Package 'anytime'

February 9, 2017

Type Package

Title Anything to 'POSIXct' or 'Date' Converter

Version 0.2.1

Date 2017-02-09

Author Dirk Eddelbuettel

Maintainer Dirk Eddelbuettel <edd@debian.org>

Description Convert input in any one of character, integer, numeric, factor, or ordered type into 'POSIXct' (or 'Date') objects, using one of a number of predefined formats, and relying on Boost facilities for date and time parsing.

URL http://dirk.eddelbuettel.com/code/anytime.html

BugReports https://github.com/eddelbuettel/anytime/issues

License GPL (>= 2)

Encoding UTF-8

Depends R (>= 3.2.0)

Imports Rcpp (>= 0.12.9)

LinkingTo Rcpp (>= 0.12.9), BH

Suggests gettz

RoxygenNote 6.0.0

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-02-09 15:24:06

R topics documented:

anytime-package *Anything to 'POSIXct' or 'Date' Converter*

Description

Convert input in any one of character, integer, numeric, factor, or ordered type into 'POSIXct' (or 'Date') objects, using one of a number of predefined formats, and relying on Boost facilities for date and time parsing.

Details

R excels at computing with dates, and times. Using *typed* representation for your data is highly recommended not only because of the functionality offered but also because of the added safety stemming from proper representation.

But there is a small nuisance cost in interactive work as well as in programming. How often have we told `as.POSIXct()` that the origin is (of course) the epoch. Do we really have to say it again? Similarly, when parsing dates that are *somewhat* in YYYYMMDD format, do we really need to bother converting from integer or numeric or character or factor or ordered with one of dozen separators and/or month forms: YYYY-MM-DD, YYYY/MM/DD, YYYYMMDD, YYYY-mon-DD and so on?

So there may have been a need for a *general purpose* converter returning a proper `POSIXct` (or `Date`) object nomatter the input (provided it was somewhat parseable). `anytime()` tries to be that function.

The actual conversion is done by a combination of Boost lexical_cast to go from *(almost) anything* to string representation which is then parsed by Boost Date_Time.

Conversion is done by looping over a fixed set of formats until a matching one is found, or returning an error if none is found. The current set of conversion formulae is accessible in the source code, and can now also be accessed in R via `getFormats()`. Formats can be added via the `addFormats()` function.

Details on the Boost date format symbols are provided by the Boost date_time documentation and similar (but identical) to what `strftime` uses.

Author(s)

Dirk Eddelbuettel

References

Boost date_time: http://www.boost.org/doc/libs/1_61_0/doc/html/date_time.html

Formats used: https://github.com/eddelbuettel/anytime/blob/master/src/anytime.cpp#L30-L70

Boost format documentation: http://www.boost.org/doc/libs/1_61_0/doc/html/date_time/date_time_io.html#date_time.format_flags

Examples

```
library(anytime)
options(digits.secs=6)                ## for fractional seconds below
Sys.setenv(TZ=anytime:::getTZ())      ## helper function to try to get TZ

## integer
anytime(20160101L + 0:2)

## numeric
anytime(20160101 + 0:2)

## factor
anytime(as.factor(20160101 + 0:2))

## ordered
anytime(as.ordered(20160101 + 0:2))

## Dates: Character
anytime(as.character(20160101 + 0:2))

## Dates: alternate formats
anytime(c("20160101", "2016/01/02", "2016-01-03"))

## Datetime: ISO with/without fractional seconds
anytime(c("2016-01-01 10:11:12", "2016-01-01 10:11:12.345678"))

## Datetime: ISO alternate (?) with 'T' separator
anytime(c("20160101T101112", "20160101T101112.345678"))

## Short month '%b' (and full month is supported too)
anytime(c("2016-Sep-01 10:11:12", "Sep/01/2016 10:11:12", "Sep-01-2016 10:11:12"))

## Datetime: Mixed format (cf http://stackoverflow.com/questions/39259184)
anytime(c("Thu Sep 01 10:11:12 2016", "Thu Sep 01 10:11:12.345678 2016"))
```

`anytime` *Parse POSIXct objects from input data*

Description

These function use the Boost Date_Time library to parse datetimes (and dates) from strings, integers, factors or even numeric values (which are cast to strings internally). They return a vector of `POSIXct` objects (or `Date` objects in the case of `anydate`). `POSIXct` objects represent dates and time as (possibly fractional) seconds since the 'epoch' of January 1, 1970. A timezone can be set, if none is supplied 'UTC' is set.

Usage

```
anytime(x, tz = getTZ(), asUTC = FALSE)
```

```
anydate(x, tz = getTZ(), asUTC = FALSE)

utctime(x, tz = getTZ())

utcdate(x, tz = getTZ())
```

Arguments

x	A vector of type character, integer or numeric with date(time) expressions to be parsed and converted.
tz	A string with the timezone, defaults to the result of the (internal) getTZ function if unset. The getTZ function returns the timezone values stored in local package environment, and set at package load time. Also note that this argument applies to the *output*: the returned object will have this timezone set. The timezone is *not* used for the parsing which will always be to localtime, or to UTC is the asUTC variable is set (as it is in the related functions link{utctime} amd utcdate). So one can think of the argument as 'shift parsed time object to this timezone'. This is similar to what format() in base R does, but our return value is still a POSIXt object instead of a character value.
asUTC	A logical value indicating if parsing should be to UTC; default is false implying localtime.

Details

A number of fixed formats are tried in succession. These include the standard ISO format 'YYYY-MM-DD HH:MM:SS' as well as different local variants including several forms popular in the United States. Two-digits years and clearly ambigous formats such as '03/04/05' are ignored. In the case of parsing failure a NA value is returned.

Fractional seconds are supported as well. As R itself only supports microseconds, the Boost compile-time option for nano-second resolution has not been enabled.

Value

A vector of POSIXct elements, or, in the case of anydate, a vector of Date objects.

Notes

By default, the (internal) conversion to (fractional) seconds since the epoch is relative to the locatime of this system, and therefore not completely independent of the settings of the local system. This is to strike a balance between ease of use and functionality. A more-full featured conversion could be possibly be added with support for arbitrary reference times, but this is (at least) currently outside the scope of this package. See the **RcppCCTZ** package which offers some timezone-shifting and differencing functionality. As of version 0.0.5 one can also parse relative to UTC avoiding the localtime issue,

Times and timezones can be tricky. This package offers a heuristic approach, it is likely that some input formats may not be parsed, or worse, be parsed incorrectly. This is not quite a Bobby Tables situation but care must always be taken with user-supplied input.

The Boost Date_Time library cannot parse single digit months or days. So while '2016/09/02' works (as expected), '2016/9/2' will not. Other non-standard formats may also fail.

The is a known issue (discussed at length in issue ticket 5) where Australian times are off by an hour. This seems to affect only Windows, not Linux.

When given a vector, R will coerce it to the type of the first element. Should that be `NA`, surprising things can happen: `c(NA, Sys.Date())` forces both values to `numeric` and the date will not be parsed correctly (as its integer value becomes numeric before our code sees it). On the other hand, `c(Sys.Date(), NA)` works as expected parsing as type Date with one missing value. See issue ticket 11) for more.

Operating System Impact

On Windows systems, accessing the `isdst` flag on dates or times before January 1, 1970, can lead to a crash. Therefore, the lookup of this value has been disabled for those dates and times, which could therefore be off by an hour (the common value that needs to be corrected). It should not affect dates, but may affect datetime objects.

Author(s)

Dirk Eddelbuettel

References

This StackOverflow answer provided the initial idea: http://stackoverflow.com/a/3787188/143305.

See Also

anytime-package

Examples

```
## See the source code for a full list of formats, and the
## or the reference in help('anytime-package') for details
times <- c("2004-03-21 12:45:33.123456",
           "2004/03/21 12:45:33.123456",
           "20040321 124533.123456",
           "03/21/2004 12:45:33.123456",
           "03-21-2004 12:45:33.123456",
           "2004-03-21",
           "20040321",
           "03/21/2004",
           "03-21-2004",
           "20010101")
anytime(times)
anydate(times)
utctime(times)
utcdate(times)

## show effect of tz argument
```

```
anytime("2001-02-03 04:05:06")
## adjust parsed time to given TZ argument
anytime("2001-02-03 04:05:06", tz="America/Los_Angeles")
## somewhat equvalent base R functionality
format(anytime("2001-02-03 04:05:06"), tz="America/Los_Angeles")
```

`getFormats` *Functions to retrieve (or set) formats used for parsing dates.*

Description

The time and date parsing and conversion relies on trying a (given and fixed) number of timeformats. The format used is the one employed by the underlying implementation of the Boost date_time library.

Usage

```
getFormats()

addFormats(fmt)
```

Arguments

`fmt`	A vector of character values in the form understood by Boost date_time

Value

Nothing in the case of `addFormats`; a character vector of formats in the case of `getFormats`

Author(s)

Dirk Eddelbuettel

See Also

`anytime-package` and references therein

Examples

```
getFormats()
addFormats(c("%d %b %y",      # two-digit date [not recommended], textual month
             "%a %b %d %Y"))  # weekday weeknumber four-digit year
```

`iso8601` *Format a Datetime object: ISO 8601, RFC 2822 or RFC 3339*

Description

ISO 8601, RFC 2822 and RFC 3339 are a standards for date and time representation covering the formatting of date and time (with or without possible fractional seconds) and timezone information.

Usage

```
iso8601(pt)

rfc2822(pt)

rfc3339(pt)
```

Arguments

`pt`	A `POSIXt` Datetime or a `Date` object

Value

A character object formatted according to ISO 8601, RFC 2822 or RFC 3339

ISO 8601

ISO 8601 is described in some detail in https://en.wikipedia.org/wiki/ISO_8601 and covers multiple date and time formats.

Here, we interpret it more narrowly focussing on a single format each for datetimes and dates. We return datetime object formatted as '2016-09-01 10:11:12' and date object as '2016-09-01'.

RFC 2822

RFC 2822 is described in some detail in https://www.ietf.org/rfc/rfc2822.txt and https://en.wikipedia.org/wiki/Email#Internet_Message_Format. The Date and Time formating cover only a subset of the specification in that RFC.

Here, we use it to provide a single format each for datetimes and dates. We return datetime object formatted as 'Thu, 01 Sep 2016 10:11:12.123456 -0500' and date object as 'Thu, 01 Sep 2016'.

RFC 3339

RFC 3339 is described in some detail in https://tools.ietf.org/html/rfc3339 It refines both earlier standards.

Here, we use it to format datetimes and dates as single and compact strings. We return datetime object formatted as '2016-09-01T10:11:12.123456-0500' and date object as '2016-09-01'.

Author(s)

Dirk Eddelbuettel

References

https://en.wikipedia.org/wiki/ISO_8601, https://www.ietf.org/rfc/rfc2822.txt, https://en.wikipedia.org/wiki/Email#Internet_Message_Format, https://tools.ietf.org/html/rfc3339

Examples

```
iso8601(anytime("2016-09-01 10:11:12.123456"))
iso8601(anydate("2016-Sep-01"))

rfc2822(anytime("2016-09-01 10:11:12.123456"))
rfc2822(anydate("2016-Sep-01"))

rfc3339(anytime("2016-09-01 10:11:12.123456"))
rfc3339(anydate("2016-Sep-01"))
```

Index