Package ‘rdflib’

December 19, 2023

Title Tools to Manipulate and Query Semantic Data
Version 0.2.8

Description The Resource Description Framework, or ‘RDF’ is a widely used data representation model that forms the cornerstone of the Semantic Web. ‘RDF’ represents data as a graph rather than the familiar data table or rectangle of relational databases. The ‘rdflib’ package provides a friendly and concise user interface for performing common tasks on ‘RDF’ data, such as reading, writing and converting between the various serializations of ‘RDF’ data, including ‘rdfxml’, ‘turtle’, ‘nquads’, ‘ntriples’, and ‘json-ld’; creating new ‘RDF’ graphs, and performing graph queries using ‘SPARQL’. This package wraps the low level ‘redland’ R package which provides direct bindings to the ‘redland’ C library. Additionally, the package supports the newer and more developer friendly ‘JSON-LD’ format through the ‘jsonld’ package. The package interface takes inspiration from the Python ‘rdflib’ library.

License MIT + file LICENSE
Encoding UTF-8

URL https://github.com/ropensci/rdflib
BugReports https://github.com/ropensci/rdflib/issues
Imports redland, methods, utils, stringi, readr, dplyr, tidyr

RoxygenNote 7.2.3

Suggests magrittr, covr, testthat, knitr, rmarkdown, jqr, DT, tibble, purrr, lubridate, htr, xml2, jsonlite, repurrsive, nycflights13, spelling, jsonld

VignetteBuilder knitr
Language en-US

NeedsCompilation no

Author Carl Boettiger [aut, cre, cph] (<https://orcid.org/0000-0002-1642-628X>), Bryce Mecum [rev] (<https://orcid.org/0000-0002-0381-3766>), Anna Krystalli [rev] (<https://orcid.org/0000-0002-2378-4915>), Viktor Senderov [ctb] (<https://orcid.org/0000-0003-3340-5963>)

1
## Description

The Resource Description Framework, or RDF is a widely used data representation model that forms the cornerstone of the Semantic Web. "RDF" represents data as a graph rather than the familiar data table or rectangle of relational databases.

## Details

It has three main goals:

- Easily read, write, and convert between all major RDF serialization formats
- Support SPARQL queries to extract data from an RDF graph into a data.frame
- Support JSON-LD format as a first-class citizen in RDF manipulations

For more information, see the Wikipedia pages for RDF, SPARQL, and JSON-LD:

To learn more about rdflib, start with the vignettes: `browseVignettes(package = "rdflib")`

Configurations via `options()`

`rdf_print_format`:
- NULL or "nquads" (default)
- any valid serializer name: e.g. "rdfxml", "jsonld", "turtle", "ntriples"

`rdf_base_uri`:
- Default base URI to use (when serializing JSON-LD only at this time) default is "localhost://"

`rdf_max_print`:
- maximum number of lines to print from rdf, default 10

**Author(s)**

**Maintainer**: Carl Boettiger <cboettig@gmail.com> (ORCID) [copyright holder]

Other contributors:
- Bryce Mecum (ORCID) [reviewer]
- Anna Krystalli (ORCID) [reviewer]
- Viktor Senderov <vsenderov@gmail.com> (ORCID) [contributor]

**See Also**

Useful links:
- [https://github.com/ropensci/rdflib](https://github.com/ropensci/rdflib)
- Report bugs at [https://github.com/ropensci/rdflib/issues](https://github.com/ropensci/rdflib/issues)

---

**as_rdf**

*Coerce an object into RDF*

**Description**

Coerce an object into RDF

**Usage**

```r
as_rdf(
  x,
  rdf = NULL,
  prefix = NULL,
  base = getOption("rdf_base_uri", "localhost://"),
  context = NULL,
  key_column = NULL
)
```
Arguments

x  an object to coerce into RDF (list, list-like, or data.frame)
rdf  An existing rdf object, (by default a new object will be initialized)
prefix  A default vocabulary (URI prefix) to assume for all predicates
base  A base URI to assume for blank subject nodes
context  a named list mapping any string to a URI
key_column  name of a column which should be treated as the primary key in a table. must be unique

Examples

as_rdf(mtcars)
as_rdf(list(repo = "rdflib", owner = list("id", "ropensci")))

c.rdf  

Concatenate rdf Objects.

Description

All subsequent rdf objects will be appended to the first rdf object Note: this does not free memory from any of the individual rdf objects Note: It is generally better to avoid the use of this function by passing an existing rdf object to and rdf_parse or rdf_add objects. Multiple active rdf objects can cause problems when using disk-based storage backends.

Usage

## S3 method for class 'rdf'
c(...)

Arguments

...  objects to be concatenated

rdf  

Initialize an rdf Object

Description

Initialize an rdf Object
Usage

```
rdf(
    storage = c("memory", "BDB", "sqlite", "postgres", "mysql", "virtuoso"),
    host = NULL,
    port = NULL,
    user = NULL,
    password = NULL,
    database = NULL,
    charset = NULL,
    dir = NULL,
    dsn = "Local Virtuoso",
    name = "rdflib",
    new_db = FALSE,
    fallback = TRUE
)
```

Arguments

- **storage**: Storage backend to use; see details
- **host**: host address for mysql, postgres, or virtuoso storage
- **port**: port for mysql (mysql storage defaults to mysql standard port, 3306) or postgres (postgres storage defaults to postgres standard port, 4321)
- **user**: user name for postgres, mysql, or virtuoso
- **password**: password for postgres, mysql, or virtuoso
- **database**: name of the database to be created/used
- **charset**: charset for virtuoso database, if desired
- **dir**: directory of where to write sqlite or berkeley database.
- **dsn**: Virtuoso dsn, either "Local Virtuoso" or "Remote Virtuoso"
- **name**: name for the storage object created. Default is usually fine.
- **new_db**: logical, default FALSE. Create new database or connect to existing?
- **fallback**: logical, default TRUE. If requested storage system cannot initialize, should `rdf()` fall back on memory (default) or throw an error (fallback=FALSE)?

Details

An rdf Object is a list of class 'rdf', consisting of three pointers to external C objects managed by the redland library. These are the world object: basically a top-level pointer for all RDF models, and a model object: a collection of RDF statements, and a storage object, indicating how these statements are stored.

rdflib defaults to an in-memory hash-based storage structure, which should be best for most use cases. For very large triplestores, disk-based storage will be necessary. Enabling external storage devices will require additional libraries and custom compiling. See the storage vignette for details.

Value

- an rdf object
**Examples**

```r
x <- rdf()
```

---

**rdf_add**  
*Add RDF Triples*

**Description**

add a triple (subject, predicate, object) to the RDF graph

**Usage**

```r
rdf_add(
  rdf, 
  subject, 
  predicate, 
  object, 
  subjectType = as.character(NA), 
  objectType = as.character(NA), 
  datatype_uri = as.character(NA)
)
```

**Arguments**

- `rdf` an rdf object
- `subject` character string containing the subject
- `predicate` character string containing the predicate
- `object` character string containing the object
- `subjectType` the Node type of the subject, i.e. "uri", "blank"
- `objectType` the Node type of the object, i.e. "literal", "uri", "blank"
- `datatype_uri` the datatype URI to associate with a object literal value

**Details**

`rdf_add()` will automatically 'duck type' nodes (if looks like a duck...). That is, strings that look like URIs will be declared as URIs. (See URI). Predicate should always be a URI (e.g. URL or a prefix:string), cannot be blank or literal. Subjects that look like strings will be treated as Blank Nodes (i.e. will be prefixed with _:). An empty subject, "", will create a blank node with random name. Objects that look like URIs will be typed as resource nodes, otherwise as literals. An empty object "" will be treated as blank node. Set subjectType or objectType explicitly to override this behavior, e.g. to treat an object URI as a literal string. NAs are also treated as blank nodes in subject or object See examples for details.
### rdf_free

**Value**

Silently returns the updated RDF graph (rdf object). Since the rdf object simply contains external pointers to the model object in C code, note that the input object is modified directly, so you need not assign the output of rdf_add() to anything.

**References**


**Examples**

```r
rdf <- rdf()
rdf_add(rdf,
    subject="http://www.dajobe.org/",
    predicate="http://purl.org/dc/elements/1.1/language",
    object="en")
```

```r
## non-URI string in subject indicates a blank subject
## (prefixes to ":b0")
rdf_add(rdf, "b0", "http://schema.org/jobTitle", "Professor")
```

```r
## identically a blank subject.
## Note rdf is unchanged when we add the same triple twice.
rdf_add(rdf, "b0", "http://schema.org/jobTitle", "Professor",
    subjectType = "blank")
```

```r
## blank node with empty string creates a default blank node id
rdf_add(rdf, "", "http://schema.org/jobTitle", "Professor")
```

```r
## Subject and Object both recognized as URI resources:
rdf_add(rdf,
    "https://orcid.org/0000-0002-1642-628X",
    "http://schema.org/homepage",
    "http://carlboettiger.info")
```

```r
## Force object to be literal, not URI resource
rdf_add(rdf,
    "https://orcid.org/0000-0002-1642-628X",
    "http://schema.org/homepage",
    "http://carlboettiger.info",
    objectType = "literal")
```

---

**rdf_free**  
**Free Memory Associated with RDF object**
**rdf_free**

Description

Free Memory Associated with RDF object

Usage

```r
ddf_free(ddf, rm = TRUE)
```

Arguments

- `rdf` an `rdf` object
- `rm` logical, default `TRUE`. Remove pointer from `parent.frame()`? Usually a good idea since referring to a pointer after it has been removed can crash R.

Details

Free all pointers associated with an `rdf` object. Frees memory associated with the storage, world, and model objects.

Examples

```r
ddf <- ddf()
ddf_free(ddf)
rm(ddf)
```

---

**rdf_has_bdb**

Description

Check for BDB support

Usage

```r
ddf_has_bdb()
```

Value

TRUE if BDB support is detected, false otherwise

Examples

```r
ddf_has_bdb()
```
**rdf_parse**

*Parse RDF Files*

**Description**

Parse RDF Files

**Usage**

```r
rdf_parse(
  doc, 
  format = c("guess", "rdfxml", "nquads", "ntriples", "turtle", "jsonld"),
  rdf = NULL,
  base = getOption("rdf_base_uri", "localhost://"),
  ... 
)
```

**Arguments**

- `doc` : path, URL, or literal string of the rdf document to parse
- `format` : rdf serialization format of the doc, one of "rdfxml", "nquads", "ntriples", "turtle" or "jsonld". If not provided, will try to guess based on file extension and fall back on rdfxml.
- `rdf` : an existing rdf triplestore to extend with triples from the parsed file. Default will create a new rdf object.
- `base` : the base URI to assume for any relative URIs (blank nodes)
- `...` : additional parameters (not implemented)

**Value**

an rdf object, containing the redland world and model objects

**Examples**

```r
doc <- system.file("extdata", "dc.rdf", package="redland")
rdf <- rdf_parse(doc)
```
rdf_query

Perform a SPARQL Query

Description

Perform a SPARQL Query

Usage

rdf_query(rdf, query, data.frame = TRUE, ...)

Arguments

rdf
  an rdf object (e.g. from rdf_parse)
query
  a SPARQL query, as text string
data.frame
  logical, should the results be returned as a data.frame?
...
  additional arguments to a redland initialize-Query

Value

a data.frame of all query results (default.) Columns will be named according to variable names in the SPARQL query. Returned object values will be coerced to match the corresponding R type to any associated datatype URI, if provided. If a column would result in mixed classes (e.g. strings and numerics), all types in the column will be coerced to character strings. If data.frame is false, results will be returned as a list with each element typed by its data URI.

Examples

doc <- system.file("extdata", "dc.rdf", package="redland")

sparql <-

rdf <- rdf_parse(doc)
rdf_query(rdf, sparql)
Serialize an RDF Document

Usage

```r
df_serialize(
    rdf,
    doc = NULL,
    format = c("guess", "rdfxml", "nquads", "ntriples", "turtle", "jsonld"),
    namespace = NULL,
    prefix = names(namespace),
    base = getOption("rdf_base_uri", "localhost://"),
    ...
)
```

Arguments

- `rdf`: an existing rdf triplestore to extend with triples from the parsed file. Default will create a new rdf object.
- `doc`: file path to write out to. If null, will write to character.
- `format`: rdf serialization format of the doc, one of "rdfxml", "nquads", "ntriples", "turtle" or "jsonld". If not provided, will try to guess based on file extension and fall back on rdfxml.
- `namespace`: a named character containing the prefix to namespace bindings. names(namespace) are the prefixes, whereas namespace are the namespaces
- `prefix`: (optional) for backward compatibility. See namespace. It contains the matching prefixes to the namespaces in namespace and is set automatically if you provide namespace as a named character vector.
- `base`: the base URI to assume for any relative URIs (blank nodes)
- `...`: additional arguments to redland::serializeToFile

Value

`rdf_serialize` returns the output file path doc invisibly. This makes it easier to use `rdf_serialize` in pipe chains with `rdf_parse`.

Examples

```r
infile <- system.file("extdata", "dc.rdf", package="redland")
out <- tempfile("file", fileext = ".rdf")

some_rdf <- rdf_parse(infile)
```
read_nquads

read an nquads file

Description
read an nquads file

Usage
read_nquads(file, ...)

Arguments
file path to nquads file
...
additional arguments to rdf_parse()

Value
an rdf object. See rdf_parse()

Examples
tmp <- tempfile(fileext = ".nq")
library(datasets)
write_nquads(iris, tmp)
read_nquads(tmp)
write_nquads

write object out as nquads

Description

write object out as nquads

Usage

write_nquads(x, file, ...)

Arguments

x an object that can be represented as nquads
file output filename
... additional parameters, see examples

Examples

tmp <- tempfile(fileext = ".nq")
library(datasets)

## convert data.frame to nquads
write_nquads(iris, tmp)
rdf <- read_nquads(tmp)

## or starting a native rdf object
write_nquads(rdf, tempfile(fileext = ".nq"))
Index

as_rdf, 3

c.rdf, 4

rdf, 4
rdf_add, 6
rdf_free, 7
rdf_has_bdb, 8
rdf_parse, 9, 10, 11
rdf_parse(), 12
rdf_query, 10
rdf_serialize, 11
rdflib (rdflib-package), 2
rdflib-package, 2
read_nquads, 12

write_nquads, 13