

The stoRy package: a general overview

Paul Sheridan and Mikael Onsjö

Last updated: May 9, 2017

The `stoRy` package implements the hypergeometric test for over-representation of literary themes in a storyset (a list of stories) relative to a background set of stories. The package is currently implemented for the analysis of over-represented literary themes in the Star Trek TV franchise series The Original Series (TOS), The Animated Series (TAS), and The Next Generation (TNG).

1 Installation

The package is hosted on CRAN and can be installed by running the command

```
install.packages("stoRy")
```

Once installed, the package can be loaded by running the standard `library` command

```
library("stoRy")
```

2 Accessing documentation

Each function in the package is documented. Run the command

```
help(package="stoRy")
```

to see a brief overview of the package functions. The package vignette can be accessed by running

```
vignette(package="stoRy")
```

Function help files can be accessed using the usual R command. For example, help on the method `TEA` can be obtained with

```
?stoRy
```

3 Example: Star Trek TV franchise episode themes

The *stoRy* package contains themed Star Trek episodes with additional meta-data for the series The Original Series (TOS), The Animated Series (TAS), and The Next Generation (TNG). A *story* is created in order to view the themes (and meta-data) for a particular episode of interest. For example, to view the data for the classic TOS episode 'Arena' (story ID TOS1x19), initialize a *story* object as follows

```
story_id <- "TOS1x19"
mystory <- story$new(story_id)
print(mystory)
```

Themes and other meta-data can be added and removed as desired. For example, run the following command to add the theme "pride" as a minor theme

```
mystory$add_theme(theme = "pride", level = "minor")
```

The theme can be summarily removed as follows

```
mystory$remove_theme(theme = "pride")
```

Settings and keywords can be added and removed in a similar manner

```
mystory$add_setting(setting = "candy shop")
mystory$remove_setting(setting = "candy shop")
mystory$add_keyword(keyword = "Captain Kirk is climbing a mountain")
```

It is also possible to examine the individual themes contained in the *stoRy* package collection. For example, to examine the theme "utopia" run the following command sequence

```
theme_name <- "utopia"
mytheme <- theme$new(theme_name)
print(mytheme)
```

The output includes the theme definition, an illustrative example, and its place in a theme hierarchy.

4 Example: Enriched themes in the Star Trek TV series

This section is devoted to an example usage of the TEA hypergeometric test for theme over-representation. In the analysis, the hypergeometric test is applied to identify over-represented themes in each series TOS, TAS, and TNG relative to the background TOS/TAS/TNG. First, read in the 'series.smt' *storysets* file

```
file <- system.file("storysets", "series.smt", package = "stoRy")
mystorysets <- storysets$new(file)
print(mystorysets)
```

The TEA function performs the calculation relative to the background TOS/TAS/TNG by default

```
results <- TEA(mystorysets)
```

The results can be accessed as follows

```
results$TOS[1:10,c("Theme", "NSample", "NOmega", "Pvalue", "Padj")]
results$TAS[1:10,c("Theme", "NSample", "NOmega", "Pvalue", "Padj")]
results$TNG[1:10,c("Theme", "NSample", "NOmega", "Pvalue", "Padj")]
```

The columns 'NSample', 'NOmega', 'Pvalue' and 'Padj' are the number of occurrences of a theme in the storyset under examination, the number of occurrences of a theme in the background storyset, the hypergeometric test P value, and the Benjamini-Hochberg corrected P value.