

#### 3.2.1.15 unsigned int\* info::column ids

List of column identifiers (m of them)

#### 3.2.1.16 unsigned int info::iter

Iteration; only calculated by the main process and broadcasted to others

#### 3.2.1.17 PROBNODE\* info::probw1\_z

P(w1|z) of size (k \* m)

#### 3.2.1.18 PROBNODE\* info::probw2\_z

P(w2|z) of size (k \* n)

#### 3.2.1.19 PROBNODE\* info::probz

P(z) of size (k); one-dimensional array does not need a pointer

#### 3.2.1.20 PROBNODE\*\* info::probz\_w1w2

P(z|w1w2) of size (k \* m \* n)

#### 3.2.1.21 signed int info::world\_id

ID of this process

#### 3.2.1.22 signed int info::world\_size

Number of processes total

3.2 info Struct Reference 9

#### 3.2.1.23 unsigned int info::block\_start

Starting block (cluster) for this process to handle

#### 3.2.1.24 unsigned int info::block\_end

Ending block (cluster) for this process to handle

#### 3.2.1.25 unsigned int info::block\_size

Size of the block for this process to handle

#### 3.2.1.26 unsigned int info::sigfpe\_count

Number of floating point exception errors

- 3.2.1.27 time\_t info::program\_start
- 3.2.1.28 double info::run\_time
- 3.2.1.29 double info::readCO\_time
- 3.2.1.30 double info::initEM\_time
- 3.2.1.31 double info::calculateML\_time
- 3.2.1.32 double info::applyEStep\_time
- 3.2.1.33 double info::applyMStep\_time
- 3.2.1.34 double info::normalizeProbs\_time
- 3.2.1.35 double info::printCoProbs\_time

#### 3.2.1.36 time\_t info::program\_end

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

• plsa-defn.h

#### 3.3 wmstruct Struct Reference

#include <wmalloc.h>

Collaboration diagram for wmstruct:



#### **Data Fields**

- void \* ptr
- size\_t size
- char \* file
- unsigned int line
- struct wmstruct \* next

#### 3.3.1 Field Documentation

- 3.3.1.1 void\* wmstruct::ptr
- 3.3.1.2 size\_t wmstruct::size
- 3.3.1.3 char\* wmstruct::file
- 3.3.1.4 unsigned int wmstruct::line
- **3.3.1.5 struct wmstruct\* wmstruct::next** [read]

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

• wmalloc.h

## **Chapter 4**

## **File Documentation**

### 4.1 config.h File Reference

#### **Defines**

- #define HAVE\_FLOAT\_H 1
- #define HAVE\_INTTYPES\_H 1
- #define HAVE\_LIMITS\_H 1
- #define HAVE\_MALLOC 1
- #define HAVE\_MEMORY\_H 1
- #define HAVE\_REALLOC 1
- #define HAVE\_STDBOOL\_H 1
- #define HAVE\_STDINT\_H 1
- #define HAVE\_STDLIB\_H 1
- #define HAVE\_STRINGS\_H 1
- #define HAVE\_STRING\_H 1
- #define HAVE\_SYS\_STAT\_H 1
- #define HAVE\_SYS\_TYPES\_H 1
- #define HAVE\_UNISTD\_H 1
- #define HAVE\_BOOL 1
- #define PACKAGE "plsa-base"
- #define PACKAGE\_BUGREPORT "r.wan@aist.go.jp"
- #define PACKAGE\_NAME "PLSA-BASE"
- #define PACKAGE\_STRING "PLSA-BASE 1.0"
- #define PACKAGE\_TARNAME "plsa-base"
- #define PACKAGE\_VERSION "1.0"
- #define STDC\_HEADERS 1
- #define VERSION "1.0"

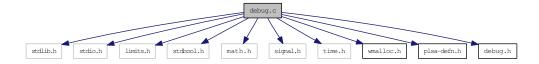
#### **4.1.1** Define Documentation

- 4.1.1.1 #define HAVE\_BOOL 1
- 4.1.1.2 #define HAVE\_FLOAT\_H 1
- 4.1.1.3 #define HAVE\_INTTYPES\_H 1
- 4.1.1.4 #define HAVE\_LIMITS\_H 1
- 4.1.1.5 #define HAVE\_MALLOC 1
- 4.1.1.6 #define HAVE\_MEMORY\_H 1
- 4.1.1.7 #define HAVE\_REALLOC 1
- 4.1.1.8 #define HAVE\_STDBOOL\_H 1
- 4.1.1.9 #define HAVE\_STDINT\_H 1
- 4.1.1.10 #define HAVE\_STDLIB\_H 1
- 4.1.1.11 #define HAVE\_STRING\_H 1
- 4.1.1.12 #define HAVE\_STRINGS\_H 1
- 4.1.1.13 #define HAVE\_SYS\_STAT\_H 1
- 4.1.1.14 #define HAVE\_SYS\_TYPES\_H 1
- 4.1.1.15 #define HAVE\_UNISTD\_H 1
- 4.1.1.16 #define PACKAGE "plsa-base"
- 4.1.1.17 #define PACKAGE\_BUGREPORT "r.wan@aist.go.jp"
- 4.1.1.18 #define PACKAGE\_NAME "PLSA-BASE"
- 4.1.1.19 #define PACKAGE\_STRING "PLSA-BASE 1.0"
- 4.1.1.20 #define PACKAGE\_TARNAME "plsa-base"
- 4.1.1.21 #define PACKAGE\_VERSION "1.0"
- 4.1.1.22 #define STDC\_HEADERS 1
- 4.1.1.23 #define VERSION "1.0"

## 4.2 debug.c File Reference

```
#include <stdlib.h>
#include <stdio.h>
#include <limits.h>
#include <stdbool.h>
#include <math.h>
#include <signal.h>
#include <time.h>
#include "wmalloc.h"
#include "plsa-defn.h"
#include "debug.h"
```

Include dependency graph for debug.c:



#### **Functions**

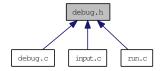
- void handler\_sigfpe ()
- void debugCheckCo (INFO \*info)
- void checkCoProb (INFO \*info)

#### **4.2.1** Function Documentation

- 4.2.1.1 void checkCoProb (INFO \* info)
- 4.2.1.2 void debugCheckCo (INFO \* info)
- 4.2.1.3 void handler\_sigfpe ()

## 4.3 debug.h File Reference

This graph shows which files directly or indirectly include this file:



#### **Functions**

- void handler\_sigfpe ()
- void debugCheckCo (INFO \*info)
- void checkCoProb (INFO \*info)

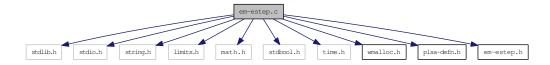
#### **4.3.1** Function Documentation

- 4.3.1.1 void checkCoProb (INFO \* info)
- 4.3.1.2 void debugCheckCo (INFO \* info)
- 4.3.1.3 void handler\_sigfpe ()

### 4.4 em-estep.c File Reference

```
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <limits.h>
#include <math.h>
#include <stdbool.h>
#include <time.h>
#include "wmalloc.h"
#include "plsa-defn.h"
#include "em-estep.h"
```

Include dependency graph for em-estep.c:



#### **Functions**

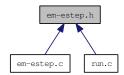
- void initEM (INFO \*info)
- void applyEStep (INFO \*info)
- PROBNODE calculateML (INFO \*info)

#### **4.4.1** Function Documentation

- 4.4.1.1 void applyEStep (INFO \* info)
- 4.4.1.2 PROBNODE calculateML (INFO \* info)
- 4.4.1.3 void initEM (INFO \* info)

## 4.5 em-estep.h File Reference

This graph shows which files directly or indirectly include this file:



#### **Functions**

- void initEM (INFO \*info)
- void applyEStep (INFO \*info)
- PROBNODE calculateML (INFO \*info)

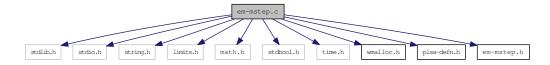
#### **4.5.1** Function Documentation

- 4.5.1.1 void applyEStep (INFO \* info)
- 4.5.1.2 PROBNODE calculateML (INFO \* info)
- 4.5.1.3 void initEM (INFO \* info)

### 4.6 em-mstep.c File Reference

```
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <limits.h>
#include <math.h>
#include <stdbool.h>
#include <time.h>
#include "wmalloc.h"
#include "plsa-defn.h"
#include "em-mstep.h"
```

Include dependency graph for em-mstep.c:



#### **Functions**

- void applyMStep (INFO \*info)
- void normalizeProbs (INFO \*info)

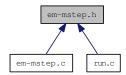
#### **4.6.1** Function Documentation

- **4.6.1.1** void applyMStep (INFO \* *info*)
- **4.6.1.2** void normalizeProbs (INFO \* *info*)

Normalize probabilities

## 4.7 em-mstep.h File Reference

This graph shows which files directly or indirectly include this file:



#### **Functions**

- void applyMStep (INFO \*info)
- void normalizeProbs (INFO \*info)

#### **4.7.1** Function Documentation

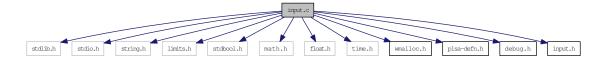
- 4.7.1.1 void applyMStep (INFO \* info)
- 4.7.1.2 void normalizeProbs (INFO \* info)

Normalize probabilities

### 4.8 input.c File Reference

```
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <limits.h>
#include <stdbool.h>
#include <math.h>
#include <float.h>
#include <time.h>
#include "wmalloc.h"
#include "plsa-defn.h"
#include "debug.h"
#include "input.h"
```

Include dependency graph for input.c:



#### **Functions**

- void initializePostInput (INFO \*info)
- bool readCO (INFO \*info)

#### 4.8.1 Function Documentation

#### **4.8.1.1** void initializePostInput (INFO \* *info*)

Initialization that depends on the input file or parameters

#### 4.8.1.2 bool readCO (INFO \* info)

Read the co-occurrence data from file. The format of the file is:

[rows][columns][row id+][column id+][w1 cos\_count (w21 c21) ... (w2n c2n)]+\*\*

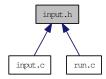
row and column ids are integer values that map to the original vocabulary. The number of values should be (info -> m) and (info -> n), respectively.

Every value is an unsigned integer in binary format, unless textmode is TRUE – if so, values are in text, separated by white space (tab).

Note: i indexes for rows (w1); j indexes for columns (w2)

## 4.9 input.h File Reference

This graph shows which files directly or indirectly include this file:



#### **Functions**

- void initializePostInput (INFO \*info)
- bool readCO (INFO \*info)

#### **4.9.1** Function Documentation

#### **4.9.1.1** void initializePostInput (INFO \* *info*)

Initialization that depends on the input file or parameters

#### 4.9.1.2 bool readCO (INFO \* info)

Read the co-occurrence data from file. The format of the file is:

[rows][columns][row id+][column id+][w1 cos\_count (w21 c21) ... (w2n c2n)]+\*\*

row and column ids are integer values that map to the original vocabulary. The number of values should be (info -> m) and (info -> n), respectively.

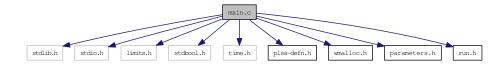
Every value is an unsigned integer in binary format, unless textmode is TRUE – if so, values are in text, separated by white space (tab).

Note: i indexes for rows (w1); j indexes for columns (w2)

#### 4.10 main.c File Reference

```
#include <stdlib.h>
#include <stdio.h>
#include <limits.h>
#include <stdbool.h>
#include <time.h>
#include "plsa-defn.h"
#include "wmalloc.h"
#include "parameters.h"
#include "run.h"
```

Include dependency graph for main.c:



#### **Functions**

• int main (int argc, char \*argv[])

#### **4.10.1** Function Documentation

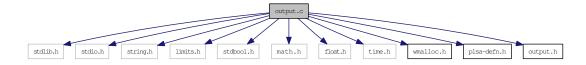
#### **4.10.1.1** int main (int *argc*, char \* *argv*[])

Main function

## 4.11 output.c File Reference

```
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <limits.h>
#include <stdbool.h>
#include <math.h>
#include <float.h>
#include <time.h>
#include "wmalloc.h"
#include "plsa-defn.h"
#include "output.h"
```

Include dependency graph for output.c:



#### **Functions**

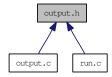
• void printCoProb (INFO \*info)

#### **4.11.1** Function Documentation

#### 4.11.1.1 void printCoProb (INFO \* info)

## 4.12 output.h File Reference

This graph shows which files directly or indirectly include this file:



#### **Functions**

• void printCoProb (INFO \*info)

#### **4.12.1** Function Documentation

4.12.1.1 void printCoProb (INFO \* info)

### 4.13 parameters.c File Reference

```
#include <getopt.h>
#include <stdlib.h>
#include <stdio.h>
#include <limits.h>
#include <string.h>
#include <stdbool.h>
#include "wmalloc.h"
#include "plsa-defn.h"
#include "parameters.h"
```

Include dependency graph for parameters.c:



#### **Defines**

• #define \_GNU\_SOURCE

#### **Functions**

- void usage (char \*progname)
- bool checkSettings (INFO \*info)
- bool processOptions (int argc, char \*argv[], INFO \*info)

#### **4.13.1** Define Documentation

4.13.1.1 #define \_GNU\_SOURCE

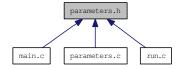
#### 4.13.2 Function Documentation

- **4.13.2.1** bool checkSettings (INFO \* *info*)
- 4.13.2.2 bool processOptions (int argc, char \* argv[], INFO \* info)
- 4.13.2.3 void usage (char \* progname)

Print out usage information

## 4.14 parameters.h File Reference

This graph shows which files directly or indirectly include this file:



#### **Functions**

- void usage (char \*progname)
- bool checkSettings (INFO \*info)
- bool processOptions (int argc, char \*argv[], INFO \*info)

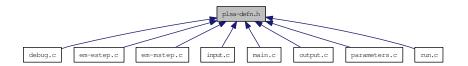
#### **4.14.1** Function Documentation

- 4.14.1.1 bool checkSettings (INFO \* info)
- 4.14.1.2 bool processOptions (int argc, char \* argv[], INFO \* info)
- 4.14.1.3 void usage (char \* progname)

Print out usage information

### 4.15 plsa-defn.h File Reference

This graph shows which files directly or indirectly include this file:



#### **Data Structures**

- struct cooccur
- struct info

#### **Defines**

- #define LN\_LIMIT 23.02585093
- #define MIN\_PROB (1.0E-24)
- #define DOLOG(X) (logf (X))
- #define DOEXP(X) (expf (X))
- #define DOLOGONE(X) (log1pf (X))
- #define DOLOG1PEXP(x) DOLOGONE(DOEXP(x))
- #define RANDOM\_FLOAT ((PROBNODE)rand () / ((PROBNODE)RAND\_MAX + (PROBNODE)1.0))
- #define  $DBL\_LESS(A, B)$  ((B A) >  $DBL\_EPSILON$ )
- #define ML\_DELTA 0.001
- #define MAINPROC 0
- #define ROUND\_DIGITS 100000000
- #define BLOCK\_SIZE(id, p, n) (BLOCK\_LOW ((id) + 1, p, n)-BLOCK\_LOW(id, p, n))
- #define PROGRESS\_MSG(A)
- #define FOPEN(FILENAME, FP, MODE)
- #define FCLOSE(FP) (void) fclose (FP);
- #define SET\_COS(W, X, Y, Z)
- #define GET\_COS(W, X) (info  $\rightarrow$  cos[W][X].x)
- #define GET\_COS\_POSITION(W, X) (info  $\rightarrow$  cos[W][X].column)
- $\bullet \ \ \text{\#define GET\_PROBW1\_Z}(X,\,Y) \ (info \rightarrow probw1\_z[X*info \rightarrow m+Y]) \\$
- #define GET\_PROBW2\_Z(X, Y) (info  $\rightarrow$  probw2\_z[X \* info  $\rightarrow$  n + Y])
- #define  $GET_PROBZ(X)$  (info  $\rightarrow$  probz[X])
- #define GET\_PROBZ\_W1W2(W, X, Y) (info  $\rightarrow$  probz\_w1w2[W][X \* info  $\rightarrow$  n + Y])
- #define logSumsInline(A, B)

#### **Typedefs**

- typedef double PROBNODE
- typedef struct cooccur COOCCUR
- typedef struct info INFO

#### **4.15.1** Define Documentation

#### 4.15.1.1 #define BLOCK\_SIZE(id, p, n) (BLOCK\_LOW ((id) + 1, p, n)-BLOCK\_LOW(id, p, n))

#### 4.15.1.2 #define $DBL\_LESS(A, B)$ ((B - A) > $DBL\_EPSILON$ )

Test if two double values are close to each other

#### **4.15.1.3** #define DOEXP(X) (expf (X))

Macro to perform the exp function

#### **4.15.1.4** #define DOLOG(X) (logf (X))

Macro to perform a log

#### 4.15.1.5 #define DOLOG1PEXP(x) DOLOGONE(DOEXP(x))

Macro to perform log (1 + expt(x))

#### 4.15.1.6 #define DOLOGONE(X) (log1pf (X))

Macro to perform log (1 + x)

#### 4.15.1.7 #define FCLOSE(FP) (void) fclose (FP);

#### 4.15.1.8 #define FOPEN(FILENAME, FP, MODE)

Value:

```
FP = fopen ((char*) FILENAME, MODE); \
  if (FP == NULL) {     fprintf (stderr, "Error %s %s.\n", (strcmp (MODE, "w") == 0) ? "creating" : "opening", FILENAME); \
     exit (EXIT_FAILURE); \
}
```

#### **4.15.1.9** #define GET\_COS(W, X) (info $\rightarrow$ cos[W][X].x)

Function to retrieve the position from the cooccurrence array

#### **4.15.1.10** #define GET\_COS\_POSITION(W, X) (info $\rightarrow$ cos[W][X].column)

Function to retrieve the cooccurrence count from the cooccurrence array

#### 4.15.1.11 #define GET PROBW1 Z(X, Y) (info $\rightarrow$ probw1 $z[X * info \rightarrow m + Y]$ )

Function to retrieve from P(w1|z); translate 2D to 1D co-ordinates

#### 4.15.1.12 #define GET\_PROBW2\_Z(X, Y) (info $\rightarrow$ probw2\_z[X \* info $\rightarrow$ n + Y])

Function to retrieve from P(w2|z); translate 2D to 1D co-ordinates

#### $\textbf{4.15.1.13} \quad \text{\#define GET\_PROBZ}(X) \ (info \rightarrow probz[X])$

Function to retrieve from P(z)

#### $\textbf{4.15.1.14} \quad \text{\#define GET\_PROBZ\_W1W2}(W, \ X, \ Y) \ (info \rightarrow probz\_w1w2[W][X*info \rightarrow n+Y])$

Function to retrieve from P(z|w1w2); translate 3D to 1D co-ordinates

#### 4.15.1.15 #define LN LIMIT 23.02585093

Accuracy of floating point values as a log (base e) value, multiplied by -1

#### 4.15.1.16 #define logSumsInline(A, B)

Value:

```
{
  register PROBNODE x, y; \
  if (A > B) {
    x = A; y = B; \
    }
  else {
    x = B; y = A; \
  }
  /* a > b */
  A = (fabs (y - x) > LN_LIMIT) ? x : x + DOLOG1PEXP (y - x); \
}
```

#### 4.15.1.17 #define MAINPROC 0

ID of the main processor is always 0

#### 4.15.1.18 #define MIN\_PROB (1.0E-24)

Minimum probability

#### 4.15.1.19 #define ML\_DELTA 0.001

Minimum difference between two maximum likelihoods

#### 4.15.1.20 #define PROGRESS MSG(A)

Value:

```
if (info -> verbose) {
  fprintf (stderr, "==\t%s\n", A); \
}
```

Define'd function to indicate program progress

# 4.15.1.21 #define RANDOM\_FLOAT ((PROBNODE)rand () / ((PROBNODE)RAND\_MAX + (PROBNODE)1.0))

Generate a random number between [0, 1); cast to floating point first to prevent overflow

#### 4.15.1.22 #define ROUND\_DIGITS 100000000

Number of digits to round; used when outputting to binary only

#### 4.15.1.23 #define SET\_COS(W, X, Y, Z)

Value:

```
{ \
  info -> cos[W][X].column = Y; \
  info -> cos[W][X].x = Z; \
}
```

Function to retrieve from the cooccurrence array

#### 4.15.2 Typedef Documentation

#### 4.15.2.1 typedef struct cooccur COOCCUR

#### 4.15.2.2 typedef struct info INFO

## 4.15.2.3 typedef double PROBNODE

Data type to use for probabilities

30 File Documentation

# 4.16 run.c File Reference

```
#include <stdlib.h>
#include <stdio.h>
#include <limits.h>
#include <stdbool.h>
#include <math.h>
#include <float.h>
#include <time.h>
#include <signal.h>
#include "wmalloc.h"
#include "plsa-defn.h"
#include "em-estep.h"
#include "em-mstep.h"
#include "input.h"
#include "output.h"
#include "parameters.h"
#include "debug.h"
#include "run.h"
```

Include dependency graph for run.c:



#### **Functions**

- INFO \* initialize ()
- void uninitialize (INFO \*info)
- bool run (INFO \*info)

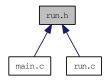
# 4.16.1 Function Documentation

- **4.16.1.1 INFO**\* initialize ()
- **4.16.1.2 bool run (INFO** \* *info*)
- **4.16.1.3** void uninitialize (INFO \* *info*)

4.17 run.h File Reference 31

# 4.17 run.h File Reference

This graph shows which files directly or indirectly include this file:



## **Functions**

- INFO \* initialize ()
- void uninitialize (INFO \*info)
- bool run (INFO \*info)

## **4.17.1** Function Documentation

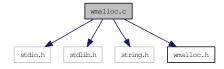
- **4.17.1.1 INFO**\* initialize ()
- **4.17.1.2 bool run (INFO** \* *info*)
- 4.17.1.3 void uninitialize (INFO \*info)

File Documentation

# 4.18 wmalloc.c File Reference

```
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include "wmalloc.h"
```

Include dependency graph for wmalloc.c:



#### **Functions**

- void \* wmalloc (size\_t y\_arg)
- void \* wrealloc (void \*x\_arg, size\_t y\_arg)
- void wfree (void \*x\_arg)
- static unsigned int hash (char \*v, signed int M)
- void initWMalloc ()
- void printWMalloc ()
- void printInUseWMalloc (void)
- void countMalloc (void \*ptr, size\_t amount, const char \*file, const unsigned int line)
- void countFree (void \*ptr)

#### **Variables**

- static unsigned int inuse\_malloc = 0
- static unsigned int max\_malloc = 0
- static WMSTRUCT \*\* wm\_array
- static char \* tempstr

#### **4.18.1** Function Documentation

- 4.18.1.1 void countFree (void \* ptr)
- 4.18.1.2 void countMalloc (void \* ptr, size\_t amount, const char \* file, const unsigned int line)
- **4.18.1.3** static unsigned int hash (char \*v, signed int M) [static]
- 4.18.1.4 void initWMalloc (void)
- 4.18.1.5 void printInUseWMalloc (void)
- 4.18.1.6 void printWMalloc (void)
- 4.18.1.7 void wfree (void  $*x\_arg$ )
- 4.18.1.8 void\* wmalloc (size\_t y\_arg)
- 4.18.1.9 void\* wrealloc (void \*  $x\_arg$ , size\_t  $y\_arg$ )

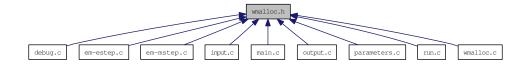
#### **4.18.2** Variable Documentation

- **4.18.2.1** unsigned int inuse\_malloc = 0 [static]
- **4.18.2.2** unsigned int max\_malloc = 0 [static]
- **4.18.2.3 char\* tempstr** [static]
- **4.18.2.4** WMSTRUCT\*\* wm\_array [static]

File Documentation

# 4.19 wmalloc.h File Reference

This graph shows which files directly or indirectly include this file:



#### **Data Structures**

• struct wmstruct

#### **Defines**

- #define WM\_SIZE 65536
- #define TEMPSTRLEN 80

# **Typedefs**

• typedef struct wmstruct WMSTRUCT

#### **Functions**

- void \* wmalloc (size\_t y\_arg)
- void \* wrealloc (void \*x\_arg, size\_t y\_arg)
- void wfree (void \*x\_arg)
- void initWMalloc (void)
- void printWMalloc (void)
- void printInUseWMalloc (void)
- void countMalloc (void \*ptr, size\_t amount, const char \*file, unsigned int line)
- void countFree (void \*ptr)

#### **4.19.1** Define Documentation

- 4.19.1.1 #define TEMPSTRLEN 80
- 4.19.1.2 #define WM\_SIZE 65536

## 4.19.2 Typedef Documentation

4.19.2.1 typedef struct wmstruct WMSTRUCT

#### **4.19.3** Function Documentation

- 4.19.3.1 void countFree (void \* ptr)
- 4.19.3.2 void countMalloc (void \* ptr, size\_t amount, const char \* file, unsigned int line)
- 4.19.3.3 void initWMalloc (void)
- 4.19.3.4 void printInUseWMalloc (void)
- 4.19.3.5 void printWMalloc (void)
- 4.19.3.6 void wfree (void  $*x\_arg$ )
- 4.19.3.7 void\* wmalloc (size\_t y\_arg)
- 4.19.3.8 void\* wrealloc (void \*  $x_arg$ , size\_t  $y_arg$ )

# **Index**

01777 00777 077	
_GNU_SOURCE	HAVE_INTTYPES_H, 12
parameters.c, 24	HAVE_LIMITS_H, 12
1.70.	HAVE_MALLOC, 12
applyEStep	HAVE_MEMORY_H, 12
em-estep.c, 15	HAVE_REALLOC, 12
em-estep.h, 16	HAVE_STDBOOL_H, 12
applyEStep_time	HAVE_STDINT_H, 12
info, 9	HAVE_STDLIB_H, 12
applyMStep	HAVE_STRING_H, 12
em-mstep.c, 17	HAVE_STRINGS_H, 12
em-mstep.h, 18	HAVE_SYS_STAT_H, 12
applyMStep_time	HAVE_SYS_TYPES_H, 12
info, 9	HAVE_UNISTD_H, 12
	PACKAGE, 12
base_fn	PACKAGE_BUGREPORT, 12
info, 7	PACKAGE_NAME, 12
block_end	PACKAGE_STRING, 12
info, 9	PACKAGE_TARNAME, 12
BLOCK SIZE	PACKAGE_VERSION, 12
plsa-defn.h, 27	STDC_HEADERS, 12
block_size	VERSION, 12
info, 9	COOCCUR
block_start	plsa-defn.h, 29
info, 8	cooccur, 5
	column, 5
calculateML	x, 5
em-estep.c, 15	cos
em-estep.h, 16	info, 8
calculateML_time	countFree
info, 9	wmalloc.c, 33
checkCoProb	wmalloc.e, 35 wmalloc.h, 35
debug.c, 13	countMalloc
debug.h, 14	
checkSettings	wmalloc.c, 33
parameters.c, 24	wmalloc.h, 35
parameters.h, 25	DDI TECC
, <del>*</del>	DBL_LESS
co_fn	plsa-defn.h, 27
info, 7	debug
column	info, 7
cooccur, 5	debug.c, 13
column_ids	checkCoProb, 13
info, 8	debugCheckCo, 13
config.h, 11	handler_sigfpe, 13
HAVE_BOOL, 12	debug.h, 14
HAVE_FLOAT_H, 12	checkCoProb, 14

debugCheckCo, 14	HAVE_BOOL
handler_sigfpe, 14	config.h, 12
debugCheckCo	HAVE_FLOAT_H
debug.c, 13	config.h, 12
	_
debug.h, 14	HAVE_INTTYPES_H
DOEXP	config.h, 12
plsa-defn.h, 27	HAVE_LIMITS_H
DOLOG	config.h, 12
plsa-defn.h, 27	HAVE_MALLOC
DOLOG1PEXP	config.h, 12
plsa-defn.h, 27	HAVE_MEMORY_H
DOLOGONE	config.h, 12
	_
plsa-defn.h, 27	HAVE_REALLOC
	config.h, 12
em-estep.c, 15	HAVE_STDBOOL_H
applyEStep, 15	config.h, 12
calculateML, 15	HAVE_STDINT_H
initEM, 15	config.h, 12
em-estep.h, 16	HAVE_STDLIB_H
-	
applyEStep, 16	config.h, 12
calculateML, 16	HAVE_STRING_H
initEM, 16	config.h, 12
em-mstep.c, 17	HAVE_STRINGS_H
applyMStep, 17	config.h, 12
normalizeProbs, 17	HAVE_SYS_STAT_H
em-mstep.h, 18	config.h, 12
applyMStep, 18	HAVE_SYS_TYPES_H
normalizeProbs, 18	config.h, 12
	HAVE_UNISTD_H
FCLOSE	config.h, 12
plsa-defn.h, 27	
file	INFO
wmstruct, 10	plsa-defn.h, 29
FOPEN	info, 6
plsa-defn.h, 27	applyEStep_time, 9
pisa-defii.ii, 27	applyMStep_time, 9
CET COS	
GET_COS	base_fn, 7
plsa-defn.h, 27	block_end, 9
GET_COS_POSITION	block_size, 9
plsa-defn.h, 27	block_start, 8
GET_PROBW1_Z	calculateML_time, 9
plsa-defn.h, 27	co_fn, <b>7</b>
GET PROBW2 Z	column_ids, 8
<del>-</del>	cos, 8
plsa-defn.h, 27	
GET_PROBZ	debug, 7
plsa-defn.h, 28	initEM_time, 9
GET_PROBZ_W1W2	iter, 8
plsa-defn.h, 28	m, 7
	maxiter, 7
handler_sigfpe	n, 7
debug.c, 13	no_output, 7
_	no_output, /
debug.h, 14	normaliza Droba tima
h a a h	normalizeProbs_time, 9
hash	num_clusters, 7
hash wmalloc.c, 33	

probw1_z, 8	MAINPROC		
probw2_z, 8	plsa-defn.h, 28		
probz, 8	max_malloc		
probz_w1w2, 8	wmalloc.c, 33		
program_end, 9	maxiter		
program_start, 9	info, 7		
readCO_time, 9	MIN_PROB		
rounding, 7	plsa-defn.h, 28		
row_ids, 8	ML_DELTA		
run_time, 9	plsa-defn.h, 28		
seed, 7			
sigfpe_count, 9	n		
textio, 7	info, 7		
verbose, 7	next		
world_id, 8	wmstruct, 10		
world_size, 8	no_output		
initEM	info, 7		
em-estep.c, 15	normalizeProbs		
em-estep.h, 16	em-mstep.c, 17		
initEM_time	em-mstep.h, 18		
info, 9	normalizeProbs_time		
initialize	info, 9		
run.c, 30	num clusters		
run.h, 31	info, <b>7</b>		
initializePostInput	,		
input.c, 19	output.c, 22		
input.h, 20	printCoProb, 22		
initWMalloc	output.h, 23		
wmalloc.c, 33	printCoProb, 23		
wmalloc.h, 35	•		
input.c, 19	PACKAGE		
initializePostInput, 19	config.h, 12		
readCO, 19	PACKAGE_BUGREPORT		
input.h, 20	config.h, 12		
initializePostInput, 20	PACKAGE_NAME		
readCO, 20	config.h, 12		
inuse_malloc	PACKAGE_STRING		
wmalloc.c, 33	config.h, 12		
iter	PACKAGE_TARNAME		
info, 8	config.h, 12		
1110, 0	PACKAGE_VERSION		
line	config.h, 12		
wmstruct, 10	parameters.c, 24		
LN_LIMIT	_GNU_SOURCE, 24		
plsa-defn.h, 28	checkSettings, 24		
logSumsInline	processOptions, 24		
plsa-defn.h, 28	usage, 24		
pion definit, 20	parameters.h, 25		
m	checkSettings, 25		
info, 7	processOptions, 25		
main	usage, 25		
main.c, 21	plsa-defn.h, 26		
main.c, 21	BLOCK_SIZE, 27		
main, 21	COOCCUR, 29		
1114111, 21	COOCCOR, 27		

DBL_LESS, 27	ptr
DOEXP, 27	wmstruct, 10
DOLOG, 27	
DOLOG1PEXP, 27	RANDOM_FLOAT
DOLOGONE, 27	plsa-defn.h, 29
FCLOSE, 27	readCO
FOPEN, 27	input.c, 19
GET_COS, 27	input.h, 20
GET_COS_POSITION, 27	readCO_time
GET_COS_TOSTHON, 27 GET PROBW1 Z, 27	info, 9
GET_ROBW1_Z, 27 GET_PROBW2_Z, 27	ROUND_DIGITS
GET_PROBX, 28	plsa-defn.h, 29
	-
GET_PROBZ_W1W2, 28	rounding
INFO, 29	info, 7
LN_LIMIT, 28	row_ids
logSumsInline, 28	info, 8
MAINPROC, 28	run
MIN_PROB, 28	run.c, 30
ML_DELTA, 28	run.h, 31
PROBNODE, 29	run.c, 30
PROGRESS_MSG, 28	initialize, 30
RANDOM_FLOAT, 29	run, 30
ROUND_DIGITS, 29	uninitialize, 30
SET_COS, 29	run.h, 31
printCoProb	initialize, 31
output.c, 22	run, 31
output.h, 23	uninitialize, 31
printCoProbs_time	run_time
info, 9	info, 9
printInUseWMalloc	11110, >
wmalloc.c, 33	seed
wmalloc.h, 35	info, 7
printWMalloc	SET_COS
wmalloc.c, 33	plsa-defn.h, 29
wmalloc.h, 35	sigfpe_count
PROBNODE	info, 9
plsa-defn.h, 29	size
probw1_z	wmstruct, 10
info, 8	STDC_HEADERS
probw2_z	config.h, 12
info, 8	
probz	tempstr
info, 8	wmalloc.c, 33
probz_w1w2	TEMPSTRLEN
info, 8	wmalloc.h, 35
processOptions	textio
parameters.c, 24	info, 7
parameters.h, 25	
program_end	uninitialize
info, 9	run.c, 30
program_start	run.h, 31
info, 9	usage
PROGRESS_MSG	parameters.c, 24
plsa-defn.h, 28	parameters.h, 25
r	r

verbose info, 7		wmalloc.c, 33 wmalloc.h, 35
VERSION		
config.h, 12	X	_
		cooccur, 5
wfree		
wmalloc.c, 33		
wmalloc.h, 35		
wm_array		
wmalloc.c, 33		
WM_SIZE		
wmalloc.h, 35		
wmalloc		
wmalloc.c, 33		
wmalloc.h, 35		
wmalloc.c, 32		
countFree, 33		
countMalloc, 33		
hash, 33		
initWMalloc, 33		
inuse_malloc, 33		
max_malloc, 33		
printInUseWMalloc, 33		
printWMalloc, 33		
tempstr, 33		
wfree, 33		
wm_array, 33		
wmalloc, 33		
wrealloc, 33		
wmalloc.h, 34		
countFree, 35		
countMalloc, 35		
initWMalloc, 35		
printInUseWMalloc, 35		
printWMalloc, 35		
TEMPSTRLEN, 35		
wfree, 35		
WM_SIZE, 35		
wmalloc, 35		
WMSTRUCT, 35		
wrealloc, 35		
WMSTRUCT		
wmalloc.h, 35		
wmstruct, 10		
file, 10		
line, 10		
next, 10		
ptr, 10		
size, 10		
world_id		
info, 8		
world_size		
info, 8		
wrealloc		
WICAHUC		