Package ‘UniProt.ws’
March 23, 2017

Type Package
Title R Interface to UniProt Web Services
Version 2.14.0
Depends methods, utils, RSQLite, RCurl, BiocGenerics (>= 0.13.8)
Imports AnnotationDbi
Suggests RUnit
Author Marc Carlson
Maintainer Bioconductor Package Maintainer <maintainer@bioconductor.org>
Description A collection of functions for retrieving, processing and repackaging the UniProt web services.
test_uniProt.ws_package.R
License Artistic License 2.0
biocViews Annotation, Infrastructure, GO, KEGG, BioCarta
LazyLoad yes
NeedsCompilation no

R topics documented:

UniProt.ws-objects .......................................................... 1
UNIPROTKB .................................................................. 4

Index

UniProt.ws-objects UniProt.ws objects and their related methods and functions

Description

UniProt.ws is the base class for interacting with the Uniprot web services from Bioconductor.
In much the same way as an AnnotationDb object allows access to select for many other annotation packages, UniProt.ws is meant to allow usage of select methods and other supporting methods to enable the easy extraction of data from the Uniprot web services.

select, columns and keys are used together to extract data via an UniProt.ws object.
columns shows which kinds of data can be returned for the UniProt.ws object.
keytypes allows the user to discover which keytypes can be passed in to select or keys via the
keytype argument.
keys returns keys for the database contained in the UniProt.ws object. By default it will return
the primary keys for the database, which are UNIPROTKB keys, but if used with the keytype
argument, it will return the keys from that keytype.
select will retrieve the data as a data.frame based on parameters for selected keys and columns
and keytype arguments.
The UniProt.ws will be loaded whenever you load the UniProt.ws package. This object will be
set up to retrieve information from Homo sapiens by default, but this value can be changed to any of
the species supported by Uniprot. The species and taxId methods allow users to see what species
is currently being accessed, and taxId<- allows them to change this value.
species shows the genus and species label currently attached to the UniProt.ws objects database.
taxId shows the NCBI taxonomy ID currently attached to the AnnotationDb objects database.
Using the equivalently names replace method (taxId<-) allows the user to change the taxon ID,
and the species represented along with it.
availableUniprotSpecies is a helper function to list out the available Species along with their
official taxonomy IDs that are available by Uniprot. Because there are so many species represented
at UniProt, there is also a pattern argument that can be used to restrict the range of things returned
to be only those whose species names match the search term. Please remember when using this
argument that the Genus is always capitalized and the species never is.
lookupUniprotSpeciesFromTaxId is another helper that will look up the species of any tax ID
that is supported by Uniprot.

Usage

    columns(x)
    keytypes(x)
    select(x, keys, columns, keytype, ...)
    species(object)
    taxId(x)

    availableUniprotSpecies(pattern, n=Inf)
    lookupUniprotSpeciesFromTaxId(taxId)
    UniProt.ws(taxId, ...)

Arguments

x        the UniProt.ws object.
object   the UniProt.ws object.
keys     the keys to select records for from the database. All possible keys are returned
         by using the keys method.
columns  the columns or kinds of things that can be retrieved from the database. As with
         keys, all possible columns are returned by using the columns method.
keytype  the keytype that matches the keys used. For the select methods, this is used to
         indicate the kind of ID being used with the keys argument. For the keys method
         this is used to indicate which kind of keys are desired from keys
pattern  A string passed in to limit the results
n        the maximim number of results to return.
**UniProt.ws-objects**

```r
taxId a taxonomy id
...
other arguments
```

**Value**

`keys, columns, keytypes, species` and `lookupUniprotSpeciesFromTaxId` each return a character vector of possible values.

`taxId` returns a numeric value that corresponds to the taxonomy ID.

`select` and `availableUniprotSpecies` each return a data.frame.

**Author(s)**

Marc Carlson

**See Also**

`select`

**Examples**

```r
## Make a UniProt.ws object
up <- UniProt.ws(taxId=9606)

## look at the object
up

## get the current species
species(up)

## look up available species with their tax ids
availableUniprotSpecies("musculus")

## get the current taxId
taxId(up)

## look up the species that goes with a tax id
lookupUniprotSpeciesFromTaxId(9606)

## set the taxId to something else
taxId(up) <- 10090
up

## list the possible key types
head(keytypes(up))

## list the columns that can be retrieved
head(columns(up))

## list all possible keys of type entrez gene ID.
## (this process is not instantaneous)
if(interactive()){
  egs = keys(up, "ENTREZ_GENE")
}

## use select to extract some data
```
Description

This manual page enumerates the kinds of data represented by the values returned when the user calls columns or keytypes.

Details

All the possible values for columns and keytypes are listed below. Users will have to actually use these methods to learn which of the following possible values actually apply in their case.

- :UNIPROTKB The central ID for UniProt and swissprot
- :UNIPARC UniParc
- :UNIREF50 UniRef50
- :UNIREF90 UniRef90
- :UNIREF100 UniRef100
- :EMBL/GENBANK/DDBJ EMBL/GenBank/DDBJ
- :EMBL/GENBANK/DDBJ_CDS EMBL/GenBank/DDBJ CDS
- :PIR PIR
- :UNIGENE UniGene
- :ENTREZ_GENE Entrez Gene (GeneID)
- :GI_NUMBER* GI number
- :IPI IPI
- :REFSEQ_PROTEIN RefSeq Protein
- :REFSEQ_NUCLEOTIDE RefSeq Nucleotide
- :PDB PDB
- :DISPROT DisProt
- :HSSP HSSP
- :DIP DIP
- :MINT MINT
- :ALLERGOME Allergome
- :MEROPS MEROPS
- :PEROXIBASE PeroxiBase
- :PPTASEDB PptaseDB
- :REBASE REBASE
- :TCDB TCDB
:ORPHANET Orphanet
:PHARMGKB PharmGKB
:POMBASE PomBase
:PSEUDOCAP PseudoCAP
:RGD RGD
:SGD SGD
:TAIR TAIR
:TUBERCULIST TubercuList
:WORMBASE WormBase
:WORMBASE_TRANS WORMBASE_TRANSCRIPT WormBase Transcript
:WORMBASE_PROTEIN WormBase Protein
:XENBASE Xenbase
:ZFIN ZFIN
:EGGNOG eggNOG
:GENETREE GeneTree
:HOVERGEN HOVERGEN
:KO KO
:OMA OMA
:ORTHODB OrthoDB
:PROTCLUSTDB ProtClustDB
:BIOCYC BioCyc
:REACTOME Reactome
:UNIPATHWAY UniPathWay
:CLEANEX CleanEx
:GERMONTREX GermOnline
:DRUGBANK DrugBank
:GENOMERNAI GenomeRNAi
:NEXTBIO NextBio
:CITATION citations
:CLUSTERS clusters
:COMMENTS comments
:DOMAINS domains
:DOMAIN domain
:EC ec ID
:ID ID
:EXISTENCE existence
:FAMILIES families
:FEATURES features
:GENES genes
:GO go term
Author(s)
Marc Carlson

Examples

library(UniProt.ws)
up <- UniProt.ws(taxId=9606)
## List the possible values for columns
columns(up)
## List the possible values for keytypes
keytypes(up)
## get some values back
## list all possible keys of type entrez gene ID.
## (this process is not instantaneous)
if(interactive()){
  keys <- head(keys(up, keytype="UNIPROTKB"))
  keys
}
select(up, keys=c("P31946","P62258"), columns=c("PDB","SEQUENCE"),
keytype="UNIPROTKB")
## Index

*Topic **classes**
  UniProt.ws-objects, 1

*Topic **manip**
  UNIPROTKB, 4

*Topic **methods**
  UniProt.ws-objects, 1

*Topic **utilities**
  UNIPROTKB, 4

3D (UNIPROTKB), 4

AARHUS/GHENT-2DPAGE (UNIPROTKB), 4
AGD (UNIPROTKB), 4
ALLERGOME (UNIPROTKB), 4
ARACHNOSERVER (UNIPROTKB), 4
availableUniprotSpecies (UniProt.ws-objects), 1

BIOCYC (UNIPROTKB), 4

CGD (UNIPROTKB), 4
CITATION (UNIPROTKB), 4
class:UniProt.ws (UniProt.ws-objects), 1
CLEANEX (UNIPROTKB), 4
CLUSTERS (UNIPROTKB), 4
cols (UniProt.ws-objects), 1
columns (UniProt.ws-objects), 1
columns,UniProt.ws-method (UniProt.ws-objects), 1

COMMENTS (UNIPROTKB), 4
CONOSERVER (UNIPROTKB), 4

CYGD (UNIPROTKB), 4

DICTYBASE (UNIPROTKB), 4
DIP (UNIPROTKB), 4
DISPROT (UNIPROTKB), 4
DMDM (UNIPROTKB), 4
DNASU (UNIPROTKB), 4
DOmain (UNIPROTKB), 4
DOMAINS (UNIPROTKB), 4
DRUGBANK (UNIPROTKB), 4

EC (UNIPROTKB), 4
ECHOBASE (UNIPROTKB), 4
ECO2DBASE (UNIPROTKB), 4
ECOGENE (UNIPROTKB), 4

EGGNOG (UNIPROTKB), 4
EMBL/GENBANK/DDBJ (UNIPROTKB), 4
EMBL/GENBANK/DDBJ_CDS (UNIPROTKB), 4
ENSEMBL (UNIPROTKB), 4
ENSEMBL_GENOMES (UNIPROTKB), 4
ENSEMBL_GENOMES PROTEIN (UNIPROTKB), 4
ENSEMBL_GENOMES TRANSCRIPT (UNIPROTKB), 4

ENSEMBL_PROTEIN (UNIPROTKB), 4
ENSEMBL_TRANSCRIPT (UNIPROTKB), 4
ENTREZ_GENE (UNIPROTKB), 4
EUHCVD (UNIPROTKB), 4
EUPATHDB (UNIPROTKB), 4
EXISTENCE (UNIPROTKB), 4

FAMILIES (UNIPROTKB), 4
FEATURES (UNIPROTKB), 4
FLYBASE (UNIPROTKB), 4

GENECARDS (UNIPROTKB), 4
GENEFARM (UNIPROTKB), 4
GENES (UNIPROTKB), 4
GENETREE (UNIPROTKB), 4
GENOLIST (UNIPROTKB), 4
GENOMERNAI (UNIPROTKB), 4
GERMONLINE (UNIPROTKB), 4
GI_NUMBER* (UNIPROTKB), 4
GO (UNIPROTKB), 4
GO-ID (UNIPROTKB), 4

H--INVDB (UNIPROTKB), 4
HGNC (UNIPROTKB), 4
HOGENOM (UNIPROTKB), 4
HOVERGEN (UNIPROTKB), 4
HPA (UNIPROTKB), 4
HSSP (UNIPROTKB), 4

ID (UNIPROTKB), 4
INTERACTOR (UNIPROTKB), 4
INTERPRO (UNIPROTKB), 4
IPI (UNIPROTKB), 4

KEGG (UNIPROTKB), 4
keys (UniProt.ws-objects), 1
INDEX

keys, UniProt.ws-method (UniProt.ws-objects), 1
keytypes (UniProt.ws-objects), 1
keytypes, UniProt.ws-method (UniProt.ws-objects), 1
KEYWORD- ID (UNIPROT KB), 4
KEYWORDS (UNIPROTKB), 4
KO (UNIPROTKB), 4
LAST- MODIFIED (UNIPROTKB), 4
LEGIONIST (UNIPROTKB), 4
LENGTH (UNIPROTKB), 4
LEPROMA (UNIPROTKB), 4
lookupUniprotSpeciesFromTaxId (UniProt.ws-objects), 1
MAIZEGDB (UNIPROTKB), 4
MEROPS (UNIPROTKB), 4
MGI (UNIPROTKB), 4
MIM (UNIPROTKB), 4
MINT (UNIPROTKB), 4
NEXTBIO (UNIPROTKB), 4
NEXTPROT (UNIPROTKB), 4
OMA (UNIPROTKB), 4
ORGANISM (UNIPROTKB), 4
ORGANISM- ID (UNIPROTKB), 4
ORPHANET (UNIPROTKB), 4
ORTHODB (UNIPROTKB), 4
PATHWAY (UNIPROTKB), 4
PATRIC (UNIPROTKB), 4
PDB (UNIPROTKB), 4
PEROXIBASE (UNIPROTKB), 4
PHARMGKB (UNIPROTKB), 4
PHOSSITE (UNIPROTKB), 4
PIR (UNIPROTKB), 4
POMBASE (UNIPROTKB), 4
PPTASEDB (UNIPROTKB), 4
PROTCLUSTDB (UNIPROTKB), 4
PROTEIN NAMES (UNIPROTKB), 4
PSEUDOCAP (UNIPROTKB), 4
REACTOME (UNIPROTKB), 4
REBASE (UNIPROTKB), 4
REFSEQ_ NUCLEOTIDE (UNIPROTKB), 4
REFSEQ_ PROTEIN (UNIPROTKB), 4
REVIEWED (UNIPROTKB), 4
RGD (UNIPROTKB), 4
SCORE (UNIPROTKB), 4
select (UniProt.ws-objects), 1

select, UniProt.ws-method (UniProt.ws-objects), 1
SEQUENCE (UNIPROTKB), 4
SGD (UNIPROTKB), 4
show, UniProt.ws-method (UniProt.ws-objects), 1
species (UniProt.ws-objects), 1
species, UniProt.ws-method (UniProt.ws-objects), 1
TAXID (UNIPROTKB), 4
taxId (UniProt.ws-objects), 1
taxId, UniProt.ws-method (UniProt.ws-objects), 1
taxId<- (UniProt.ws-objects), 1
taxId<-, UniProt.ws-method (UniProt.ws-objects), 1
TAXONOMY- LINEAGE (UNIPROTKB), 4
TCDB (UNIPROTKB), 4
TIGR (UNIPROTKB), 4
TOOLS (UNIPROTKB), 4
TUBERCULIST (UNIPROTKB), 4
UCSC (UNIPROTKB), 4
UNIGENE (UNIPROTKB), 4
UNIPARC (UNIPROTKB), 4
UNIPATHWAY (UNIPROTKB), 4
UniProt.ws (UniProt.ws-objects), 1
UniProt.ws-class (UniProt.ws-objects), 1
UniProt.ws-objects, 1
UNIPROTKB, 4
UNIREF100 (UNIPROTKB), 4
UNIREF50 (UNIPROTKB), 4
UNIREF90 (UNIPROTKB), 4
VECTORBASE (UNIPROTKB), 4
VERSION (UNIPROTKB), 4
WORLD- 2DPAGE (UNIPROTKB), 4
WORMBASE (UNIPROTKB), 4
WORMBASE_ PROTEIN (UNIPROTKB), 4
WORMBASE_ TRANSCRIPT (UNIPROTKB), 4
XENBASE (UNIPROTKB), 4
ZFIN (UNIPROTKB), 4