Package 'arpa'

July 19, 2015

Type Package

Title Parses ARPA language model files

Version 0.1

Date 2015-07-18

Author Adam Acosta

Maintainer Adam Acosta <adam.acosta@gatech.edu>

Description Parse language models read from ARPA files to R objects.

Depends hash, stringi

Suggests testthat

License MIT

URL https://github.com/adamacosta/arpa/wiki

NeedsCompilation no

R topics documented:

`bigrams,ngram.model-method`

Return the bigrams of a language model

Description

Return the bigrams of a language model

Usage

```
## S4 method for signature 'ngram.model'
bigrams(object)
```

Arguments

object	An ngram.model object

Value

bigrams The bigrams of the model

Author(s)

Adam Acosta

---

contains,ngram.model,character-method

Tests whether or not a language model contains an ngram

Description

Tests whether or not a language model contains an ngram

Usage

```
## S4 method for signature 'ngram.model,character'
contains(object, key)
```

Arguments

object	An ngram.model object
key	A character string

Value

boolean Whether or not the string is in the language model

Author(s)

Adam Acosta

ngram.model-class *An S4 class to represent an ngram language model*

Description

Provides an efficient, fast way to map ngrams in a language model to their log probability, allowing for easy next-word prediction.

Slots

`unigrams` A hash table mapping unigrams to their log probability

`bigrams` A hash table mapping bigrams to their log probability

`trigrams` A hash table mapping trigrams to their log probability

Author(s)

Adam Acosta

read.arpa *Read ARPA file*

Description

Reads an ARPA file and returns the language model. See http://www.speech.sri.com/projects/srilm/manpages/ngram-format.5.html for a description of the ARPA file format.

Usage

```
read.arpa(input = "", header = TRUE, verbose = FALSE, nrow = -1L,
  skip = 0L, ugrams = -1L, bgrams = -1L, tgrams = -1L)
```

Arguments

`input`	A filename
`header`	boolean indicating whether or not the file has a header. Default is TRUE
`verbose`	A boolean indicating whether you want the function to print information to the console as it is parsing the file. Default is FALSE.
`nrow`	The number of rows in the file. This is optional, as the format itself dictates that the file header must give the number of each ngram, from which the number of lines in the file can be inferred. If this is passed as a parameter, pass the number of unigrams, bigrams, and trigrams as well, and it will speed up the parsing process.
`skip`	The number of rows, if any, to skip in the file. Default is 0.

ugrams	See below
bgrams	See below
tgrams	integer values indicating the number of unigrams, bigrams, and trigrams. Should be in the header but can be passed to the function directly to avoid header parsing overhead.

Value

An ngram.model object, stored internally as a list of three hash tables mapping each ngram to its log probability.

Author(s)

Adam Acosta

---

trigrams,ngram.model-method

Return the trigrams of a language model

Description

Return the trigrams of a language model

Usage

```
## S4 method for signature 'ngram.model'
trigrams(object)
```

Arguments

object	An ngram.model object

Value

trigrams The trigrams of the model

Author(s)

Adam Acosta

`unigrams,ngram.model-method`
Return the unigrams of a language model

Description

Return the unigrams of a language model

Usage

```
## S4 method for signature 'ngram.model'
unigrams(object)
```

Arguments

`object`	An ngram.model object

Value

unigrams The unigrams of the model

Author(s)

Adam Acosta