Package 'mldr.datasets'

May 8, 2026

Title R Ultimate Multilabel Dataset Repository

Version 0.4.2

Date 2019-01-16

Description Large collection of multilabel datasets along with the functions needed to export them to several formats, to make partitions, and to obtain bibliographic information.

URL https://github.com/fcharte/mldr.datasets

Depends R (>= 3.0.0)

Imports utils

License LGPL (>= 3) | file LICENSE

LazyData true

RoxygenNote 6.1.0

Suggests mldr

Encoding UTF-8

NeedsCompilation no

Author David Charte [cre] (ORCID: <https://orcid.org/0000-0002-4830-9512>), Francisco Charte [aut] (ORCID: <https://orcid.org/0000-0002-3083-8942>), Antonio J. Rivera [aut]

Maintainer David Charte <fdavidcl@ugr.es>

Repository CRAN

Date/Publication 2019-01-17 13:20:03 UTC

Contents

`available.mldrs` *Obtain additional datasets available to download*

Description

`available.mldrs` retrieves the most up to date list of additional datasets. Those datasets are not included into the package, but can be downloaded and saved locally.

Usage

```
available.mldrs()
```

Value

A data.frame with the available multilabel datasets

Examples

```
## Not run:
library(mldr.datasets)
names <- available.mldrs()$Name

## End(Not run)
```

`bibtex` *Dataset with BibTeX entries*

Description

Multilabel dataset from the text domain.

Usage

```
bibtex(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 7395 instances, 1836 attributes and 159 labels

Source

Katakis, I. and Tsoumakas, G. and Vlahavas, I., "Multilabel Text Classification for Automated Tag Suggestion", in Proc. ECML PKDD08 Discovery Challenge, Antwerp, Belgium, pp. 75-83, 2008

Examples

```
## Not run:
bibtex <- bibtex()  # Check and load the dataset
toBibtex(bibtex)
bibtex$measures

## End(Not run)
```

birds *Dataset with sounds produced by birds and the species they belong to*

Description

Multilabel dataset from the sound domain.

Usage

```
birds
```

Format

An mldr object with 645 instances, 260 attributes and 19 labels

Source

Briggs, F. and Lakshminarayanan, B. and Neal, L. and Fern, X. Z. and Raich, R. and Hadley, S. J. K. and Hadley, A. S. and Betts, M. G., "Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach", The Journal of the Acoustical Society of America, (6)131, pp. 4640–4650, 2012

Examples

```
## Not run:
toBibtex(birds)
birds$measures

## End(Not run)
```

bookmarks *Dataset with data from web bookmarks and their categories*

Description

Multilabel dataset from the text domain.

Usage

```
bookmarks(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 87856 instances, 2150 attributes and 208 labels

Source

Katakis, I. and Tsoumakas, G. and Vlahavas, I., "Multilabel Text Classification for Automated Tag Suggestion", in Proc. ECML PKDD08 Discovery Challenge, Antwerp, Belgium, pp. 75-83, 2008

Examples

```
## Not run:
bookmarks <- bookmarks()  # Check and load the dataset
toBibtex(bookmarks)
bookmarks$measures

## End(Not run)
```

---

cal500 *Dataset with music data along with labels for emotions, instruments, genres, etc.*

Description

Multilabel dataset from the music domain.

Usage

```
cal500
```

Format

An mldr object with 502 instances, 68 attributes and 174 labels

Source

Turnbull, Douglas and Barrington, Luke and Torres, David and Lanckriet, Gert, "Semantic annotation and retrieval of music and sound effects", Audio, Speech, and Language Processing, IEEE Transactions on, (2)16, pp. 467-476, 2008

Examples

```
## Not run:
toBibtex(cal500)
cal500$measures

## End(Not run)
```

check_n_load.mldr *(Defunct) Check if an mldr object is locally available and download it if needed*

Description

This function checks if the mldr object whose name is given as input is locally available, loading it in memory. If necessary, the dataset will be downloaded from the GitHub repository and saved locally.

Usage

```
check_n_load.mldr(mldr.name)
```

Arguments

mldr.name	Name of the dataset to load

Examples

```
## Not run:
library(mldr.datasets)
check_n_load.mldr("bibtex")
bibtex$measures

## End(Not run)
```

corel16k001 *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k001(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 13766 instances, 500 attributes and 153 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k001 <- corel16k001()  # Check and load the dataset
toBibtex(corel16k001)
corel16k001$measures

## End(Not run)
```

`corel16k002` *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k002(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 13761 instances, 500 attributes and 164 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k002 <- corel16k002()  # Check and load the dataset
toBibtex(corel16k002)
corel16k002$measures

## End(Not run)
```

corel16k003 *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k003(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 13760 instances, 500 attributes and 154 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k003 <- corel16k003()  # Check and load the dataset
toBibtex(corel16k003)
corel16k003$measures

## End(Not run)
```

corel16k004 *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k004(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 13837 instances, 500 attributes and 162 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k004 <- corel16k004()  # Check and load the dataset
toBibtex(corel16k004)
corel16k004$measures

## End(Not run)
```

---

corel16k005 *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k005(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 13847 instances, 500 attributes and 160 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k005 <- corel16k005()  # Check and load the dataset
toBibtex(corel16k005)
corel16k005$measures

## End(Not run)
```

`corel16k006` *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k006(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 13859 instances, 500 attributes and 162 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k006 <- corel16k006()  # Check and load the dataset
toBibtex(corel16k006)
corel16k006$measures

## End(Not run)
```

corel16k007 *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k007(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 13915 instances, 500 attributes and 174 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k007 <- corel16k007()  # Check and load the dataset
toBibtex(corel16k007)
corel16k007$measures

## End(Not run)
```

corel16k008 *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k008(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 13864 instances, 500 attributes and 168 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k008 <- corel16k008()  # Check and load the dataset
toBibtex(corel16k008)
corel16k008$measures

## End(Not run)
```

---

corel16k009 *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k009(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 13884 instances, 500 attributes and 173 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k009 <- corel16k009()  # Check and load the dataset
toBibtex(corel16k009)
corel16k009$measures

## End(Not run)
```

`corel16k010` *Datasets with data from the Corel image collection. There are 10 subsets in corel16k*

Description

Multilabel dataset from the image domain.

Usage

```
corel16k010(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 13618 instances, 500 attributes and 144 labels

Source

Barnard, K. and Duygulu, P. and Forsyth, D. and de Freitas, N. and Blei, D. M. and Jordan, M. I., "Matching words and pictures", Journal of Machine Learning Research, Vol. 3, pp. 1107–1135, 2003

Examples

```
## Not run:
corel16k010 <- corel16k010()  # Check and load the dataset
toBibtex(corel16k010)
corel16k010$measures

## End(Not run)
```

corel5k *Dataset with data from the Corel image collection*

Description

Multilabel dataset from the image domain.

Usage

```
corel5k(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 5000 instances, 499 attributes and 374 labels

Source

Duygulu, P. and Barnard, K. and de Freitas, J.F.G. and Forsyth, D.A., "Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary", Computer Vision, ECCV 2002, pp. 97-112, 2002

Examples

```
## Not run:
corel5k <- corel5k()  # Check and load the dataset
toBibtex(corel5k)
corel5k$measures

## End(Not run)
```

delicious *Dataset generated from the del.icio.us site bookmarks*

Description

Multilabel dataset from the text domain.

Usage

```
delicious(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 16105 instances, 500 attributes and 983 labels

Source

Tsoumakas, G. and Katakis, I. and Vlahavas, I., "Effective and Efficient Multilabel Classification in Domains with Large Number of Labels", in Proc. ECML/PKDD Workshop on Mining Multidimensional Data, Antwerp, Belgium, MMD08, pp. 30–44, 2008

Examples

```
## Not run:
delicious <- delicious()  # Check and load the dataset
toBibtex(delicious)
delicious$measures

## End(Not run)
```

`density` *Calculate the density level of the dataset*

Description

This function calculates the ratio of nonzero-valued elements over the total of elements.

Usage

```
density(mld)
```

Arguments

`mld`	An `"mldr"` object

Examples

```
library(mldr.datasets)
density(emotions)
```

emotions — *Dataset with features extracted from music tracks and the emotions they produce*

Description

Multilabel dataset from the music domain.

Usage

```
emotions
```

Format

An mldr object with 593 instances, 72 attributes and 6 labels

Source

Wieczorkowska, A. and Synak, P. and Ra's, Z., "Multi-Label Classification of Emotions in Music", Intelligent Information Processing and Web Mining, Vol. 35, Chap. 30, pp. 307-315, 2006

Examples

```
## Not run:
toBibtex(emotions)
emotions$measures

## End(Not run)
```

enron — *Dataset with email messages and the folders where the users stored them*

Description

Multilabel dataset from the text domain.

Usage

```
enron(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 1702 instances, 1001 attributes and 53 labels

Source

Klimt, B. and Yang, Y., "The Enron Corpus: A New Dataset for Email Classification Research", in Proc. ECML04, Pisa, Italy, pp. 217-226, 2004

Examples

```
## Not run:
enron <- enron()  # Check and load the dataset
toBibtex(enron)
enron$measures

## End(Not run)
```

`eurlexdc_test` *List with 10 folds of the test data from the EUR-Lex directory codes dataset*

Description

Multilabel dataset from the text domain.

Usage

```
eurlexdc_test(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 1935 instances, 5000 attributes and 412 labels

Source

Mencia, E. L. and Furnkranz, J., "Efficient pairwise multilabel classification for large-scale problems in the legal domain", Machine Learning and Knowledge Discovery in Databases, pp. 50–65, 2008

Examples

```
## Not run:
eurlexdc_test <- eurlexdc_test()  # Check and load the dataset
toBibtex(eurlexdc_test[[1]])
eurlexdc_test[[1]]$measures

## End(Not run)
```

`eurlexdc_tra` *List with 10 folds of the train data from the EUR-Lex directory codes dataset*

Description

Multilabel dataset from the text domain.

Usage

```
eurlexdc_tra(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 17413 instances, 5000 attributes and 412 labels

Source

Mencia, E. L. and Furnkranz, J., "Efficient pairwise multilabel classification for large-scale problems in the legal domain", Machine Learning and Knowledge Discovery in Databases, pp. 50–65, 2008

Examples

```
## Not run:
eurlexdc_tra <- eurlexdc_tra()  # Check and load the dataset
toBibtex(eurlexdc_test[[1]])
eurlexdc_test[[1]]$measures

## End(Not run)
```

eurlexev_test — *List with 10 folds of the test data from the EUR-Lex EUROVOC descriptors dataset*

Description

Multilabel dataset from the text domain.

Usage

```
eurlexev_test(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 1935 instances, 5000 attributes and 3993 labels

Source

Mencia, E. L. and Furnkranz, J., "Efficient pairwise multilabel classification for large-scale problems in the legal domain", Machine Learning and Knowledge Discovery in Databases, pp. 50–65, 2008

Examples

```
## Not run:
eurlexev_test <- eurlexev_test()  # Check and load the dataset
toBibtex(eurlexev_test[[1]])
eurlexev_test[[1]]$measures

## End(Not run)
```

eurlexev_tra — *List with 10 folds of the train data from the EUR-Lex EUROVOC descriptors dataset*

Description

Multilabel dataset from the text domain.

Usage

```
eurlexev_tra(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 17413 instances, 5000 attributes and 3993 labels

Source

Mencia, E. L. and Furnkranz, J., "Efficient pairwise multilabel classification for large-scale problems in the legal domain", Machine Learning and Knowledge Discovery in Databases, pp. 50–65, 2008

Examples

```
## Not run:
eurlexev_tra <- eurlexev_tra()  # Check and load the dataset
toBibtex(eurlexev_tra[[1]])
eurlexev_tra[[1]]$measures

## End(Not run)
```

---

eurlexsm_test *List with 10 folds of the test data from the EUR-Lex subject matters dataset*

Description

Multilabel dataset from the text domain.

Usage

```
eurlexsm_test(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 1935 instances, 5000 attributes and 201 labels

Source

Mencia, E. L. and Furnkranz, J., "Efficient pairwise multilabel classification for large-scale problems in the legal domain", Machine Learning and Knowledge Discovery in Databases, pp. 50–65, 2008

Examples

```
## Not run:
eurlexsm_test <- eurlexsm_test()  # Check and load the dataset
toBibtex(eurlexsm_test[[1]])
eurlexsm_test[[1]]$measures

## End(Not run)
```

`eurlexsm_tra` *List with 10 folds of the train data from the EUR-Lex subject matters dataset*

Description

Multilabel dataset from the text domain.

Usage

```
eurlexsm_tra(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 17413 instances, 5000 attributes and 201 labels

Source

Mencia, E. L. and Furnkranz, J., "Efficient pairwise multilabel classification for large-scale problems in the legal domain", Machine Learning and Knowledge Discovery in Databases, pp. 50–65, 2008

Examples

```
## Not run:
eurlexsm_tra <- eurlexsm_tra()  # Check and load the dataset
toBibtex(eurlexsm_tra[[1]])
eurlexsm_tra[[1]]$measures

## End(Not run)
```

flags *Dataset with features correspoinding to world flags*

Description

Multilabel dataset from the image domain.

Usage

```
flags
```

Format

An mldr object with 194 instances, 19 attributes and 7 labels

Source

Goncalves, E. C. and Plastino, A. and Freitas, A. A., "A genetic algorithm for optimizing the label ordering in multi-label classifier chains", Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on, pp. 469-476, 2013

Examples

```
## Not run:
toBibtex(flags)
flags$measures

## End(Not run)
```

genbase *Dataset with genes data and their functional expression*

Description

Multilabel dataset from the biology domain.

Usage

```
genbase
```

Format

An mldr object with 662 instances, 1186 attributes and 27 labels

Source

Diplaris, S. and Tsoumakas, G. and Mitkas, P. and Vlahavas, I., "Protein Classification with Multiple Algorithms", in Proc. 10th Panhellenic Conference on Informatics, Volos, Greece, PCI05, pp. 448–456, 2005

Examples

```
## Not run:
toBibtex(genbase)
genbase$measures

## End(Not run)
```

`get.mldr` *Get a multilabel dataset by name*

Description

`get.mldr` obtains a multilabel dataset, either by finding it inside the package data, in the download directory or by downloading it.

Usage

```
get.mldr(name, download.dir = if
  (is.null(getOption("mldr.download.dir"))) tempdir() else
  getOption("mldr.download.dir"))
```

Arguments

`name`	Name of the dataset to load
`download.dir`	The path to the download directory, can be also set through `options()`

Examples

```
## Not run:
library(mldr.datasets)
# customize the download directory
options(mldr.download.dir = "./datasets")
# retrieve the bibtex dataset, as an mldr object, into a variable
bibtex <- get.mldr("bibtex")
bibtex$measures

## End(Not run)
```

imdb *Dataset generated from the IMDB film database*

Description

Multilabel dataset from the text domain.

Usage

```
imdb(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 120919 instances, 1001 attributes and 28 labels

Source

Read, J. and Pfahringer, B. and Holmes, G. and Frank, E., "Classifier chains for multi-label classification", Machine Learning, (3)85, pp. 333-359, 2011

Examples

```
## Not run:
imdb <- imdb()  # Check and load the dataset
toBibtex(imdb)
imdb$measures

## End(Not run)
```

iterative.stratification.holdout *Hold-out partitioning of an mldr object*

Description

Iterative stratification

Implemented from the algorithm explained in: Konstantinos Sechidis, Grigorios Tsoumakas, and Ioannis Vlahavas. 2011. On the stratification of multi-label data. In Proceedings of the 2011 European conference on Machine learning and knowledge discovery in databases - Volume Part III (ECML PKDD'11), Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis (Eds.), Vol. Part III. Springer-Verlag, Berlin, Heidelberg, 145-158.

Usage

```
iterative.stratification.holdout(mld, p = 60, seed = 10,
  get.indices = FALSE)
```

Arguments

mld	The mldr object to be partitioned
p	The percentage of instances to be selected for the training partition
seed	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
get.indices	A logical value indicating whether to return lists of indices or lists of "mldr" objects

Value

An mldr.folds object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called train and test

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
parts.emotions <- iterative.stratification.holdout(emotions, p = 70)
summary(parts.emotions$train)
summary(parts.emotions$test)

## End(Not run)
```

---

iterative.stratification.kfolds

Partition an mldr object into k folds

Description

Iterative stratification

Implemented from the algorithm explained in: Konstantinos Sechidis, Grigorios Tsoumakas, and Ioannis Vlahavas. 2011. On the stratification of multi-label data. In Proceedings of the 2011 European conference on Machine learning and knowledge discovery in databases - Volume Part III (ECML PKDD'11), Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis (Eds.), Vol. Part III. Springer-Verlag, Berlin, Heidelberg, 145-158.

Usage

```
iterative.stratification.kfolds(mld, k = 5, seed = 10,
  get.indices = FALSE)
```

Arguments

mld	The mldr object to be partitioned
k	The number of folds to be generated. By default is 5
seed	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
get.indices	A logical value indicating whether to return lists of indices or lists of "mldr" objects

Value

An mldr.folds object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called train and test

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
folds.emotions <- iterative.stratification.kfolds(emotions)
summary(folds.emotions[[1]]$train)
summary(folds.emotions[[1]]$test)

## End(Not run)
```

---

iterative.stratification.partitions

Generic partitioning of an mldr object

Description

Iterative stratification

Implemented from the algorithm explained in: Konstantinos Sechidis, Grigorios Tsoumakas, and Ioannis Vlahavas. 2011. On the stratification of multi-label data. In Proceedings of the 2011 European conference on Machine learning and knowledge discovery in databases - Volume Part III (ECML PKDD'11), Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis (Eds.), Vol. Part III. Springer-Verlag, Berlin, Heidelberg, 145-158.

Usage

```
iterative.stratification.partitions(mld, is.cv = FALSE, r, seed = 10,
  get.indices = FALSE)
```

Arguments

mld	The mldr object to be partitioned
is.cv	Option to enable treatment of partitions as cross-validation test folds
r	A vector of percentages of instances to be selected for each partition
seed	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
get.indices	A logical value indicating whether to return lists of indices or lists of "mldr" objects

Value

An mldr.folds object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called train and test

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
parts.emotions <- iterative.stratification.partitions(emotions, r = c(35, 25, 40))
summary(parts.emotions[[2]])

## End(Not run)
```

---

langlog *Dataset with data from the Language forum discussion*

Description

Multilabel dataset from the text domain.

Usage

```
langlog
```

Format

An mldr object with 1460 instances, 1004 attributes and 75 labels

Source

Read, Jesse, "Scalable multi-label classification", University of Waikato, 2010

Examples

```
## Not run:
toBibtex(langlog)
langlog$measures

## End(Not run)
```

`mediamill` *Dataset with features extracted from video sequences and semantic concepts assigned as labels*

Description

Multilabel dataset from the video domain.

Usage

```
mediamill(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 43907 instances, 120 attributes and 101 labels

Source

Snoek, C. G. M. and Worring, M. and van Gemert, J. C. and Geusebroek, J. M. and Smeulders, A. W. M., "The challenge problem for automated detection of 101 semantic concepts in multimedia", in Proc. 14th ACM International Conference on Multimedia, MULTIMEDIA06, pp. 421-430, 2006

Examples

```
## Not run:
mediamill <- mediamill()  # Check and load the dataset
toBibtex(mediamill)
mediamill$measures

## End(Not run)
```

medical *Dataset generated from medical reports*

Description

Multilabel dataset from the text domain.

Usage

```
medical
```

Format

An mldr object with 978 instances, 1449 attributes and 45 labels

Source

Crammer, K. and Dredze, M. and Ganchev, K. and Talukdar, P. P. and Carroll, S., "Automatic Code Assignment to Medical Text", in Proc. Workshop on Biological, Translational, and Clinical Language Processing, Prague, Czech Republic, BioNLP07, pp. 129-136, 2007

Examples

```
## Not run:
toBibtex(medical)
medical$measures

## End(Not run)
```

mldrs *(Defunct) Obtain and show a list of additional datasets available to download*

Description

The function downloads from GitHub the most up to date list of additional datasets. Those datasets are not included into the package, but can be downloaded and saved locally.

Usage

```
mldrs()
```

Examples

```
## Not run:
library(mldr.datasets)
mldrs()

## End(Not run)
```

ng20 *Dataset with news messages and the news groups they belong to*

Description

Multilabel dataset from the text domain. The original name of the dataset is 20ng

Usage

```
ng20
```

Format

An mldr object with 19300 instances, 1006 attributes and 20 labels

Source

Ken Lang, "Newsweeder: Learning to filter netnews", in Proc. 12th International Conference on Machine Learning, pp. 331-339, 1995

Examples

```
## Not run:
toBibtex(ng20)
ng20$measures

## End(Not run)
```

nuswide_BoW *Dataset obtained from the NUS-WIDE database with BoW representation*

Description

Multilabel dataset from the image domain.

Usage

```
nuswide_BoW(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 269648 instances, 501 attributes and 81 labels

Source

Chua, Tat-Seng and Tang, Jinhui and Hong, Richang and Li, Haojie and Luo, Zhiping and Zheng, Yantao, "NUS-WIDE: a real-world web image database from National University of Singapore", in Proc. of the ACM international conference on image and video retrieval, pp. 48, 2009

Examples

```
## Not run:
nuswide_BoW <- nuswide_BoW()  # Check and load the dataset
toBibtex(nuswide_BoW)
nuswide_BoW$measures

## End(Not run)
```

`nuswide_VLAD` *Dataset obtained from the NUS-WIDE database with cVLAD+ representation*

Description

Multilabel dataset from the image domain.

Usage

```
nuswide_VLAD(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 269648 instances, 129 attributes and 81 labels

Source

Chua, Tat-Seng and Tang, Jinhui and Hong, Richang and Li, Haojie and Luo, Zhiping and Zheng, Yantao, "NUS-WIDE: a real-world web image database from National University of Singapore", in Proc. of the ACM international conference on image and video retrieval, pp. 48, 2009

Examples

```
## Not run:
nuswide_VLAD <- nuswide_VLAD()  # Check and load the dataset
toBibtex(nuswide_VLAD)
nuswide_VLAD$measures

## End(Not run)
```

ohsumed *Dataset generated from a subset of the Medline database*

Description

Multilabel dataset from the text domain.

Usage

```
ohsumed(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 13929 instances, 1002 attributes and 23 labels

Source

Joachims, Thorsten, "Text Categorization with Suport Vector Machines: Learning with Many Relevant Features", in Proc. 10th European Conference on Machine Learning, pp. 137-142, 1998

Examples

```
## Not run:
ohsumed <- ohsumed()  # Check and load the dataset
toBibtex(ohsumed)
ohsumed$measures

## End(Not run)
```

random.holdout *Hold-out partitioning of an mldr object*

Description

Random partitioning

Usage

```
random.holdout(mld, p = 60, seed = 10, get.indices = FALSE)
```

Arguments

mld	The mldr object to be partitioned
p	The percentage of instances to be selected for the training partition
seed	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
get.indices	A logical value indicating whether to return lists of indices or lists of "mldr" objects

Value

An mldr.folds object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called train and test

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
parts.emotions <- random.holdout(emotions, p = 70)
summary(parts.emotions$train)
summary(parts.emotions$test)

## End(Not run)
```

---

random.kfolds *Partition an mldr object into k folds*

Description

This method randomly partitions the given dataset into k folds, providing training and test partitions for each fold.

Usage

```
random.kfolds(mld, k = 5, seed = 10, get.indices = FALSE)
```

Arguments

mld	The mldr object to be partitioned
k	The number of folds to be generated. By default is 5
seed	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
get.indices	A logical value indicating whether to return lists of indices or lists of "mldr" objects

Value

An mldr.folds object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called train and test

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
folds.emotions <- random.kfolds(emotions)
summary(folds.emotions[[1]]$train)
summary(folds.emotions[[1]]$test)

## End(Not run)
```

---

random.partitions *Generic partitioning of an mldr object*

Description

Random partitioning

Usage

```
random.partitions(mld, is.cv = FALSE, r, seed = 10,
  get.indices = FALSE)
```

Arguments

mld	The mldr object to be partitioned
is.cv	Option to enable treatment of partitions as cross-validation test folds
r	A vector of percentages of instances to be selected for each partition
seed	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
get.indices	A logical value indicating whether to return lists of indices or lists of "mldr" objects

Value

An mldr.folds object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called train and test

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
parts.emotions <- random.partitions(emotions, r = c(35, 25, 40))
summary(parts.emotions[[2]])

## End(Not run)
```

`rcv1sub1` *Dataset from the Reuters corpus (subset 1)*

Description

Multilabel dataset from the text domain.

Usage

```
rcv1sub1(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 6000 instances, 47236 attributes and 101 labels

Source

Lewis, D. D. and Yang, Y. and Rose, T. G. and Li, F., "RCV1: A new benchmark collection for text categorization research", The Journal of Machine Learning Research, Vol. 5, pp. 361-397, 2004

Examples

```
## Not run:
rcv1sub1 <- rcv1sub1()  # Check and load the dataset
toBibtex(rcv1sub1)
rcv1sub1$measures

## End(Not run)
```

rcv1sub2 *Dataset from the Reuters corpus (subset 2)*

Description

Multilabel dataset from the text domain.

Usage

```
rcv1sub2(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 6000 instances, 47236 attributes and 101 labels

Source

Lewis, D. D. and Yang, Y. and Rose, T. G. and Li, F., "RCV1: A new benchmark collection for text categorization research", The Journal of Machine Learning Research, Vol. 5, pp. 361-397, 2004

Examples

```
## Not run:
rcv1sub2 <- rcv1sub2()  # Check and load the dataset
toBibtex(rcv1sub2)
rcv1sub2$measures

## End(Not run)
```

rcv1sub3 *Dataset from the Reuters corpus (subset 3)*

Description

Multilabel dataset from the text domain.

Usage

```
rcv1sub3(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 6000 instances, 47236 attributes and 101 labels

Source

Lewis, D. D. and Yang, Y. and Rose, T. G. and Li, F., "RCV1: A new benchmark collection for text categorization research", The Journal of Machine Learning Research, Vol. 5, pp. 361-397, 2004

Examples

```
## Not run:
rcv1sub3 <- rcv1sub3()  # Check and load the dataset
toBibtex(rcv1sub3)
rcv1sub3$measures

## End(Not run)
```

`rcv1sub4` *Dataset from the Reuters corpus (subset 4)*

Description

Multilabel dataset from the text domain.

Usage

```
rcv1sub4(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 6000 instances, 47229 attributes and 101 labels

Source

Lewis, D. D. and Yang, Y. and Rose, T. G. and Li, F., "RCV1: A new benchmark collection for text categorization research", The Journal of Machine Learning Research, Vol. 5, pp. 361-397, 2004

Examples

```
## Not run:
rcv1sub4 <- rcv1sub4()  # Check and load the dataset
toBibtex(rcv1sub4)
rcv1sub4$measures

## End(Not run)
```

rcv1sub5 *Dataset from the Reuters corpus (subset 5)*

Description

Multilabel dataset from the text domain.

Usage

```
rcv1sub5(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 6000 instances, 47235 attributes and 101 labels

Source

Lewis, D. D. and Yang, Y. and Rose, T. G. and Li, F., "RCV1: A new benchmark collection for text categorization research", The Journal of Machine Learning Research, Vol. 5, pp. 361-397, 2004

Examples

```
## Not run:
rcv1sub5 <- rcv1sub5()  # Check and load the dataset
toBibtex(rcv1sub5)
rcv1sub5$measures

## End(Not run)
```

reutersk500 *Dataset from the Reuters Corpus with the 500 most relevant features selected*

Description

Multilabel dataset from the text domain.

Usage

```
reutersk500(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 6000 instances, 500 attributes and 103 labels

Source

Read, Jesse, "Scalable multi-label classification", University of Waikato, 2010

Examples

```
## Not run:
reutersk500 <- reutersk500()  # Check and load the dataset
toBibtex(reutersk500)
reutersk500$measures

## End(Not run)
```

---

scene *Dataset from images with different natural scenes*

Description

Multilabel dataset from the image domain.

Usage

```
scene(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 2407 instances, 294 attributes and 6 labels

Source

Boutell, M. and Luo, J. and Shen, X. and Brown, C., "Learning multi-label scene classification", Pattern Recognition, (9)37, pp. 1757–1771, 2004

Examples

```
## Not run:
scene <- scene()
toBibtex(scene)
scene$measures

## End(Not run)
```

`slashdot` *Dataset generated from slashdot.org site entries*

Description

Multilabel dataset from the text domain.

Usage

```
slashdot
```

Format

An mldr object with 3782 instances, 1079 attributes and 22 labels

Source

Read, J. and Pfahringer, B. and Holmes, G. and Frank, E., "Classifier chains for multi-label classification", Machine Learning, (3)85, pp. 333–359, 2011

Examples

```
## Not run:
toBibtex(slashdot)
slashdot$measures

## End(Not run)
```

`sparsity` *Calculate the sparsity level of the dataset*

Description

This function calculates the ratio of zero-valued elements over the total of elements. It is useful to decide whether to export in a dense or sparse format.

Usage

```
sparsity(mld)
```

Arguments

mld	An "mldr" object

Examples

```
library(mldr.datasets)
sparsity(emotions)
```

---

stackex_chemistry *Dataset from the Stack Exchange's chemistry forum*

Description

Multilabel dataset from the text domain.

Usage

```
stackex_chemistry(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 6961 instances, 540 attributes and 175 labels

Source

Charte, Francisco and Rivera, Antonio J. and del Jesus, Maria J. and Herrera, Francisco, "QUINTA: A question tagging assistant to improve the answering ratio in electronic forums", in EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), IEEE, pp. 1-6, 2015

Examples

```
## Not run:
stackex_chemistry <- stackex_chemistry()  # Check and load the dataset
toBibtex(stackex_chemistry)
stackex_chemistry$measures

## End(Not run)
```

stackex_chess *Dataset from the Stack Exchange's chess forum*

Description

Multilabel dataset from the text domain.

Usage

```
stackex_chess
```

Format

An mldr object with 1675 instances, 585 attributes and 227 labels

Source

Charte, Francisco and Rivera, Antonio J. and del Jesus, Maria J. and Herrera, Francisco, "QUINTA: A question tagging assistant to improve the answering ratio in electronic forums", in EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), IEEE, pp. 1-6, 2015

Examples

```
## Not run:
toBibtex(stackex_chess)
stackex_chess$measures

## End(Not run)
```

stackex_coffee *Dataset from the Stack Exchange's coffee forum*

Description

Multilabel dataset from the text domain.

Usage

```
stackex_coffee(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 225 instances, 1763 attributes and 123 labels

Source

Charte, Francisco and Rivera, Antonio J. and del Jesus, Maria J. and Herrera, Francisco, "QUINTA: A question tagging assistant to improve the answering ratio in electronic forums", in EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), IEEE, pp. 1-6, 2015

Examples

```
## Not run:
stackex_coffee <- stackex_coffee()
toBibtex(stackex_coffee)
stackex_coffee$measures

## End(Not run)
```

---

stackex_cooking *Dataset from the Stack Exchange's cooking forum*

Description

Multilabel dataset from the text domain.

Usage

```
stackex_cooking(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 10491 instances, 577 attributes and 400 labels

Source

Charte, Francisco and Rivera, Antonio J. and del Jesus, Maria J. and Herrera, Francisco, "QUINTA: A question tagging assistant to improve the answering ratio in electronic forums", in EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), IEEE, pp. 1-6, 2015

Examples

```
## Not run:
stackex_cooking <- stackex_cooking()  # Check and load the dataset
toBibtex(stackex_cooking)
stackex_cooking$measures

## End(Not run)
```

stackex_cs *Dataset from the Stack Exchange's computer science forum*

Description

Multilabel dataset from the text domain.

Usage

```
stackex_cs(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 9270 instances, 635 attributes and 274 labels

Source

Charte, Francisco and Rivera, Antonio J. and del Jesus, Maria J. and Herrera, Francisco, "QUINTA: A question tagging assistant to improve the answering ratio in electronic forums", in EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), IEEE, pp. 1-6, 2015

Examples

```
## Not run:
stackex_cs <- stackex_cs()  # Check and load the dataset
toBibtex(stackex_cs)
stackex_cs$measures

## End(Not run)
```

stackex_philosophy *Dataset from the Stack Exchange's philosophy forum*

Description

Multilabel dataset from the text domain.

Usage

```
stackex_philosophy(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 3971 instances, 842 attributes and 233 labels

Source

Charte, Francisco and Rivera, Antonio J. and del Jesus, Maria J. and Herrera, Francisco, "QUINTA: A question tagging assistant to improve the answering ratio in electronic forums", in EUROCON 2015 - International Conference on Computer as a Tool (EUROCON), IEEE, pp. 1-6, 2015

Examples

```
## Not run:
stackex_philosophy <- stackex_philosophy()  # Check and load the dataset
toBibtex(stackex_philosophy)
stackex_philosophy$measures

## End(Not run)
```

---

stratified.holdout *Hold-out partitioning of an mldr object*

Description

Stratified partitioning

Implementation of the algorithm defined in: Charte, F., Rivera, A., del Jesus, M. J., & Herrera, F. (2016, April). On the impact of dataset complexity and sampling strategy in multilabel classifiers performance. In International Conference on Hybrid Artificial Intelligence Systems (pp. 500-511). Springer, Cham.

Usage

```
stratified.holdout(mld, p = 60, seed = 10, get.indices = FALSE)
```

Arguments

mld	The mldr object to be partitioned
p	The percentage of instances to be selected for the training partition
seed	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
get.indices	A logical value indicating whether to return lists of indices or lists of "mldr" objects

Value

An `mldr.folds` object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called `train` and `test`

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
parts.emotions <- stratified.holdout(emotions, p = 70)
summary(parts.emotions$train)
summary(parts.emotions$test)

## End(Not run)
```

`stratified.kfolds` *Partition an mldr object into k folds*

Description

This method partitions the given dataset into k folds using a stratified strategy, providing training and test partitions for each fold.

Implementation of the algorithm defined in: Charte, F., Rivera, A., del Jesus, M. J., & Herrera, F. (2016, April). On the impact of dataset complexity and sampling strategy in multilabel classifiers performance. In International Conference on Hybrid Artificial Intelligence Systems (pp. 500-511). Springer, Cham.

Usage

```
stratified.kfolds(mld, k = 5, seed = 10, get.indices = FALSE)
```

Arguments

`mld`	The `mldr` object to be partitioned
`k`	The number of folds to be generated. By default is 5
`seed`	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
`get.indices`	A logical value indicating whether to return lists of indices or lists of `"mldr"` objects

Value

An `mldr.folds` object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called `train` and `test`

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
folds.emotions <- stratified.kfolds(emotions)
summary(folds.emotions[[1]]$train)
summary(folds.emotions[[1]]$test)

## End(Not run)
```

`stratified.partitions` *Generic partitioning of an mldr object*

Description

Stratified partitioning

Generalization of the algorithm defined in: Charte, F., Rivera, A., del Jesus, M. J., & Herrera, F. (2016, April). On the impact of dataset complexity and sampling strategy in multilabel classifiers performance. In International Conference on Hybrid Artificial Intelligence Systems (pp. 500-511). Springer, Cham.

Usage

```
stratified.partitions(mld, is.cv = FALSE, r, seed = 10,
  get.indices = FALSE)
```

Arguments

`mld`	The `mldr` object to be partitioned
`is.cv`	Option to enable treatment of partitions as cross-validation test folds
`r`	A vector of percentages of instances to be selected for each partition
`seed`	The seed to initialize the random number generator. By default is 10. Change it if you want to obtain partitions containing different samples, for instance to use a 2x5 fcv strategy
`get.indices`	A logical value indicating whether to return lists of indices or lists of `"mldr"` objects

Value

An `mldr.folds` object. This is a list containing k elements, one for each fold. Each element is made up of two mldr objects, called `train` and `test`

Examples

```
## Not run:
library(mldr.datasets)
library(mldr)
parts.emotions <- stratified.partitions(emotions, r = c(35, 25, 40))
summary(parts.emotions[[2]])

## End(Not run)
```

`tmc2007` *Dataset from airplanes failures reports*

Description

Multilabel dataset from the text domain.

Usage

```
tmc2007(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 28596 instances, 49060 attributes and 22 labels

Source

Srivastava, A. N. and Zane-Ulman, B., "Discovering recurring anomalies in text reports regarding complex space systems", Aerospace Conference, pp. 3853-3862, 2005

Examples

```
## Not run:
tmc2007 <- tmc2007()  # Check and load the dataset
toBibtex(tmc2007)
tmc2007$measures

## End(Not run)
```

tmc2007_500 *Dataset from airplanes failures reports (500 most relevant features extracted)*

Description

Multilabel dataset from the text domain.

Usage

```
tmc2007_500(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 28596 instances, 500 attributes and 22 labels

Source

Srivastava, A. N. and Zane-Ulman, B., "Discovering recurring anomalies in text reports regarding complex space systems", Aerospace Conference, pp. 3853-3862, 2005

Examples

```
## Not run:
tmc2007_500 <- tmc2007_500()  # Check and load the dataset
toBibtex(tmc2007_500)
tmc2007_500$measures

## End(Not run)
```

toBibtex.mldr *BibTeX entry associated to an mldr object*

Description

Gets the content of the bibtex member of the mldr object and returns it

Usage

```
## S3 method for class 'mldr'
toBibtex(object, ...)
```

Arguments

object	The mldr object whose BibTeX entry is needed
...	Additional parameters from the generic toBibtex function not used by toBibtex.mldr

Value

A string with the BibTeX entry

Examples

```
## Not run:
library(mldr.datasets)
cat(toBibtex(emotions))

## End(Not run)
```

---

write.mldr *Export an mldr object or set of mldr objects to different file formats*

Description

Writes one or more files in the specified formats with the content of the `mldr` or `mldr.folds` given as parameter

Usage

```
write.mldr(mld, format = c("MULAN", "MEKA"), sparse = FALSE,
  basename = ifelse(!is.null(mld$name) && nchar(mld$name) > 0,
  regmatches(mld$name, regexpr("(\\w)+", mld$name)), "unnamed_mldr"),
  noconfirm = FALSE, ...)
```

Arguments

mld	The `mldr/mldr.folds` object to be exported
format	A vector of strings stating the desired file formats. It can contain the values `'MULAN'`, `'MEKA'`, `'KEEL'`, `'CSV'` and `'LIBSVM'`
sparse	Boolean value indicating if sparse representation has to be used for ARFF-based file formats
basename	Base name for the files. `'unnamed_mldr'` is used by default
noconfirm	Use TRUE to skip confirmation of file writing
...	Additional options for the exporting functions (e.g. `chunk_size`, the number of instances to write at a time)

Examples

```
## Not run:
library(mldr.datasets)
write.mldr(emotions, format = c('CSV', 'KEEL'))

## End(Not run)
```

---

yahoo_arts *Dataset generated from the Yahoo! web site index (arts category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_arts(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 7484 instances, 23146 attributes and 26 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_arts <- yahoo_arts()  # Check and load the dataset
toBibtex(yahoo_arts)
yahoo_arts$measures

## End(Not run)
```

yahoo_business *Dataset generated from the Yahoo! web site index (business category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_business(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 11214 instances, 21924 attributes and 30 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_business <- yahoo_business()  # Check and load the dataset
toBibtex(yahoo_business)
yahoo_business$measures

## End(Not run)
```

yahoo_computers *Dataset generated from the Yahoo! web site index (computers category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_computers(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 12444 instances, 34096 attributes and 33 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_computers <- yahoo_computers()  # Check and load the dataset
toBibtex(yahoo_computers)
yahoo_computers$measures

## End(Not run)
```

---

yahoo_education *Dataset generated from the Yahoo! web site index (arts education)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_education(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 12030 instances, 27534 attributes and 33 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_education <- yahoo_education()  # Check and load the dataset
toBibtex(yahoo_education)
yahoo_education$measures

## End(Not run)
```

`yahoo_entertainment` *Dataset generated from the Yahoo! web site index (arts entertainment)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_entertainment(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 12730 instances, 32001 attributes and 21 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_entertainment <- yahoo_entertainment()  # Check and load the dataset
toBibtex(yahoo_entertainment)
yahoo_entertainment$measures

## End(Not run)
```

yahoo_health *Dataset generated from the Yahoo! web site index (health category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_health(...)
```

Arguments

...	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 8205 instances, 30605 attributes and 32 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_health <- yahoo_health()  # Check and load the dataset
toBibtex(yahoo_health)
yahoo_health$measures

## End(Not run)
```

yahoo_recreation *Dataset generated from the Yahoo! web site index (recreation category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_recreation(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 12828 instances, 30324 attributes and 22 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_recreation <- yahoo_recreation()  # Check and load the dataset
toBibtex(yahoo_recreation)
yahoo_recreation$measures

## End(Not run)
```

---

yahoo_reference *Dataset generated from the Yahoo! web site index (reference category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_reference(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 8027 instances, 39679 attributes and 33 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_reference <- yahoo_reference()  # Check and load the dataset
toBibtex(yahoo_reference)
yahoo_reference$measures

## End(Not run)
```

`yahoo_science` *Dataset generated from the Yahoo! web site index (science category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_science(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 6428 instances, 37187 attributes and 40 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_science <- yahoo_science()  # Check and load the dataset
toBibtex(yahoo_science)
yahoo_science$measures

## End(Not run)
```

yahoo_social *Dataset generated from the Yahoo! web site index (social category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_social(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 12111 instances, 52350 attributes and 39 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_social <- yahoo_social()  # Check and load the dataset
toBibtex(yahoo_social)
yahoo_social$measures

## End(Not run)
```

yahoo_society *Dataset generated from the Yahoo! web site index (society category)*

Description

Multilabel dataset from the text domain.

Usage

```
yahoo_society(...)
```

Arguments

...	Additional options for the loading function (e.g. download.dir)

Format

An mldr object with 14512 instances, 31802 attributes and 27 labels

Source

Ueda, N. and Saito, K., "Parametric mixture models for multi-labeled text", Advances in neural information processing systems, pp. 721–728, 2002

Examples

```
## Not run:
yahoo_society <- yahoo_society()  # Check and load the dataset
toBibtex(yahoo_society)
yahoo_society$measures

## End(Not run)
```

---

yeast *Dataset with protein profiles and their categories*

Description

Multilabel dataset from the biology domain.

Usage

```
yeast(...)
```

Arguments

`...`	Additional options for the loading function (e.g. `download.dir`)

Format

An mldr object with 2417 instances, 103 attributes and 14 labels

Source

Elisseeff, A. and Weston, J., "A Kernel Method for Multi-Labelled Classification", Advances in Neural Information Processing Systems, Vol. 14, pp. 681–687, 2001

Examples

```
## Not run:
yeast <- yeast()  # Check and load the dataset
toBibtex(yeast)
yeast$measures

## End(Not run)
```

Index