Package 'isdparser'

October 23, 2018

Title Parse 'NOAA' Integrated Surface Data Files

Description Tools for parsing 'NOAA' Integrated Surface Data ('ISD') files, described at <https://www.ncdc.noaa.gov/isd>. Data includes for example, wind speed and direction, temperature, cloud data, sea level pressure, and more. Includes data from approximately 35,000 stations worldwide, though best coverage is in North America/Europe/Australia. Data is stored as variable length ASCII character strings, with most fields optional. Included are tools for parsing entire files, or individual lines of data.

Version 0.3.0

License MIT + file LICENSE

Encoding UTF-8

URL https://github.com/ropensci/isdparser

BugReports https://github.com/ropensci/isdparser/issues

LazyData true

VignetteBuilder knitr

Imports tibble (>= 1.2), data.table (>= 1.10.0)

Suggests roxygen2 (>= 6.1.0), testthat, rmarkdown, knitr

RoxygenNote 6.1.0

X-schema.org-applicationCategory Climate

X-schema.org-keywords climate, NOAA, data, ISD, stations

X-schema.org-isPartOf https://ropensci.org

NeedsCompilation no

Author Scott Chamberlain [aut, cre] (<https://orcid.org/0000-0003-1444-9135>)

Maintainer Scott Chamberlain <myrmecocystus@gmail.com>

Repository CRAN

Date/Publication 2018-10-23 20:00:03 UTC

R topics documented:

`isdparser-package` *Parse NOAA ISD Files*

Description

Parse NOAA ISD Files

Data format

Each record (data.frame row or individual list element) you get via `isd_parse` or `isd_parse_line` has all data combined. Control data fields are first, then mandatory fields, then additional data fields and remarks. Control and mandatory fields have column names describing what they are, while additional data fields have a length three character prefix (e.g., AA1) linking the fields to the documentation for the **Additional Data Section** at ftp://ftp.ncdc.noaa.gov/pub/data/noaa/ish-format-document.pdf

Data size

Each line of an ISD data file has maximum of 2,844 characters.

Control Data

The beginning of each record provides information about the report including date, time, and station location information. Data fields will be in positions identified in the applicable data definition. The control data section is fixed length and is 60 characters long.

Mandatory data

Each line of an ISD data file starts with mandatory data section. The mandatory data section contains meteorological information on the basic elements such as winds, visibility, and temperature. These are the most commonly reported parameters and are available most of the time. The mandatory data section is fixed length and is 45 characters long.

Additional data

Each line of an ISD data file has an optional additional data section, which follows the mandatory data section. These additional data contain information of significance and/or which are received with varying degrees of frequency. Identifiers are used to note when data are present in the record. If all data fields in a group are missing, the entire group is usually not reported. If no groups are reported the section will be omitted. The additional data section is variable in length with a minimum of 0 characters and a maximum of 637 (634 characters plus a 3 character section identifier) characters.

Remarks data

The numeric and character (plain language) remarks are provided if they exist. The data will vary in length and are identified in the applicable data definition. The remarks section has a maximum length of 515 (512 characters plus a 3 character section identifier) characters.

Missing values

Missing values for any non-signed item are filled (i.e., 999). Missing values for any signed item are positive filled (i.e., +99999).

Longitude and Latitude Coordinates

Longitudes will be reported with negative values representing longitudes west of 0 degrees, and latitudes will be negative south of the equator. Although the data field allows for values to a thousandth of a degree, the values are often only computed to the hundredth of a degree with a 0 entered in the thousandth position.

Author(s)

Scott Chamberlain <myrmecocystus@gmail.com>

---

isd_metadata *NOAA ISD metadata data.frame*

Description

This data.frame includes metadata describing all the data provided in ISD data files. And is used for transforming and scaling variables.

Format

A data frame with 643 rows and 19 columns

Details

Original csv data is in inst/extdata/isd_metadata.csv, collected from

The data.frame has the following columns:

- pos - (chr) position, if any
- category - (chr) category, one of additional-data section, control-data section, element quality data section, mandatory-data section, original observation data section, or remarks data section
- sub_category - (chr) sub category label, one of climate reference network unique data, cloud and solar data, ground surface data, hail data, marine data, network metadata, precipitation-data, pressure data, runway visual range data, sea surface temperature, soil temperature data, temperature data, weather occurrence data, weather-occurrence-data, or wind data
- abbrev - (chr) abbreviation, if any, NA for control and mandatory sections
- label - (chr) label, a top level label for the data, usually the same as the abbreviation
- sub_label - (chr) sub label, a more detailed label about the variable
- field_length - (int) field length, number of characters
- min - (chr) minimum value, if applicable, original
- min_numeric - (int) minimum value, if applicable, integer
- max - (chr) maximum value, if applicable, original
- max_numeric - (chr) maximum value, if applicable, integer
- units - (chr) units, if applicable
- scaling_factor - (chr) scaling factor, original
- scaling_factor_numeric - (int) scaling factor, integer, one of 1, 10, 100, 1000, or NA
- missing - (chr) value used to indicate missing data, original
- missing_numeric - (int) value used to indicate missing data, integer, one of 9, 99, 999, 9999, 99999, 999999, or NA
- description - (chr) short description of variable
- dom - (chr) long description of variable with categories
- dom_parsed_json - (list) NA if no categries, or a named list with category labels and their values

---

`isd_parse` *Parse NOAA ISD/ISH data files*

Description

Parse NOAA ISD/ISH data files

Usage

```
isd_parse(path, additional = TRUE, parallel = FALSE,
  cores = getOption("cl.cores", 2), progress = FALSE)
```

Arguments

path	(character) file path. required
additional	(logical) include additional and remarks data sections in output. Default: TRUE
parallel	(logical). do processing in parallel. Default: FALSE
cores	(integer) number of cores to use: Default: 2. We look in your option "cl.cores", but use default value if not found.
progress	(logical) print progress - ignored if parallel=TRUE. The default is FALSE because printing progress adds a small bit of time, so if processing time is important, then keep as FALSE

Value

A tibble (data.frame)

References

ftp://ftp.ncdc.noaa.gov/pub/data/noaa

See Also

isd_parse_line

Examples

```
path <- system.file('extdata/104270-99999-1928.gz', package = "isdparser")

(res <- isd_parse(path))

# with progress
(res2 <- isd_parse(path, progress = TRUE))

# only control + mandatory sections
(res <- isd_parse(path, additional = FALSE))

## Not run:
# in parallel
(out <- isd_parse(path, parallel = TRUE))

## End(Not run)
```

---

isd_parse_line *Parse NOAA ISD/ISH data files - line by line*

Description

Parse NOAA ISD/ISH data files - line by line

Usage

```
isd_parse_line(x, additional = TRUE, as_data_frame = TRUE)
```

Arguments

x	(character) a single ISD line
additional	(logical) include additional and remarks data sections in output. Default: TRUE
as_data_frame	(logical) output a tibble. Default: FALSE

Value

A tibble (data.frame)

References

ftp://ftp.ncdc.noaa.gov/pub/data/noaa

See Also

isd_parse

Examples

```
path <- system.file('extdata/024130-99999-2016.gz', package = "isdparser")
lns <- readLines(path, encoding = "latin1")
isd_parse_line(lns[1])
isd_parse_line(lns[1], FALSE)

res <- lapply(lns[1:1000], isd_parse_line)
library("data.table")
library("tibble")
as_data_frame(
 rbindlist(res, use.names = TRUE, fill = TRUE)
)

# only control + mandatory sections
isd_parse_line(lns[10], additional = FALSE)
isd_parse_line(lns[10], additional = TRUE)
```

---

isd_transform *Transform ISD data variables*

Description

Transform ISD data variables

Usage

```
isd_transform(x)
```

Arguments

x	(data.frame/tbl_df) data.frame/tbl from isd_parse or data.frame/tbl or list from isd_parse_line

Details

This function helps you clean your ISD data. isd_parse and isd_parse_line give back data without modifying the data. However, you'll likely want to transform some of the variables, in terms of the variable class (character to numeric), accounting for the scaling factor (variable X may need to be multiplied by 1000 according to the ISD docs), and missing values (unfortunately, missing value standards vary across ISD data).

Value

A tibble (data.frame) or list

operations performed

- scale latitude by factor of 1000
- scale longitude by factor of 1000
- scale elevation by factor of 10
- scale wind speed by factor of 10
- scale temperature by factor of 10
- scale temperature dewpoint by factor of 10
- scale air pressure by factor of 10
- scale precipitation by factor of 10
- convert date to a Date class with as.Date
- change wind direction to numeric
- change total characters to numeric

See Also

isd_parse, isd_parse_line

Examples

```
path <- system.file('extdata/104270-99999-1928.gz', package = "isdparser")
(res <- isd_parse(path))
isd_transform(res)

lns <- readLines(path, encoding = "latin1")
# data.frame
(res <- isd_parse_line(lns[1]))
isd_transform(res)
# list
(res <- isd_parse_line(lns[1], as_data_frame = FALSE))
isd_transform(res)
```

Index