

quantedaData

September 17, 2014

Type Package

Title Data files for quanteda

Version 0.9

Date 2014-09-15

Author Ken Benoit and Paul Nulty

Maintainer Ken Benoit <kbenoit@lse.ac.uk> and Paul Nulty <p.nulty@lse.ac.uk>

Description Data files to accompany the quanteda text analysis package

Encoding UTF-8

License GPL-3

Requires quanteda

Depends quanteda

R topics documented:

amicusCorpus	2
exampleString	2
iebudgets	2
LBGexample	3
movies	4
quantedaData	4
ukManifestos	4
Index	5

amicusCorpus	<i>amicus curiae</i> briefs from <i>Bakke</i> (1978) and <i>Bollinger</i> (2008)
--------------	--

Description

Texts of amicus curiae briefs labelled as being either pro-petitioner or pro-respondent in US Supreme court cases on affirmative action, *Bakke* (1978) and *Bollinger* (2008), taken from Evans et al (2007).

Format

amicusCorpus is corpus with 100 texts, and docvar set for petitioner/respondent

amicusTexts is the same set of texts as a named character vector

References

Evans, Michael, et al. 2007. "Recounting the courts? Applying automated content analysis to enhance empirical legal research." *Journal of Empirical Legal Studies* 4(4): 1007-1039.

exampleString	<i>A paragraph of text for testing various text-based functions</i>
---------------	---

Description

This is a long paragraph (2,914 characters) of text taken from an Irish budget speech by Joe Higgins

Format

character vector with one element

Examples

```
data(exampleString)
clean(exampleString)
```

iebudgets	<i>Irish budget speeches</i>
-----------	------------------------------

Description

Speeches and document-level variables from Irish budget debates held annually, for the years 2008-2012.

Usage

```
data(iebudgets)
data(ieTexts)
data(ieDocvars)
data(ie2010Corpus)
```

Format

iebudgets is the corpus object for the 2008-2012 budget speeches, with document-level variables for year, debate, serial number, first and last name of the speaker, and the speaker's party.

ieTexts is a character vector of five budget speeches for experimenting with corpus construction.

ieDocvars is a data frame of document-level variables for experimenting with corpus construction.

ie2010Corpus is the subset for the 2010 budget speeches from Lowe and Benoit (2013)

References

Alexander Herzog and Kenneth Benoit. 2013. "**The Most Unkindest Cuts: Government Cohesion and Economic Crisis.**" Paper prepared for presentation at the 3rd Annual General Conference of the European Political Science Association, 20-22 June 2013, Barcelona.

Lowe, Will and Kenneth Benoit. 2013. "" *Political Analysis* 21: 298–313.

Examples

```
# load the Irish budget speeches
data(iebudgets)
ie2010corp <- subset(iebudgets, year==2010) # just the 2010 speeches
summary(ie2010corp)
summary(subset(iebudgets, no=="02")) # just the Finance Minister (always first)
                                     # note that "no" is a factor, not integer

# create a corpus (just five speeches)
mycorpus <- corpus(ieTexts, docvars=ieDocvars)
summary(mycorpus)
```

LBGexample

Example data from Laver Benoit and Garry (2003)

Description

Example data to demonstrate the Wordscores algorithm, from Laver Benoit and Garry (2003)

Format

A **dfm** object with 6 documents and 37 features

Details

This is the example word count data from Laver, Benoit and Garry's (2003) Table 1. Documents R1 to R5 are assumed to have known positions: -1.5, -0.75, 0, 0.75, 1.5. Document V1 is assumed unknown, and will have a raw text score of approximately -0.45 when computed as per LBG (2003).

References

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "**Estimating policy positions from political text using words as data.**" *American Political Science Review* 97(2): 311-331.

movies	<i>movie reviews from Pang, Lee, and Vaithyanathan (2002)</i>
--------	---

Description

A corpus object containing 2000 movie reviews classified by positive or negative sentiment

References

<http://dl.acm.org/citation.cfm?id=1118704>

quantedaData	<i>An R package containing texts and corpora for quanteda.</i>
--------------	--

Description

A set of texts and corpus objects for use with the quanteda R package.

Author(s)

Ken Benoit and Paul Nulty

See Also

[quanteda](#)

ukManifestos	<i>A corpus object containing 105 UK Manifestos</i>
--------------	---

Description

A corpus object containing 105 UK Manifestos from 1945-2005, with party and year attributes

Index

amicus (amicusCorpus), [2](#)

amicusCorpus, [2](#)

dfm, [3](#)

exampleString, [2](#)

ie2010Corpus (iebudgets), [2](#)

iebudgets, [2](#)

ieDocvars, (iebudgets), [2](#)

ieTexts, (iebudgets), [2](#)

LBGexample, [3](#)

movies, [4](#)

quanteda, [4](#)

quantedaData, [4](#)

quantedaData-package (quantedaData), [4](#)

ukManifestos, [4](#)