

# quantedaData

September 19, 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
ieTextsHeaders . . . . .	2
movies . . . . .	3
quantedaData . . . . .	4
<b>Index</b>	<b>5</b>

---

amicusCorpus

*amicus curiae* briefs from *Bakke* (1978) and *Bollinger* (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

*A paragraph of text for testing various text-based functions*

---

### 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)
```

---

ieTextsHeaders

*Irish budget speeches*

---

### Description

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

## 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)
```

---

`movies`

*movie reviews from Pang, Lee, and Vaithyanathan (2002)*

---

## Description

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

## References

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

---

quantedaData

*An R package containing texts and corpora for [quanteda](#).*

---

**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](#)

# Index

amicus (amicusCorpus), [2](#)  
amicusCorpus, [2](#)  
amicusTexts (amicusCorpus), [2](#)  
  
exampleString, [2](#)  
  
ie2010Corpus (ieTextsHeaders), [2](#)  
iebudgets (ieTextsHeaders), [2](#)  
ieDocvars (ieTextsHeaders), [2](#)  
ieDocvars, (ieTextsHeaders), [2](#)  
ieTexts (ieTextsHeaders), [2](#)  
ieTexts, (ieTextsHeaders), [2](#)  
ieTextsHeaders, [2](#)  
  
movies, [3](#)  
  
quanteda, [4](#)  
quantedaData, [4](#)  
quantedaData-package (quantedaData), [4](#)