Package ‘UniProt.ws’
January 31, 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:

<table>
<thead>
<tr>
<th>Topic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UniProt.ws-objects</td>
<td>1</td>
</tr>
<tr>
<td>UNIPROTKB</td>
<td>4</td>
</tr>
</tbody>
</table>

Index

<table>
<thead>
<tr>
<th>Topic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UniProt.ws-objects</td>
<td>UniProt.ws objects and their related methods and functions</td>
</tr>
</tbody>
</table>

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**

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
```
res <- select(up,
    keys = c("22627","22629"),
    columns = c("PDB","UNIGENE","SEQUENCE"),
    keytype = "ENTREZ_GENE")
head(res)

---

**UNIPROTKB**

*Descriptions of available values for columns and keytypes.*

**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
:PHOSSITE  PhosSite
:DMDM  DMDM
:AARHUS/GHENT-2DPAGE  Aarhus/Ghent-2DPAGE
:ECO2DBASE  ECO2DBASE
:WORLD-2DPAGE  World-2DPAGE
:DNASU  DNASU
:ENSEMBL  Ensembl
:ENSEMBL_PROTEIN  Ensembl Protein
:ENSEMBL_TRANSCRIPT  Ensembl Transcript
:ENSEMBL_GENOMES  Ensembl Genomes
:ENSEMBL_GENOMES_PROTEIN  Ensembl Genomes Protein
:ENSEMBL_GENOMES_TRANSCRIPT  Ensembl Genomes Transcript
:KEGG  KEGG
:PATRIC  PATRIC
:TIGR  TIGR
:UCSC  UCSC
:VECTORBASE  VectorBase
:AGD  AGD
:ARACHNOSERVER  ArachnoServer
:CGD  CGD
:CONOSERVER  ConoServer
:CYGD  CYGD
:DICTYBASE  dictyBase
:ECHOBASE  EchoBASE
:ECOGENE  EcoGene
:EUHCVDB  euHCVdb
:EUPATHDB  EuPathDB
:FLYBASE  FlyBase
:GENECARDS  GeneCards
:GENEFARM  GeneFarm
:GENOLIST  GenoList
:H-INVDB  H-InvDB
:HGNRC  HGNC
:HPA  HPA
:LEGIOLIST  LegioList
:LEPROMA  Leproma
:MAIZEGDB  MaizeGDB
:MIM  MIM
:MGI  MGI
:NEXT PROT  neXtProt
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:UniProt.ws (UniProt.ws-objects), 1
cleanex (UniProt.ws-objects), 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

dicybase (UNIPROTKB), 4
dip (UNIPROTKB), 4
disprot (UNIPROTKB), 4
dmm (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
<table>
<thead>
<tr>
<th>Key</th>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIZEGDB</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>MEROPS</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>MGI</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>MIM</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>MINT</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>NEXTBIO</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>NEXTPROT</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>OMA</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>ORGANISM</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>ORGANISM-ID</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>ORPHANET</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>ORTHODB</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PATHWAY</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PATRIC</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PDB</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PEROXIBASE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PHARMGKB</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PHOSSITE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PIR</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>POMBASE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PPTASEDB</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PROTEIN NAMES</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>PSEUDOCAP</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>REACTOME</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>REBASE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>REFSEQ_NUCLEOTIDE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>REFSEQ_PROTEIN</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>REVIEWED</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>RGD</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>SCORE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>select</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>SELECT</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>SEQUENCE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>SGD</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>show</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>species</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>taxId&lt;-&gt;</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>TAXONOMY-LINEAGE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>TCDB</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>TIGR</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>TOOLS</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>TUBERCULIST</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>UCSC</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>UNIGENE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>UNIPARC</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>UNIPATHWAY</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>UniProt.ws-class</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>UniProt.ws-objects</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>VERSION</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>VECTORBASE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>WORLD-2DPAGE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>WORMBASE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>WORMBASE_PROTEIN</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>WORMBASE_TRANSCRIPT</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>XENBASE</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>ZFIN</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>