

Package: poddr (via r-universe)

October 31, 2024

Title Collect Metadata for Selected Podcasts

Version 0.2.5

Description Collecting all the data, but just for The Incomparable,
Relay.fm and ATP.

License MIT + file LICENSE

URL <https://github.com/jemus42/poddr>

BugReports <https://github.com/jemus42/poddr/issues>

Imports checkmate, cli, dplyr, fs, glue, hms, lubridate, polite,
progress, purrr, readr, rlang (>= 0.1.2), rvest, stringr,
tibble, tidyr

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Repository <https://jemus42.r-universe.dev>

RemoteUrl <https://github.com/jemus42/poddr>

RemoteRef HEAD

RemoteSha b1ddfe70ae458fbbdabc1324230b3d6808120a6c

Contents

atp_get_episodes	2
atp_parse_page	3
cache_podcast_data	3
gather_people	4
incomparable_get_episodes	5
incomparable_get_shows	5
incomparable_get_subcategories	6
incomparable_parse_archive	7
incomparable_parse_stats	8
label_n	8
parse_duration	9

relay_get_episodes	10
relay_get_shows	10
relay_parse_feed	11
update_cached_data	11

Index	13
--------------	-----------

atp_get_episodes	<i>Retrieve ATP episodes</i>
------------------	------------------------------

Description

Retrieve ATP episodes

Usage

```
atp_get_episodes(page_limit = NULL, cache = TRUE)
```

Arguments

page_limit	Number of pages to scrape, from newest to oldest episode. Page 1 contains the 5 most recent episodes, and subsequent pages contain 50 episodes per page. As of December 2020, there are 10 pages total. Pass NULL (default) to get all pages.
cache	(logical(1)) Set to FALSE to disable caching.

Value

A tibble.

Examples

```
## Not run:
# Only the first page with the newest 5 episodes
atp_new <- atp_get_episodes(page_limit = 1)

# The latest and then 50 more
atp_latest <- atp_get_episodes(page_limit = 2)

# Get all episodes (use wisely)
atp_full <- atp_get_episodes()

## End(Not run)
```

atp_parse_page	<i>Parse a single ATP page</i>
----------------	--------------------------------

Description

Parse a single ATP page

Usage

```
atp_parse_page(page)
```

Arguments

page Scraped page object, e.g. from `polite::scrape()`.

Value

A tibble.

Examples

```
## Not run:
session <- polite::bow(url = "https://atp.fm")
page <- polite::scrape(session, query = list(page = 1))
atp_parse_page(page)

## End(Not run)
```

cache_podcast_data	<i>Cache episode data</i>
--------------------	---------------------------

Description

Cache episode data

Usage

```
cache_podcast_data(x, dir = "data_cache", filename = NULL, csv = TRUE)
```

Arguments

x Object to cache.
dir ["data_cache"] Directory to save data to.
filename Optional filename sans extension, if not specified the name of x is used.
csv If TRUE (default), also saves a CSV file with the same base name.

Value

Nothing

Examples

```
## Not run:  
atp_new <- atp_get_episodes(page_limit = 1)  
cache_podcast_data(atp_new, csv = FALSE)  
  
## End(Not run)
```

gather_people	<i>Gather episode datasets by people</i>
---------------	--

Description

A thin wrapper around `tidyr::pivot_longer()` and `tidyr::separate_rows()`.

Usage

```
gather_people(episodes)
```

Arguments

`episodes` A tibble containing host and guest columns, with names separated by ;.

Value

A tibble with new columns "role" and "person", one row per person.

Examples

```
## Not run:  
incomparable <- incomparable_get_episodes(incomparable_get_shows())  
incomparable_wide <- gather_people(incomparable)  
  
## End(Not run)
```

`incomparable_get_episodes`*Retrieve all episodes for The Incomparable shows*

Description

This combines `incomparable_parse_stats()` and `incomparable_parse_archive()` to retrieve full episode information including host/guest, durations including seconds, podcast subcategories and topics. Use sparingly to limit unnecessarily hammering the poor webserver!

Usage

```
incomparable_get_episodes(incomparable_shows, cache = TRUE)
```

Arguments

`incomparable_shows` Dataset of shows with title and URLs as returned by `incomparable_get_shows()`.

`cache` (logical(1)) Set to FALSE to disable caching.

Value

A tibble with one row per episode.

Examples

```
## Not run:  
incomparable_shows <- incomparable_get_shows()  
incomparable <- incomparable_get_episodes(incomparable_shows)  
  
## End(Not run)
```

`incomparable_get_shows`*Retrieve all The Incomparable shows*

Description

Parses the show overview page and returns a tibble of show names with corresponding URLs, which in turn can then be passed to `incomparable_parse_archive()` and `incomparable_parse_stats()` individually.

Usage

```
incomparable_get_shows(cache = TRUE)
```

Arguments

cache (logical(1)) Set to FALSE to disable caching.

Value

A tibble with following columns:

```
Columns: 4
 $ show      <chr>
 $ stats_url <glue>
 $ archive_url <glue>
 $ status    <chr>
```

Examples

```
## Not run:
incomparable_get_shows()

## End(Not run)
```

incomparable_get_subcategories
Extract subcategory index for given show

Description

Not actively used in other functions but could come in handy.

Usage

```
incomparable_get_subcategories(  
  archive_url = "https://www.theincomparable.com/gameshow/archive/"  
)
```

Arguments

archive_url E.g. "https://www.theincomparable.com/theincomparable/archive/".

Value

A tibble with subcategory links link and category name category

Examples

```
## Not run:
incomparable_get_subcategories("https://www.theincomparable.com/gameshow/archive/")

## End(Not run)
```

`incomparable_parse_archive`*Parse a show's archive page on The Incomparable website*

Description

Retrieves all episodes for one or more shows passed as a tibble. The archive page *does not* include full duration information, as it is limited to hours and minutes. Use `incomparable_parse_stats()` for accurate episode durations.

Usage

```
incomparable_parse_archive(archive_url)
```

Arguments

`archive_url` E.g. "https://www.theincomparable.com/theincomparable/archive/".

Value

A tibble, with following format:

```
#> dplyr::glimpse(incomparable_parse_archive(archive_url))
Columns: 12
 $ number <chr>
 $ title  <chr>
 $ date   <date>
 $ year   <dbl>
 $ month  <ord>
 $ weekday <ord>
 $ host   <chr>
 $ guest  <chr>
 $ category <chr>
 $ topic  <chr>
 $ summary <chr>
 $ network <chr>
```

Examples

```
## Not run:
archive_url <- "https://www.theincomparable.com/gameshow/archive/"
incomparable_parse_archive(archive_url)

## End(Not run)
```

 incomparable_parse_stats

Parse The Incomparable stats.txt files

Description

The stats.txt files have a slightly different format, especially the host/guest information may differ from what is returned by `incomparable_parse_archive()`, which implicitly assumes the first person mentioned to be the host of the episode. However, this data source does not include podcast subcategories (e.g. "Old Movie Club") or topic information, which is only available on the archive page.

Usage

```
incomparable_parse_stats(stats_url)
```

Arguments

stats_url URL to the stats.txt, e.g. "https://www.theincomparable.com/salvage/stats.txt".

Value

A tibble.

Examples

```
## Not run:
incomparable_parse_stats("https://www.theincomparable.com/salvage/stats.txt")

## End(Not run)
```

 label_n

Convenience function to display N

Description

Convenience function to display N

Usage

```
label_n(x, brackets = FALSE)
```

Arguments

x Data or singular value.
brackets Set TRUE to enclose result in ().

Value

A character of length 1.

Examples

```
label_n(100)
label_n(tibble::tibble(x = 1:10, y = 1:10), brackets = TRUE)
```

<code>parse_duration</code>	<i>Converting HH:MM:SS or MM:SS to hms</i>
-----------------------------	--

Description

Converting HH:MM:SS or MM:SS to hms

Usage

```
parse_duration(x)
```

Arguments

x A duration

Value

A numeric of durations in `hms::hms()`.

Note

Only needed to parse durations in The Incomparable `stats.txt` files.

Examples

```
parse_duration("32:12")
parse_duration("32:12:04")
```

relay_get_episodes *Retrieve all episodes for relay.fm shows*

Description

Retrieves all episodes for one or more shows passed as a tibble.

Usage

```
relay_get_episodes(relay_shows, cache = TRUE)
```

Arguments

relay_shows A tibble of shows, from relay_get_shows().
cache (logical(1)) Set to FALSE to disable caching.

Value

A tibble.

Examples

```
## Not run:  
relay_shows <- relay_get_shows()  
relay <- relay_get_episodes(relay_shows)  
  
## End(Not run)
```

relay_get_shows *Retrieve all relay.fm shows*

Description

Parses the show overview page and returns a tibble of show names with corresponding feed URLs, which in turn can then be passed to relay_parse_feed() individually.

Usage

```
relay_get_shows(cache = TRUE)
```

Arguments

cache (logical(1)) Set to FALSE to disable caching.

Value

A tibble with one row for each show

Examples

```
## Not run:  
relay_get_shows()  
  
## End(Not run)
```

relay_parse_feed	<i>Parse a relay.fm show feed</i>
------------------	-----------------------------------

Description

Parses a single feed and returns its content as a tibble.

Usage

```
relay_parse_feed(url)
```

Arguments

url A show's feed URL, e.g. "https://www.relay.fm/ungeniused/feed". Use relay_get_shows() to retrieve feed URLs.

Value

A tibble.

Examples

```
## Not run:  
relay_parse_feed(url = "https://www.relay.fm/ungeniused/feed")  
  
## End(Not run)
```

update_cached_data	<i>Update and cache data locally</i>
--------------------	--------------------------------------

Description

Update and cache data locally

Usage

```
update_cached_data(dir = "data_cache")
```

Arguments

dir ["data_cache"] Directory path to save cached data to.

Value

Nothing

Examples

```
## Not run:  
update_cached_data()
```

```
## End(Not run)
```

Index

atp_get_episodes, [2](#)
atp_parse_page, [3](#)

cache_podcast_data, [3](#)

gather_people, [4](#)

incomparable_get_episodes, [5](#)
incomparable_get_shows, [5](#)
incomparable_get_subcategories, [6](#)
incomparable_parse_archive, [7](#)
incomparable_parse_stats, [8](#)

label_n, [8](#)

parse_duration, [9](#)

relay_get_episodes, [10](#)
relay_get_shows, [10](#)
relay_parse_feed, [11](#)

update_cached_data, [11](#)